

Road Surface Defect Detection – From Image-based to Non-image-based: A Survey

Jongmin Yu¹, Jiaqi Jiang², Sebastiano Fichera^{3,4}, Paolo Paoletti^{3,4}, Lisa Layzell⁴,
Devansh Mehta⁴, and Shan Luo^{2†} *†‡§¶

Abstract

Ensuring traffic safety is crucial, which necessitates the detection and prevention of road surface defects. As a result, there has been a growing interest in the literature on the subject, leading to the development of various road surface defect detection methods. The methods for detecting road defects can be categorised in various ways depending on the input data types or training methodologies. The predominant approach involves image-based methods, which analyse pixel intensities and surface textures to identify defects. Despite popularity, image-based methods share the distinct limitation of vulnerability to weather and lighting changes. To address this issue, researchers have explored the use of additional sensors, such as laser scanners or LiDARs, providing explicit depth information to enable the detection of defects in terms of scale and volume. However, the exploration of data beyond images has not been sufficiently investigated. In this survey paper, we provide a comprehensive review of road surface defect detection studies, categorising them based on input data types and methodologies used. Additionally, we review recently proposed non-image-based methods and discuss several challenges and open problems associated with these techniques.

*¹J. Yu is with the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Wilberforce Rd, Cambridge CB3 0WA, United Kingdom; Emails: jy522@cam.ac.uk (was affiliated with Department of Engineering, King’s College London, Strand, London, WC2R 2LS, United Kingdom)

†¹J. Jiang and S. Luo are with the Department of Engineering, King’s College London, Strand, London WC2R 2LS, United Kingdom; Emails: {jiaqi.1.jiang, shan.luo}@kcl.ac.uk

‡²S. Fichera and P. Paoletti are with the Department of Mechanical, Materials and Aerospace Engineering, School of Engineering, University of Liverpool, Liverpool, L69 3GH, United Kingdom; Emails: {Sebastiano.Fichera, paoletti}@liverpool.ac.uk

§³S. Fichera, P. Paoletti, L. Layzell and D. Mehta are with Robotiz3d Ltd, Daresbury, WA4 4FS, United Kingdom; Emails: {sebastiano.fichera, paolo.paoletti, lisa.layzell, devansh.mehta}@robotiz3d.com

¶⁴This work was supported by the Innovate UK SMART grant “ARRES PREVENT: The World-First Autonomous Road Repair Vehicle” (10006122).

† represents the corresponding author.

1 Introduction

Road conditions are crucial for traffic safety, economic development, and growth. Defective roads increase the likelihood of vehicle accidents and pose significant threats to traffic safety. In the United Kingdom, for example, poor or damaged road surfaces were the cause of 12.6% of all car accidents in 2020, according to the Department for Transport [1]. As autonomous vehicle driving becomes increasingly prevalent, it is essential to address road surface defects to ensure safe operations and enable fully autonomous deployment [2, 3].

Road defects also have negative environmental consequences. Research suggests that the condition of road surfaces has a significant impact on carbon emissions, with higher emissions being observed when road surfaces are in poor condition [4]. Furthermore, a study conducted in [5] demonstrated that poor road surface quality leads to a 2% increase in carbon emissions and a 5% increase in fuel consumption.

Monitoring road conditions is crucial for ensuring road safety and managing road infrastructure. The traditional practice of manual defect detection and maintenance is time-consuming and expensive and poses risks to the workers who perform maintenance tasks on the road. In recent years, there has been a significant increase in interest in developing automatic road defect detection systems to address these challenges. Thanks to the remarkable progress in the field of computer vision and machine learning, there have been rapid developments in visual road defect detection methods, including both traditional image processing techniques and deep learning methods, as extensively reported in [6–11].

Image-based methods are the most widely used approach for road surface defect detection, as reported in the literature [12–24]. Typically, these methods involve extracting informative features from road surface images and then detecting defects using these features. The detection of road defects can be performed using various methodologies, such as image segmentation [23–28] or object detection [19, 21, 22, 29]. In these approaches, it is generally assumed that the pixel intensities or texture information of defects differ from normal road surfaces [12, 13, 29].

Image-based methods dominate the literature on road defect detection, and several review articles have been published to survey these methods. For instance, in [6], image processing techniques for road crack detection are summarised. In [7], a review of various platforms and image processing approaches is provided for evaluating the conditions of asphalt surfaces. In [10], researchers can find a survey of Machine Learning-based crack detection algorithms. These review articles offer valuable insight into image-based road surface defect detection, providing researchers with a better understanding of the existing methods and their strengths and weaknesses.

However, existing surveys on road defect detection mainly focus on image-based methods and do not cover the recent trends of using non-image data for this purpose. In recent years, there has been a growing interest in utilising depth maps [30, 31] or point clouds [32, 33] for road defect detection. Compared to image data, non-image data is less sensitive to changes in climate and illu-

mination conditions, making it a more robust solution. Several methods have been proposed that utilise features obtained from non-vision sensors, such as laser scanners or LiDARs [30–35]. In addition, data-fusion approaches that integrate features from images and other sensing modalities have been developed. Given the importance of road defect detection as an industrial application and the need to accelerate advancements in this field, a new investigation that reflects the recent research trend is needed.

In this paper, our goal is to offer an exhaustive exploration of existing techniques for road surface anomaly detection and pinpoint potential directions for upcoming research. We commence by elaborating on the diverse materials employed in road construction and the variety of associated road defects. We also delve into an array of sensors employed to gather relevant data for pinpointing these road defects and introduce the datasets that are openly accessible in the literature. To mirror the current trajectory of research, we divide the forefront of road anomaly identification into three distinct strategies: techniques rooted in imagery, those that use non-image data, and those that incorporate data fusion. Initially, we assess the advancements in methods centred around images, emphasising how factors like climatic variations and lighting conditions might influence their efficacy. We then present an overview of road defect detection techniques that use other sensing modalities, such as depth sensing, to overcome the limitations of image-based methods [30–35, 35]. It is worth noting that, for the first time, academic researchers and industrial practitioners have collaborated in this survey to provide a more comprehensive understanding of this field.

This paper is organised as follows. In Section 2, we present an overview of road surfacing materials and different types of road defects. Information about various sensors, such as cameras or laser scanners, is provided in Section 3. Section 4 describes various publicly available road surface defect detection datasets. In Section 5, we categorise various defect detection methods based on the input data type and their methodologies and provide detailed information on each method. Finally, we conclude this paper in Section 6.

2 Road types and defects

Road surfaces are made up of various materials, each with its own characteristics and performance. The type of road defects that occur are also related to the material of the road. In this section, we will provide a detailed introduction to various materials used for road surfacing and discuss some representative road defect types that are mainly considered in road surface defect detection studies.

2.1 Road types depending on surfacing materials

In today’s society, there is a wide variety of road surfacing materials being used. As shown in Fig. 1, these materials include soil, gravel, Kankar, asphalt and



Figure 1: Images of various types of roads based on the materials used for road surfacing: (a) Earthen road, (b) Gravel road, (c) Kankar road, (d) WBM road, (e) Asphalt road, and (f) Concrete road.

concrete. Firstly, we will briefly introduce these road surfacing materials and compare their properties and characteristics.

An **earthen road** is typically constructed using natural soil and is considered a cost-effective option due to the abundance of soil. However, the road surface made from the soil is not durable as the physical properties of the road change with weather conditions. Earthen roads are commonly constructed in areas with low traffic, such as rural and countryside regions.

Stabilised Earthen Road is a solution that overcomes the limitations of a regular earthen road. It involves adding an additional compacted layer using stabilised soil, which can improve the road’s durability and strength. Stabilised soil is created by incorporating physical, chemical, or biological agents into ordinary soil, resulting in a more robust and durable material [36].

A **gravel road**, also known as a "metal road" or "dirty road", is a type of unpaved roads that is surfaced with gravel. The gravel used to construct the road is typically obtained from a quarry or stream bed. While these roads are common in developing countries and rural areas, they are also prevalent in many developed countries [37]. In fact, according to Rajkamal *et al.* [38], more than 75% of the road network in many developing countries consists of gravel and earth roads. Gravel roads are more durable than earthen roads and can withstand climate change. However, the surface of a gravel road made from natural gravel may not be uniform, damaging the lower part of vehicles.

A **Kankar road** is a type of road constructed using Kankar, a nodular variety of limestone that is spongy in nature and contains clayey and silicious

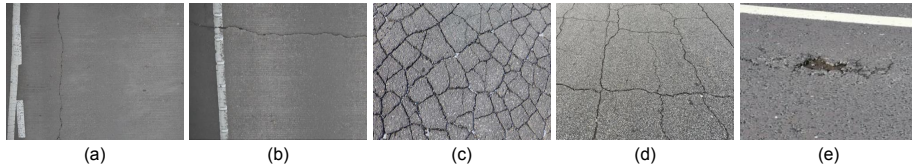


Figure 2: Example snapshots of various types of road surface defects: (a) Transverse crack, (b) Longitudinal crack, (c) Alligator crack, (d) Block crack, and (e) Pothole. It is important to note that the key difference between (a) and (b) is the angle with the road lane. If the crack runs parallel to the road lane, it is classified as a transverse crack. Conversely, if the crack is at a right angle to the lane, it is classified as a longitudinal crack.

matter. Kankar is a common surfacing material in countries with abundant limestone mines. However, the quality of the road surface depends on the quality of Kankar used. While a hard variety of Kankar may be as strong as stones, a soft variety may be weak. The construction cost of a Kankar road is similar to that of earthen and gravel roads, making it a popular choice in developing countries.

A **WBM road**, short for Water Bound Macadam road, is constructed with a stone border. The border of a WBM road is filled with sand, soil and very fine gravel called "Screening", which fastens the stone borders. The WBM road is more robust in the rain compared with an earthen road since the top-wearing surface is made up of stone gravel and a boundary. Moreover, it provides better functionality than gravel roads and Kankar roads that are constructed using single materials, as the soil helps to improve the flatness of the road. In the construction of the WBM road, a rolling machine is used to compact the stone border, while water is added to it during the rolling process. This kind of road is particularly effective and compacted in locations with heavy rainfall.

In the case of roads constructed with various pavement materials, as mentioned earlier, there is a lack of homogeneity in the pavement structure. For instance, gravel roads exhibit unevenness due to variations in gravel particle sizes. While these roads are not prone to surface cracks such as longitudinal or alligator cracks, they may develop potholes over time due to the sinking of the road caused by vehicle weight-induced pressure. Detecting such defects can be conveniently accomplished by utilising laser-based measurements to determine the distance of the road surface, as highlighted in a study by Pan *et al.* [39].

An **asphalt road** is a road surfaced with asphalt. Asphalt is a material that is produced through the refining of petroleum. To improve its durability and strength for road surfacing, it is mixed with aggregate, bitumen and sand and then heated and dried. This is then installed across a surface, such as an asphalt driveway, to form an asphalt road. Compared with the aforementioned road types, asphalt roads provide better functionality due to their smooth surface that boosts traction and skid resistance. Additionally, asphalt roads have an attractive price point due to their 100% recovery rate. Therefore, asphalt is the

primary material used to build heavy-traffic roads.

Most major roads in areas with high traffic are made of asphalt, which can make managing them expensive [40]. In fact, Highway Statistics [41] report that in 2016, approximately 94% of asphalt surfaces were made of asphalt. As a result, advanced maintenance technology is available for asphalt roads, unlike other roads such as earthen, gravel, or Kankar roads. This is reflected in the majority of studies which propose detection techniques specifically for defects occurring on asphalt roads.

A **concrete road** is made by mixing aggregates such as crushed rock, sand, cement, and water. The cement serves as the binder, holding the aggregate together as it dries into a hard, brittle surface that can crack and break, especially if the underlying surface is uneven. To address this issue, a mixture of cement and stone gravel is used when constructing concrete roads, making them more resistant to wear and tear problems like rutting, cracking, stripping loss of texture, and potholes commonly found in flexible pavement surfaces.

Concrete road paving has become increasingly popular in recent years, rivaling asphalt paving in popularity. This trend is mainly due to concrete roads' superior durability, longevity, and weather resistance. Unlike asphalt roads, which last for around 10 years, concrete roads have a lifespan of over 20 years due to their robustness. Additionally, concrete roads have low thermal expansion and contraction properties, which prevent surface defects from developing [42]. However, a major challenge with concrete roads is their repair, which requires the replacement of entire slabs rather than simple patching. Furthermore, the installation and maintenance costs for concrete roads are higher than those for asphalt.

2.2 Types of road defects

Road surfaces can develop various types of defects, including potholes, surface deterioration, edge failure, cracking, crazing, rutting, and subsidence, as illustrated in Fig. 2. These defects can arise due to various factors, such as climate change, vehicle accidents, and the natural ageing of road surfaces. In this survey paper, we provide detailed information on the typical patterns of road surface defects.

Longitudinal and transverse cracks are two common types of cracks found in roads. Longitudinal cracks run parallel to the road centerline, while transverse cracks run perpendicular to the traffic direction. Transverse cracks typically appear on both sides of the road and may eventually develop into horizontal cracks through the road. In some cases, these cracks may be accompanied by a small number of branching cracks.

A variety of factors can cause longitudinal cracks. Firstly, improper joint placement or construction can be a cause. This is because joints are often challenging to compact fully, resulting in a weak spot in the pavement. Secondly, a reflecting crack from a sublayer (not including joint reflection cracking) can cause a crack to develop directly on top of an underlying fracture due to minor movement of the underlying crack that stresses the overlying layer. Furthermore,

fatigue caused by Hot-Mix Asphalt is another significant factor contributing to longitudinal cracks. Initially, signs of fatigue cracking typically manifest in the longitudinal direction, either directly within or near the wheel path. In pavements with considerable thickness relative to the applied load, cracking can initiate from the pavement surface instead of originating at the bottom of the pavement layer. This occurrence is attributed to surface-induced stress generated by wheel loads, which, when combined with an aged and more brittle asphalt surface layer, leads to the formation of cracks.

Alligator crack, also known as crocodile crack, is a type of pavement distress that results in the formation of small acute polygons resembling a turtle pattern. These cracks are caused by repeated traffic loads that lead to longitudinal fissures along the wheel paths. As these fissures continue to develop, they eventually form a distinctive pattern resembling alligator skin. This type of pavement distress is typically due to a decrease in the base, sub-base, or subgrade support's capacity to support loads as a result of poor drainage or spring thaw problems. Other contributing factors may include overload, inadequate pavement structural design, insufficient thickness to support the intended loads, poor compaction and subgrade preparation during pavement construction, and improper pavement drainage design.

Block crack is characterised by interconnected cracks that divide the pavement into roughly rectangular pieces. In some cases, block cracking results from longitudinal and transverse cracks that break the pavement into chunks of various sizes. These cracks can be as small as a foot or as large as several yards square. As the cracks deepen and spread, moisture and vegetation growth further degrade the surface, rendering it unsuitable for foot traffic or vehicle use. The shrinkage of the surface material and temperature changes over time primarily causes block cracking. Proper mixing and placement of the asphalt mix, as well as the use of suitable asphalt binders, can help prevent the occurrence of block cracks.

A **pothole** is a sharply edged hole that forms on the upper layers of a road surface, typically due to untreated alligator cracks and moisture ingress. As alligator cracking worsens, small pavement sections become loose and can move under the weight of passing vehicles, leaving a gap in the surface layer that forms a pothole. Potholes may seem minor, but they can lead to serious consequences, including billions of dollars in vehicle damages and a significant proportion of highway fatalities. Unlike other types of road damage that have a negligible impact on driving, potholes can cause temporary loss of traction and lead to accidents. It is important to repair potholes promptly to prevent further deterioration of road surfaces and maintain safety for road users. According to the American Automobile Association, around 16 million drivers in the United States experienced vehicle damage caused by potholes in the five years leading up to 2016, resulting in punctured tyres, bent wheels, and damaged suspensions, with an annual cost of \$3 billion [43]. In the United Kingdom alone, repairing all potholes on the country's roads is estimated to cost £12 billion [44].

3 Data collection protocols

There are various types of sensors used to detect road surface defect detection, with cameras being the most commonly used. However, the image captured by cameras is susceptible to changes in climate and illumination conditions. Recent studies have proposed using depth sensors based on Time of Flight (ToF) sensors or laser scanning technology, such as LiDAR, to overcome this limitation. This section introduces different types of sensors used in road defect studies. We cover the sensors used in research and describe off-the-shelf platforms that utilise sensors to collect road surface information for defect detection.

3.1 RGB Camera

RGB cameras are widely used for detecting road surface defects, with many researchers and companies in the road management and defect-fixing fields relying on them to capture images of road surfaces [23–28]. From simple mobile cameras mounted on cell phones [45, 46] to ultra-high-resolution cameras of 4K or higher [29, 47–49], various types of RGB cameras have been used in both research and commercial products for road defect detection. Typically, methods for detecting road surface defects using RGB images aim to identify differences in texture and pixel intensities between the normal road surface and those with defects. Several studies have utilised RGB cameras for road surface defect detection, including [23, 49].

Mobile cameras have become popular for collecting road surface images due to their affordability and accessibility. Since almost everyone owns a mobile device, the biggest advantage is that it does not require additional costs to purchase a camera. Mei and Gul presented a deep neural network for detecting pavement cracks, using the camera module on an iPhone 5 to collect the road surface images [50]. The Crack500 [45] and CrackForest [46] datasets, commonly used for road defect segmentation, are composed of images taken using cell phones. However, the image sensor scale of mobile cameras is relatively small compared to general digital cameras, resulting in lower-resolution images. For example, an image in the CrackForest dataset has a resolution of 480×320 .

General digital cameras are another popular tool for collecting road surface data due to their relatively high resolution. Hu *et al.* used a Nikon digital camera to capture over 3,000 road surface pictures with a resolution of 2976×3968 [51]. Similarly, Yusof *et al.* used a digital camera to generate road surface image samples for training a convolutional neural network that detects cracks on asphalt pavement [52]. These cameras are useful in research and installed in off-the-shelf platforms for road defect detection.

One of the main advantages of using image sensors for road defect detection is their availability and the wealth of existing studies, which facilitates the application of various models. However, as discussed earlier, image-based road defect detection methods rely on detecting defects by analysing the difference between the texture and pixel values of the normal road surface and the defect. As shown in Fig. 3(a), road defects typically appear darker than normal

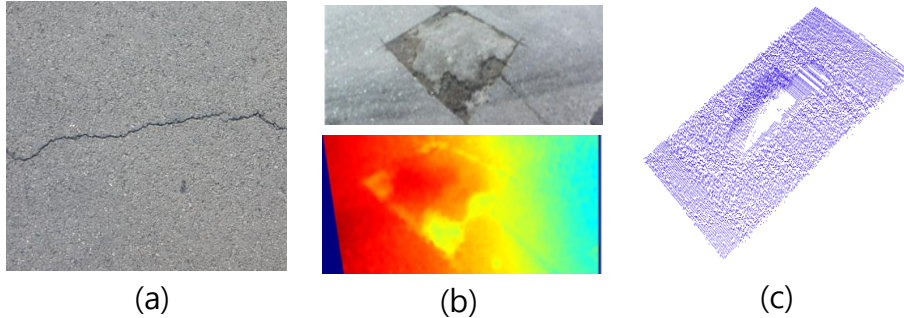


Figure 3: Example snapshots of data collected from various sensors, including RGB cameras, stereo cameras, and Time of Flight (ToF) based depth sensors: (a) An image from the CrackForest dataset [46] taken by an iPhone 5’s mobile camera. (b) An image of a road surface and the corresponding disparity map obtained by Fan *et al.* [53]. (c) Visualisation results of point clouds collected by Zhu *et al.* [54].

road surfaces due to their deeper nature. These challenges can compromise the robustness of image-based methods in certain scenarios, including lighting conditions (e.g., day versus night) and alterations in road surface texture resulting from weather changes. Moreover, changes in road materials or the presence of reflective surfaces can also impact the reliability of information captured by image sensors. Addressing these challenges often necessitates a substantial volume of image data, which can present a significant obstacle. This is why various explicit depth sensors, such as ToF-based or structural light-based sensors, are starting to become an alternative to cameras. However, visual cameras are still a predominant way to gather information for road defect detection since they are low-cost and easy to access. Numerous image-based road defect detection methods have been proposed and will be elaborated on later.

3.2 Stereo cameras

Numerous studies, including those by Li *et al.* [55], Zhang *et al.* [56], Thompson *et al.* [57], and Fan *et al.* [58], have demonstrated the feasibility of robust road defect detection using 3D methods. One commonly employed approach to obtain 3D information is through the use of a stereo camera, which captures images of the same subject from multiple angles and utilises the disparity that is the distance between two sets of coordinates of the cameras for the same scene point. This method offers a cost advantage as it allows for the estimation of 3D information without needing an additional sensor like a laser scanner to acquire depth information. Li *et al.* [55] employed a stereo camera consisting of two USB cameras for pothole detection, while Zhang *et al.* . Utilised a stereo camera constructed using two PointGrey Flea 3 cameras. Figure 3(b) presents an example of data collected from a stereo camera.

While using a stereo camera to obtain 3D information has its advantages, there are also drawbacks. One limitation is the difficulty in handling textureless or reflective surfaces. When a surface lacks distinctive features, the stereo-matching algorithm may struggle to find reliable correspondences, leading to inaccurate depth estimation in such areas.

Furthermore, deploying a practical stereo camera system requires manual calibration. Precise calibration is essential to ensure accurate depth estimation. Calibration involves determining the cameras' intrinsic parameters (focal length, distortion) and extrinsic parameters (relative position and orientation). Calibration errors can introduce inaccuracies into the depth maps, making the calibration process critical but sometimes time-consuming.

Despite these disadvantages, stereo cameras remain a popular choice for depth estimation due to their affordability, simplicity, and effectiveness within their limitations. However, it is important to consider these challenges and potential limitations when using stereo cameras for 3D information acquisition.

3.3 ToF-based and Structured light-based depth sensors

Several road defect detection approaches using depth sensors based on Time-of-Flight (ToF) and structured light-based methods have been introduced in recent years. Unlike stereoscopic cameras, which estimate depth based on the disparity of images from two cameras, these sensors measure distance directly by projecting laser or structured infrared rays onto the road surface. This makes them more robust to external environmental conditions and eliminates the need for heuristic parameters for camera calibration.

Among the structured light sensors commonly used for detecting road defects, one prevalent depth-sensing device is the Microsoft Kinect [59]. The Kinect utilises infrared sensors and has an approximate depth measurement accuracy of $\pm 5\%$ [57, 60, 61]. Another popular choice is the RealSense L515¹, a structured-light-based depth sensor extensively used for road defect detection [62, 63].

Compared to depth sensors based on stereo cameras, structured-light-based depth sensors such as the Kinect and RealSense L515 offer accurate depth measurements of the road surface and defects. This high level of accuracy is crucial for effectively detecting and characterizing road defects.

LiDARs have been used to detect road defects [58, 64, 65] among the ToF-based depth sensors. Similar to the structured-light-based depth sensors, these sensors measure the explicit distance between the sensor and the road surface, making them a more straightforward way to detect road damage compared to other data collection approaches. Data collected from LiDARs is represented as point clouds, which are discrete sets of data points in space, as shown in Fig. 3(c). This format enables the structural form of road defects to be expressed in three dimensions, in contrast to other data formats that can only capture the external form of road defects.

¹<https://www.intelrealsense.com/lidar-camera-1515/>

However, not all LiDARs are suitable for road surface assessment. For near-field perception, some LiDARs have a long perception range that may not be effective. Several commercial LiDARs are specialised for road assessment, such as the Velarray M1600². The XenoTrack³ and Trimble X7 Scanner⁴ are two other commercial LiDARs designed for road assessment. Trimble Geospatial⁵ offers the high-speed 3D laser scanning system, Trimble X7 Scanner, which simplifies adoption and increases efficiency, as well as complete field-to-finish mobile mapping solutions, such as Trimble MX9⁶ and Trimble MX50. Compared to Trimble MX9, Trimble MX50 has a slightly higher laser scan accuracy (2mm) and precision (2.5mm). Velodyne’s M1600 provides near-field perception up to 30 meters away and a wide vertical field of vision of 32 degrees. The L515 is a LiDAR depth camera that consistently delivers high accuracy across a range of 0.20m to 9m and provides over 23 million precise depth pixels per second, with a depth resolution of 1024 x 768 at 30 frames per second. On the other hand, the XenoTrack LiDAR system from Xenomatix⁷ is designed for road assessment and has high precision and high-resolution short-range capabilities that enable granular surface measurements and reveal intricate road geometries. The XenoTrack is mountable on vehicles and can be used instantly without complex setup procedures.

There are several off-the-shelf laser systems available for crack measurement, and one of the most popular is the Laser Crack Measurement System (LCMS[®]-2)⁸ from Pavemetrics⁹. LCMS-2 is a single-pass 3D sensor for pavement inspection that can be used to automatically geo-tag, measure, detect and quantify all key functional parameters of pavement. It provides a complete 1mm resolution automated pavement condition survey and can dramatically reduce labour costs.

4 Road defect detection datasets

In the early days of the research on road surface defect detection, datasets were not a major concern as most detection methods were based on image processing and did not require learning algorithms. However, with the rise of data-driven approaches using machine learning and deep learning, the importance of datasets has grown rapidly. While most publicly available datasets are still image-based, the use of depth sensors in research has resulted in an increasing number of datasets containing depth information. In this survey paper, we focus only on publicly accessible datasets, as private datasets are not accessible. This section provides an overview of various publicly available datasets for road

²<https://velodynelidar.com/products/velarray-m1600/>

³<https://xenomatix.com/lidar/xenotrack/>

⁴<https://geospatial.trimble.com/x7-scanner>

⁵<https://geospatial.trimble.com/>

⁶<https://geospatial.trimble.com/products-and-solutions/mobile-mapping>

⁷<https://xenomatix.com/>

⁸<https://www.pavemetrics.com/applications/road-inspection/lcms2-en/>

⁹<https://www.pavemetrics.com/>

defect detection.

4.1 RGB image-based datasets

Image-based datasets can be categorised into two types based on their annotation methods. The first type provides pixel-level annotations for road defects, which gives pixel-wise information to distinguish between normal road surfaces and defects. These datasets usually consist of road surface images and corresponding mask images that define whether a pixel belongs to a defect or not. The second type provides bounding-box annotations, which do not provide detailed information at the pixel level. However, compared to pixel-level annotations, which can only indicate two classes (i.e., general roads and defects), bounding-box annotations can be used to classify specific types of defects by comprehensively collecting information contained in the bounding boxes.

4.1.1 Pixel-level annotated datasets

Over the last decade, various pixel-level annotated datasets have been made available for pavement distress detection. The CrackForest dataset [46] is one such dataset, which comprises of 118 images of cracks on urban road surfaces in Beijing. These images were captured using an iPhone 5, resized to 480×320 pixels, and labelled with pixel-wise masks. The CRACK500 dataset [45] contains 250, 50, and 200 images for training, validation, and testing, respectively, and is further divided into 12 cropped parts for each image to increase the dataset size. This results in a dataset of 1,896, 328, and 1,124 images for training, validation, and testing, respectively. The German Asphalt Pavement Distress (GAPs) dataset [47] is a high-quality dataset that addresses the comparability issue in pavement distress detection. This standardised dataset contains 1,969 grayscale images with various classes of distress, such as cracks, potholes, and inlaid patches. The images have a resolution of 1920×1080 pixels with a per-pixel resolution of $1.2 \text{ mm} \times 1.2 \text{ mm}$. The Cracktree260 dataset is an extension of the dataset used in [48] and consists of 260 road pavement images captured by an area-array camera. The images are rotated with nine angles (from 0 to 90 degrees, at intervals of 10), flipped vertically and horizontally for each angle, and cropped five times on each flipped image (four in the corners and one in the centre). The CRKWH100 dataset [29] contains 100 road pavement images captured by a line-array camera with a ground sampling distance of 1 millimetre.

The NHA12D [66] dataset was collected by National Highways Surveying vehicles and consists of 80 pavement images, 40 of which depict concrete pavement and 40 depict asphalt pavement. Each image has a resolution of 1920×1080 and captures the view of a single lane on the A12, with some images including additional objects such as vehicles, road markings, road studs, and water stains. Concrete pavements have transversal and longitudinal joints every certain meter, which can pose a challenge for crack detection models to differentiate. For

different types of pavements, 25 images were taken using cameras with a vertical viewpoint, and the other 15 were taken with a forwarding camera.

The EdmCrack600 dataset [67] is widely considered one of the most challenging datasets due to several factors taken into consideration during the data collection. These include changes in weather conditions, significant environmental effects, and noise such as shadows, occlusion, stains, texture differences, and low contrast due to overexposure. Furthermore, blurring effects caused by the moving vehicle and poor lighting conditions add to the complexity of the dataset.

The CQU-BPDD dataset [68] consists of 60,056 bituminous pavement images captured automatically by in-vehicle cameras of a professional pavement inspection vehicle from various locations in southern China at different times. Each image corresponds to a 2×3 metres pavement patch of highways and has a resolution of 1200×900 pixels. The dataset contains seven different distress types, including transverse crack, massive crack, alligator crack, crack pouring, longitudinal crack, ravelling, repair, and normal pavement. Due to the large number of images, it is not feasible to provide pixel-level annotation. Therefore, the CQU-BPDD dataset provides patch-level annotation. Although the dataset does not provide precise pixel-wise annotation for defining defect locations, it does provide a naive location based on a small window or patch. The dataset also includes some challenging conditions, such as uneven illumination, which is manifested as weak light in some areas, and a variety of backgrounds in natural environments, such as crossings, foreign bodies, and ruts. Table 1 summarises the pixel-wise annotated datasets mentioned earlier in this paper.








4.1.2 Bounding-box annotated datasets

While pixel-wise annotated datasets provide a variety of data, datasets in the form of bounding boxes are still limited in number. Bounding boxes offer approximate locations of cracks, with a wider field of view compared to pixel-level-annotated datasets. The example images in Table 1 mainly capture close-up views of the road surface taken at right angles, while the example images in Table 1 showcasing bounding box datasets have a much wider field of view, including the vehicle’s front. Therefore, the availability of bounding box datasets is crucial for developing algorithms that can detect cracks in a wider range of scenarios.

The RDD2018 dataset [69] is a collection of 9,053 road damage images captured using a smartphone mounted on a car. The dataset includes 15,435 instances of road surface damage that were captured under different weather and illuminance conditions. Each image in the dataset has a resolution of 600×600 , and is annotated with a bounding box that locates the damage on the road surface and identifies its type.

The RDD2019 dataset [70] is an updated version of RDD2018 [69]. The dataset underwent a thorough review, relabelling and new annotated images were added to create RDD2019. This updated version contains a total of 13,135

Table 1: Publicly available datasets for road surface defect segmentation.

| Name | Format | Sensor | Resolution | Label type | Surface | Description | Sample |
|--|--------|-------------------------|------------|----------------------------|---------------------|--|---|
| CrackForest (CFD) [46] | RGB | Mobile cam (iPhone 5) | 480×320 | Pixel-wise (Binary class) | Asphalt | <ul style="list-style-type: none"> • 118 images • Contain noises such as shadows, oil spots, and water stains |  |
| Crack500 [45] | RGB | Mobile cam | 2000×1500 | Pixel-wise (Binary class) | Asphalt | <ul style="list-style-type: none"> • 250 images |  |
| German Asphalt Pavement Distress (GAPs) [47] | RGB | high-resolution cameras | 1920×1080 | Pixel-wise (Binary class) | Asphalt | <ul style="list-style-type: none"> • 1,969 grey valued images • various classes of distress such as cracks, potholes, inlaid patches |  |
| CrackTree260 [48] | RGB | Area-array camera | 800×500 | Pixel-wise (Binary class) | Asphalt | <ul style="list-style-type: none"> • 260 images • the crack is defined by-rotate the images with 9 different angles |  |
| CRKWH100 [29] | RGB | Line-array camera | 512×512 | Pixel-wise (Binary class) | Asphalt | <ul style="list-style-type: none"> • 100 road pavement images • Images captured by a line-array camera under visible-light illumination. |  |
| NHA12D [66] | RGB | High-resolution camera | 1920×1080 | Pixel-wise (Binary class) | Asphalt Concrete | <ul style="list-style-type: none"> • 80 pavement images • 40 concrete pavement images and 40 asphalt pavement images |  |
| EdmCrack600 [67] | RGB | GoPro 7 | 1920×1080 | Pixel-wise (Binary class) | Asphalt | <ul style="list-style-type: none"> • 600 pavement images • The data aim to cover various factors such as weather conditions, illumination conditions, and shadows from other objects. |  |
| CQU-BPDD [68] | RGB | In-vehicle cameras | 1200×900 | Patch-label (Binary class) | Asphalt | <ul style="list-style-type: none"> • 60,056 pavement images • Illumination is uneven, which is manifested as weak light in some areas • Consider a variety of backgrounds in real environments, such as crossings, foreign bodies, ruts, etc. |  |
| Stone331 [49] | RGB | Area-array camera | 1024×1024 | Pixel-wise (Binary class) | Stone | <ul style="list-style-type: none"> • 331 images of stone surfaces • detect stone crack observed in a cutting surface on a stone. |  |

images, which is an increase from 9,053 in RDD2018, and has 30,989 annotations, which is a substantial increase from 15,435 annotations in RDD2018. RDD2019 includes road images captured in Japan, and each image is annotated with a bounding box that locates the damage on the road and its type.

RDD2020 [71] is a large-scale dataset that comprises 26,336 road images captured in India, Japan, and the Czech Republic. The dataset includes over 31,000 instances of road damage and is designed for developing deep learning-based methods to automatically detect and classify various types of road damage. RDD2020 covers a range of common road damage types, such as longitudinal cracks, transverse cracks, alligator cracks, and potholes. The images in RDD2020 were captured using cost-effective vehicle-mounted smartphones, enabling a low-cost method for monitoring road pavement surface conditions. The dataset includes images with resolutions of either 600×600 or 720×720 , providing a detailed view of the road surface. RDD2020 can be a useful resource for municipalities and road agencies seeking to develop cost-effective methods for monitoring and maintaining road conditions.

RDD2022 [72] is the latest and most extensive version of the RDD datasets. Compared to its predecessors, RDD2022 offers the largest number of image samples, including 47,420 road surface images that contain 55,007 instances of road defects. Additionally, the dataset includes images with various resolutions, such as 512×512 , 600×600 , 712×712 , and 3650×2044 , which enables researchers to evaluate the performance of algorithms across different resolutions. RDD2022 provides an essential resource for researchers working on developing and training machine learning algorithms to detect and classify different types of road defects accurately.

It is worth noting that several bounding-box-annotated datasets have been used for road defect detection studies; however, this survey paper focuses on providing information on publicly available datasets only.

4.2 Depth sensor-based datasets



Recently, many defect detection studies have utilised explicit depth sensors such as laser scanners and LiDAR, resulting in the release of various datasets containing point clouds or depth maps. Point clouds consist of a discrete set of data points in space and are usually represented by a list of 3D points. In contrast, a depth map can be considered as a structured point cloud, where the structure is more similar to that of an image frame, and each pixel value defines a depth value. Compared to point clouds, which are unstructured and only provide the coordinates of each point, depth maps offer a more structured representation of the data. Some point cloud datasets also contain colour codes for each point. These datasets provide valuable resources for researchers developing algorithms to process and analyse point cloud data for defect detection and other applications.

The SHREC 2022 dataset [57] is a comprehensive dataset created specifically for semantic segmentation of potholes and cracks on road surfaces. The dataset consists of 4,340 image/mask pairs, which were compiled from 5 publicly

Table 2: Publicly available datasets for road surface defect detection.

| Name | Format | Sensor | Resolution | Label type | Surface | Description | Sample |
|-----------------|--------|--|--|-----------------------------------|---------|---|---|
| RDD2018 [69] | RGB | LG Nexus 5Xa | 600×600 | Bounding boxes (Multi-classes) | Asphalt | <ul style="list-style-type: none"> 9,053 road damage images 15,435 instances of road surface damage |  |
| RDD2019 [70] | RGB | In-vehicle camera | 600×600 | Bounding boxes (Multi-classes) | Asphalt | <ul style="list-style-type: none"> 13,135 road images 30,989 instances of road damages Nine categories, including road damages and wheel marks |  |
| RDD2020 [71] | RGB | LG Nexus 5X Samsung Galaxy J6 | 600×600 720×720 | Bounding boxes (Multi-classes) | Asphalt | <ul style="list-style-type: none"> 26,336 road images collected from India, Japan, and the Czech Republic 31,000 instances of road damages Four damage categories: Longitudinal Cracks, Transverse Cracks, Alligator Cracks and Potholes. |  |
| RDD2022 [72] | RGB | Smartphones High-resolution Cameras Google Street View | 512×512 600×600 720×720 3650×2044 | Bounding boxes (Multi-classes) | Asphalt | <ul style="list-style-type: none"> 47,420 road images from Japan, India, the Czech Republic, Norway, the United States, and China 55,000 instances of road damage Four damage categories: Longitudinal Cracks, Transverse Cracks, Alligator Cracks and Potholes. |  |

Table 3: Publicly available datasets containing depth information for road surface defect segmentation.

| Name | Format | Sensor | Resolution | Label type | Surface | Description | Sample |
|---------------------|--------|----------------------|----------------------|------------------------------|---------|--|---|
| SHREC2022 [57] | RGB+D | Luxonis OAK-D camera | Multiple resolutions | Pixel-wise (Binary class) | Asphalt | <ul style="list-style-type: none"> 4340 image pairs (RGB image and segmentation mask) 797 non-annotated RGB-D video clips Depth is estimated by disparity map |  |
| Pothole-600 [58] | RGB+D | ZED stereo camera | 400×400 | Pixel-wise (Binary class) | Asphalt | <ul style="list-style-type: none"> 600 of image and depth map pairs Depth is estimated by disparity map |  |

available datasets and augmented by a small number of manually segmented images. The images in the dataset vary in resolution and have been divided into training, validation, and test sets in 77/11/12 ratios. Furthermore, the dataset includes 797 non-annotated RGB-D video clips, each containing an RGB video and a disparity map video. Participants can extract additional images from these clips to enhance the dataset.

The Pothole-600 dataset [58] was collected using a ZED stereo camera. Three datasets containing 67 pairs of stereo images were created using Search Range Propagation (SRP) to estimate the road disparity images [73]. The image resolutions of datasets 1, 2 and 3 are 1028×1730 , 1030×1720 , and 1028×1710 pixels, respectively, resulting in a total of 79 potholes captured.

Overall, the above depth-sensor-based datasets are a valuable resource for researchers and practitioners working on the problem of semantic segmentation of road surface potholes and cracks. However, on the other side, to our best knowledge, there are currently no publicly available datasets that exclusively consist of point cloud data. Instead, a few datasets have been released that include both images and their corresponding depth maps. Although several methods using point clouds have been proposed, these datasets have not been publicly released. Unfortunately, there is a lack of publicly available datasets for point cloud analysis, which hinders the development and evaluation of new methods in this field.

5 Road surface defect detection

This section comprehensively reviews the various methods for detecting road surface defects. Due to the limited number of studies available on this topic, we refer to various reviews related to defect detection, such as texture fault detection [74] and flat steel crack detection [75]. Most researchers in the field mainly rely on image-based techniques that use RGB images to detect defects on material surfaces. They typically group defect detection techniques into different categories based on certain characteristics, although the exact taxonomy can vary depending on the researcher. One commonly used categorisation scheme is based on the type of approaches used. For example, Timm *et al.* [74] broadly divides techniques for detecting texture faults into two categories: local and global. Meanwhile, defect detection techniques are often classified as classification-based, local-abnormality-based, and template-matching-based methods. Another categorisation scheme used by Youkachen *et al.* [76] groups defect detection techniques into statistical, network-based, proximity-based, deviation-based, and statistical models.

The reviews mentioned above primarily focus on methodologies for defect detection using image-based methods, and their taxonomy was limited to this type of input data. However, in this paper, we aim to broaden the taxonomy by not only considering categories for classifying the defect detection methodologies but also considering the types of input data used. Specifically, we classify road surface defect detection methods based on their input data type, *i.e.* image-

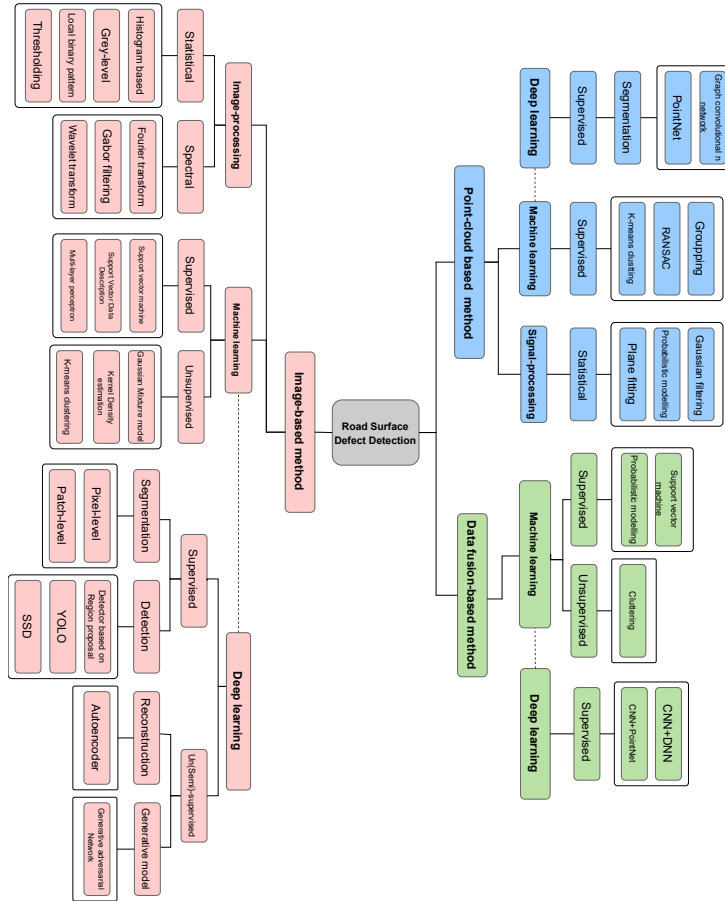


Figure 4: The taxonomy of road surface defect detection methods is structured as follows: The red-coloured categories represent the methodologies that use RGB images. The blue-coloured categories indicate the class of point cloud-based methods. The green-coloured categories define the methods using RGB images and depth values simultaneously. Deep learning is a sub-category of machine learning. However, since many methods based on deep learning have been proposed, it is connected to machine learning methods by dotted lines to represent methodological independence.

based methods, point cloud-based methods, and data fusion-based methods. We then categorise these methods based on their underlying methodologies, including image processing-based, machine learning-based, and deep learning-based methods. To visualise this taxonomy structure, Fig. 4.2 provides an overview of the road surface defect detection classification scheme. By incorporating both the input data types and the underlying methodologies used, our taxonomy provides a more comprehensive framework for understanding and comparing different approaches to road surface defect detection.

5.1 Image-based methods

Compared to other sensors, such as laser scanners and LiDARs, cameras are a relatively inexpensive way to capture road surface images. As a result, RGB images are the most commonly used input data for road surface defect detection. We categorise image-based methods into three groups: image-processing-based, machine-learning-based, and deep-learning-based. In this section, we provide detailed descriptions of each of these methods. Image-processing-based methods typically involve using various algorithms to extract features from road surface images, which are then used to detect defects. Machine learning-based methods involve training a model on a dataset of labelled road surface images to learn patterns associated with different types of defects. Deep learning-based methods use neural networks with multiple layers to automatically learn features and patterns from raw images. Table 4 provides an overview of the most commonly used image-based methods for road defect detection.

5.1.1 Image processing-based approaches

The group of image-processing-based methods encompasses a variety of techniques that leverage statistical or spectral methods to analyse digital images. In this section, we introduce four statistical methods and three spectral methods, providing an overview of each.

Statistical: Statistical methods for surface defect detection can be broadly classified into four categories: thresholding-based, local binary pattern-based, grey-level-based, and histogram-based methods. Among these, thresholding methods are the most commonly used for detecting defective regions on flat surfaces and are widely employed in Automated Visual Inspection (AVI) systems. However, traditional thresholding methods that rely on a fixed pixel value for defect detection are susceptible to random noise and non-uniform illuminations. To overcome this limitation, Djukic and Spuzic [77] proposed an adaptive thresholding approach that estimates the probability distribution of pixel intensities. This dynamic thresholding approach accurately separates true defects from random noise, and has been successfully applied in several real-world scenarios.

The local binary pattern (LBP) operator is another widely used classical technique for characterising local texture features of images. One of the main advantages of LBP is its invariance to rotation and grey-level changes. Initially proposed in 1994 by Ojala *et al.* [78], LBP has been frequently used for detect-

ing defects on flat steel surfaces [79–81]. However, the original LBP has some limitations, including weak global descriptive ability and sensitivity to noise. To address these issues, several LBP variants have been developed by modifying the threshold or scale of the original LBP. These variants have been widely used in defect detection of flat steel surfaces. For example, Wang *et al.* [82] proposed an LBP-inspired feature extractor that estimates the variations in four directions, including horizontal, vertical, and two diagonal directions, simultaneously. LBP variants can be applied as lightweight feature descriptors in both defect detection and classification tasks. Looking ahead, future trends in automated visual inspection systems involve the development of more noise-robust and scale-invariant LBP variants or LBP-like descriptors. This would enable more accurate and efficient detection of defects in a wide range of materials and conditions.

Grey-level statistics is an alternative approach that can be used for defect detection. Direct thresholding techniques may prove ineffective for defect detection in low-contrast images, making it necessary to analyse the distribution of grey levels in the image before performing the thresholding operation. Yang *et al.* [83] utilised features such as the mean value and distribution of pixels from the steel surface background to simultaneously differentiate between bright and dark defect objects. To ensure insensitivity to noise, Choi and Kim [84] employed a spectral-based approach to estimate the distribution of the background and then refined the defective regions locally to obtain probabilistic estimates. This method is superior to previous defect detection methods and provides robust outcomes even in noisy environments. However, the applicability of these methods for surface defect detection is limited by the diversity of surface defects encountered in various scenarios. Therefore, developing more sophisticated grey-level statistics-based methods is necessary to effectively handle various types of surface defects encountered in different scenarios. These methods can potentially offer more accurate and reliable results in defect detection and classification tasks, especially in low-contrast images.

Spectral: Despite statistical approaches that have been extensively researched for surface defect detection, they may not be reliable for detecting subtle defects with low-contrast intensity transitions. This is particularly true when illumination changes or pseudo defects are frequently present. As a result, there is a growing need for emerging Automated Visual Inspection (AVI) methods to detect surface defects in real-world environments.

Spectral approaches can be broadly classified into transform-based and filter-based methods. Transform-based methods employ Fourier Transform (FT) to achieve translation, expansion, and rotation invariance of image features. However, images from the real world often require additional processing to enhance image quality, such as denoising and contrast enhancement. For instance, longitudinal cracks on complex backgrounds can be detected by computing features with translational invariance using the Fourier amplitude spectrum of each subband [85]. Another unsupervised method based on Phase-Only Transform (PHOT) was proposed in [86] inspired by discrete FT, effectively highlighting irregular patterns to indicate defects. This novel method demonstrated its effec-

tiveness and versatility on various textured surfaces such as wood, steel, ceramic, and silicon wafers. Although transform-based methods have numerous advantages in terms of translation, expansion, and rotation invariance, they may have limitations in handling noise and illumination variation.

Wavelet transform offers the advantage of adjusting the time-frequency window and adapting it with the frequency change in the window’s centre. Additionally, the characteristics of wavelets are more compatible with the human visual system, enabling the extraction of information from signals and the execution of multiscale analysis of functions or signals through scaling and shifting operations. Therefore, wavelet-based methods have been extensively used for image analysis, including surface defect detection.

Filter-based methods operate in the frequency domain to represent an image, which can lead to the neglect of local descriptive information in the spatial domain. This limitation can be addressed by Gabor filters, which modulate a Gaussian kernel function on a sinusoidal wave with a specific frequency, resulting in both spatial and frequency domain analysis [87]. A simple 2-D Gabor filter can achieve localised and oriented frequency analysis [88]. However, it is important to note that Gabor filters can be computationally expensive, and there may be better options available for optimisation, such as Finite Impulse Response (FIR) filters, which offer more freely adjustable parameters compared to Infinite Impulse Response (IIR) and Gabor filters, as suggested in [89].

5.1.2 Machine learning-based methods

The machine learning group comprises a range of approaches that utilise machine learning algorithms such as Support Vector Machine (SVM) [91] or Neural Networks [92, 93] to analyse images based on hand-crafted features (in contrast to the features automatically learned by neural networks in deep learning based methods discussed later in Section 5.1.3). These approaches can be broadly categorised into two groups: supervised and unsupervised methods, depending on whether annotated data is available or not.

Supervised: Supervised methods use explicit annotation to define whether the input image contains defect patterns or not. SVMs have been used in [91] to distinguish between potholes or normal road surfaces in cropped image patches. Similarly, neural networks were used for path-level defect detection in [92] and [93]. In these methods, the input image is first divided into smaller patches, and hand-crafted features are extracted from each patch. The features are then used to train machine learning-based models, which are subsequently deployed to identify defects in the road surface. While these approaches demonstrated better performances than simple image-processing-based methods, the hand-crafted features they rely on may not be robust to changes in illumination conditions caused by weather or day-and-night variations. Moreover, these methods are limited in covering complex feature distributions.

Unsupervised: Unsupervised methods aim to detect road defects without explicit annotation to represent them. Instead, these methods derive a model from representing normal road surface features [94, 95, 105–107]. They assume

Table 4: List of image-based methods.

| Cat | Subcat | Reference | Year | Based method | Output protocol | Details | |
|-------------------------------|-------------------------------|-----------------------------|--------------------------|---------------------------|---|--|--|
| Image processing-based method | Statistical | Djukic and Spuzic [77] | 2007 | Kernel estimation | Pixel-level (Binary) | Estimate the probability distribution of pixel intensities and apply adaptive thresholding to detect defects. | |
| | | Wang <i>et al.</i> [82] | 2019 | LBP | Patch-level (Binary) | Propose lightweight feature descriptor. Estimating the variations of four directions simultaneously, which are horizontal, vertical, and two diagonal directions, so that the features are more robust to directional variation. | |
| | | Yang <i>et al.</i> [83] | 2008 | Gray-level statistic | Pixel-level (Binary) | Utilise the statistics of features <i>i.e.</i> mean value and distribution of pixels. Separate bright and dark areas on a road surface to distinguish between normal road surfaces and defects. | |
| | Spectral | Aiger and Talbot [86] | 2010 | Fourier transform | Patch-level (Binary) | Propose an unsupervised method based on the Fourier transform. The proposed method can persist only irregular patterns to present defects. | |
| | | Liu and Yan [90] | 2014 | Wavelet transform | Patch-level (Binary) | Present a novel wavelet-based image filtering algorithm based on anisotropic diffusion for detecting road defect detection. | |
| | Machine learning-based method | Supervised | Lin and Liu [91] | 2010 | SVM | Patch-level (Binary) | Distinguish whether a cropped image patch contains a pothole or a normal road surface. |
| Saar <i>et al.</i> [92] | | | 2010 | Multi-layer Perceptron | Patch-level (Binary) | Derive neural networks which can provide patch-level defect detection. The neural network is trained based on hand-crafted features. | |
| Yusof <i>et al.</i> [93] | | | 2019 | Multi-layer perceptron | Patch-level (Binary) | Similar to Saar <i>et al.</i> [92], deploy the defect detection method based on multi-layer perceptron. Hand-crafted features are used to train parameters. | |
| Unsupervised | | Oliveira <i>et al.</i> [94] | 2012 | Gaussian Mixture Model | Patch-level (Binary) | Model the probabilistic distribution of features extracted from normal road surfaces using multiple Gaussian distributions. Detect defects using a pre-defined threshold. | |
| | | Buza <i>et al.</i> [95] | 2013 | Clustering | Patch-level (Binary) | Leverage K-means clustering to compile a normal road surface model. Detect road defects by computing the distance between the input feature and centroids. | |
| | | Oliveira <i>et al.</i> [94] | 2012 | Kernel density estimation | Pixel-level (Binary) | Directly estimate a probabilistic model for the normal road surface using kernel density estimation. Detect road defects using a pre-defined threshold. | |
| Deep learning-based method | | Supervised | Zou <i>et al.</i> [25] | 2018 | U-Net | Pixel-level (Binary) | Pixel-level road defect segmentation model. Provide binary classification results representing likelihoods for normal road surfaces and defects. |
| | | | Konig <i>et al.</i> [24] | 2019 | U-Net | Pixel-level (Binary) | Utilise an attention mechanism and residual convolutional blocks to improve discriminative power for the defect segmentation model. |
| | | | Zhang <i>et al.</i> [28] | 2020 | GAN | Patch-level (Binary) | Propose a deep generative adversarial network for pavement crack detection, which is trained end-to-end with a crack-patch-only supervised method to overcome the local minimum problem. |
| | Chen <i>et al.</i> [19] | | 2017 | CNN | Pixel-level (Binary) | Present a CNN architecture for pixel-level defect detection based on video data. a Naive Bayes data fusion scheme is proposed to reduce false positives. | |
| | Dong <i>et al.</i> [96] | | 2023 | CNN | Pixel-level (Binary) | propose a pavement damage segmentation framework consisting of a Siamese network, an explicit crack segmentation network, and a feature fusion method. | |
| | Chen <i>et al.</i> [97] | | 2023 | CNN | Pixel-level (Binary) | Present a CNN architecture using an attention mechanism for multi-scale defect detection. The multi-scale attention method improves the robustness of defect detection methods for various image scales. | |
| | Wang <i>et al.</i> [98] | | 2023 | CNN | Pixel-level (Binary) | propose a relevance-aware and cross-reasoning network (RCN) for road defect segmentation, which can segment defects using merely non-defective images for training. | |
| | Fang <i>et al.</i> [21] | | 2020 | RCNN | Instant-level (Multi-classes) | Present instant-level road defect detection method. The proposed method is based on Faster-RCNN (Region-based Convolutional Neural Networks). Provide detection results identified by multiple defect classes. | |
| | Unsupervised | Mujeeb <i>et al.</i> [99] | 2019 | Autoencoder | Pixel-level (Binary) | Train an autoencoder using normal road surface images. Detect road defects by comparing a reconstruction error and a pre-defined threshold. | |
| | | Yu <i>et al.</i> [100] | 2020 | GAN | Pixel-level (Binary) | Build an adversarial network to derive a transformation model between a normal road surface and its corresponding frequency-domain image. Defects are detected by comparing the difference in the frequency domain. | |
| | | Yu <i>et al.</i> [101] | 2023 | U-Net | Pixel-level (Binary) | Design unsupervised defect detection using based on multi-source domain adaptation. Provide more stable and improved segmentation performance on an unlabelled dataset. | |
| | Semi-supervised | Li <i>et al.</i> [102] | 2020 | GAN | Pixel-level (Binary) | A semi-supervised method for pavement crack detection. An adversarial learning method and a full convolution discriminator are adopted, which can learn to distinguish the ground truth from segmentation predictions. | |
| Shim <i>et al.</i> [103] | | 2020 | GAN | Pixel-level (Binary) | Multiscale and adversarial learning-based semi-supervised semantic segmentation approach for crack detection in concrete surface. The multiscale segmentation neural network, discriminator neural network, and adversarial learning technique were used. | | |
| Ren <i>et al.</i> [104] | | 2023 | GAN | Pixel-level (Binary) | A pavement anomaly detection network (PAD Net) is proposed. PAD Net is a semi-supervised learning approach based on generative adversarial networks for identifying pixel-level anomalous image segments. | | |

that the training data consists entirely of normal road surfaces, and based on these assumptions, the model is trained. If the model’s output is abnormal, it is determined that there is a defect. These approaches’ quality highly depends on how sophisticated the probability model can be derived. Kernel density estimation has been used to derive a probabilistic model for features extracted from normal road surface images [106,107]. Alternatively, Oliveira *et al.* [94] propose an unsupervised road defect detection method based on the Gaussian mixture model. Buza *et al.* [95] leverage k-means clustering to derive a normal road surface model. While these methods can be useful in detecting defects without the need for explicit annotation, they may suffer from the overfitting problem when dealing with complex and diverse road surface images. Furthermore, these approaches have an inherent weakness as they assume that the feature distribution follows a Gaussian or linear distribution, which may not always hold true for actual data. This can lead to limited effectiveness and accuracy of the models when the actual data distribution significantly deviates from the assumed distribution. Therefore, it is important to consider the choice of probability models carefully and to validate their performance on real-world datasets with diverse feature distributions.

5.1.3 Deep learning-based methods

In this subsection, we will introduce deep learning-based road condition assessment from two perspectives: bounding box level detection and pixel level detection. It is worth noting that deep learning is a subset of machine learning that uses neural networks with multiple layers to learn representations of data. As such, deep learning methods are becoming increasingly popular in this field due to their ability to automatically learn features from data, reducing the need for hand-crafted feature engineering.

Supervised: Supervised methods using deep learning can be classified into two categories: pixel-level defect detection and objective-level detection. Pixel-level defect detection involves defect segmentation, where each pixel in the image is classified as a defect or non-defect. The encoder-decoder network is one of the most famous structures used for defect segmentation. For example, in [23], the U-Net [108], a classical encoder-decoder network, was used to segment cracks on the road.

Several recent studies have proposed novel deep-learning approaches for crack detection on the road surface. Zou *et al.* [25] proposed a novel encoder-decoder network that separates cracks from the background using convolutional feature learning. Similarly, Pauly *et al.* [109] leverage convolutional neural network to obtain pixel-wise classification results between normal road surfaces and road defects. Konig *et al.* [24] introduced a U-Net-based network with an attention mechanism and residual convolutional blocks that achieved state-of-the-art results in two distinct crack datasets. Ji *et al.* [26] utilised the DeepLabv3+ model [27] for automatic crack detection in asphalt pavement. In addition, Zhang *et al.* [28] proposed a deep generative adversarial network for pavement crack detection, which was trained end-to-end with a crack-patch-only super-

vised method to overcome the local minimum problem. Dong *et al.* [96] proposed a pavement damage segmentation framework consisting of a Siamese network and an explicit crack segmentation network. This framework has been extended by applying the feature fusion method [110]. Wang *et al.* [98] proposed a Relevance-aware and Cross-Reasoning network (RCN) for road defect segmentation, which can segment defects using merely non-defective images for training. The aforementioned methods do not consider the robustness of the image scale. To improve the robustness of image scale variation, Chen *et al.* [97] leveraged the multi-scale attention method, which can detect road defect performance in various image resolutions.

Defect detection at the bounding box level can be achieved through either patch scanning or end-to-end Convolutional Neural Network (CNN) approaches. The patch scanning-based method involves cropping the original high-resolution image into a set of low-resolution patches, identifying patches containing defects using a trained classifier, and applying filtering to remove false positives. In [19], a CNN architecture was employed to classify crack patches, and a registration procedure was proposed to maintain the spatiotemporal coherence of cracks in videos. A Naive Bayes data fusion scheme, which aggregates information from multiple frames, was proposed to effectively discard false positives. However, this approach requires a high density of patches, leading to increased computational load and challenging real-time applications.

Some other works use CNN detectors to predict the bounding box of cracks directly [20, 21, 29, 111]. Whether a Region Proposal Network (RPN) is used, these detectors can be divided into two families: one-stage and two-stage. A typical two-stage network is Faster-RCNN (Region-based Convolutional Neural Networks) [20], which uses an RPN to generate regions of interest in the first stage, and then performs object classification and bounding-box regression on the region proposals. Li *et al.* [29] used Faster R-CNN to detect six types of road defects under different lighting conditions. Similarly, Fang *et al.* [21] employed Faster-RCNN to identify the candidate crack regions, followed by a Bayesian integration algorithm to filter out false positives. Ma *et al.* [111] proposed a system composed of a Pavement Crack Generative Adversarial Network (PCGAN) and a crack detection and tracking network called YOLO-MF. In this method, First, The PCGAN is leveraged to generate virtual crack images. Next, YOLO-MF is developed to identify the regions of cracks.

Although two-stage networks can reduce some computational load, they are still not fast enough for real-time applications. Therefore, single-stage networks such as You Only Look Once (YOLO) [112] and Single Shot MultiBox Detector (SSD) [113] have been used to improve perception speed significantly. Single-stage networks treat object detection as a simple regression problem by taking an input image and learning the class probabilities and bounding box coordinates. YOLOv2 [114] and YOLOv3 [115] were used for crack detection of road surfaces in [22] and [116], respectively. These studies showed that cracks on structural surfaces can be detected and quantified with high accuracy in real-time when one-stage detectors are used.

Unsupervised and semi-supervised: Several unsupervised methods based

on autoencoders [117] have been proposed for road defect detection [99, 118]. These methods train models on normal road images and detect defects by checking whether the reconstruction error of an input sample exceeds a threshold. However, this reconstruction-based approach has a significant drawback: the model may generate plausible-looking samples that were not seen during training, making it difficult to distinguish whether a sample contains defects or not [119, 120]. To address this limitation, Yu *et al.* [100] proposed an adversarial network-based road defect detection method for road defect detection. Their model uses a generator network to synthesise defect samples fed to a discriminator network that distinguishes between real and synthesised defect samples. The generator is trained to produce defect samples that can fool the discriminator, while the discriminator is trained to improve its ability to distinguish between real and synthesised samples. This adversarial training process leads to a model that can detect road defects even without labelled training data. As similar to the above method, Yu *et al.* [101] presented an unsupervised defect segmentation method based on multi-source domain adaptation. Their method improves the segmentation performance on an unlabelled dataset using multiple labelled datasets.

Unsupervised methods offer the advantage of not relying on labelled data, which makes them appealing to real-world environments. However, their actual performance is noticeably inferior to that of supervised methods [100, 117, 120]. Specifically, methods based on the fine-tuning technique, where a model is re-trained using generated pseudo labels, exhibit varying performance depending on the accuracy of these labels [100, 101]. To address these limitations, semi-supervised methods have been proposed [102–104].

Li *et al.* [102] and Shim *et al.* [103] have applied adversarial learning in semi-supervised road defect detection. They employed adversarial learning to minimise domain shifts between a large-scale labelled dataset and a dataset with only a few labelled samples. This approach helps bridge the gap between labelled and unlabelled data to improve model performance. Another study by Ren *et al.* [104] utilises adversarial learning to derive a normal road surface model, enhancing the accuracy of road defect detection. By leveraging semi-supervised techniques and adversarial learning, these methods aim to improve the performance and robustness of road defect detection, especially when labelled data is limited or unavailable.

Remarks. Image-based methods excel in road pavement defect detection, offering numerous advantages. They are non-invasive, ensuring the safety of inspection personnel by eliminating physical contact with the pavement. This results in efficient, large-scale coverage, surpassing manual inspection speeds. The objective nature of image-based systems ensures consistent results, free from human error and subjective judgment. Additionally, these methods facilitate collecting and analysing historical data, which is invaluable in understanding pavement deterioration patterns and aiding future planning.

However, these methods have notable drawbacks: Defect detection accuracy depends heavily on image quality and is susceptible to poor lighting, shadows, and adverse weather conditions; They primarily detect surface defects, miss-

ing critical subsurface issues; Furthermore, privacy concerns arise with imaging technology, especially if images are captured beyond the road surface. These challenges emphasise the need for careful implementation of such systems.

In summary, the past decades have witnessed the rapid development of visual road defect detection, with traditional image processing methods and deep learning methods being the two main approaches. Deep learning-based methods, in particular, have gained more attention due to their better robustness and accuracy. Despite the progress, there are still limitations in image-based road defect detection, such as the susceptibility of visual methods to varying lighting conditions and the high computational load required to process high-resolution images. Therefore, there is a need to address these challenges and improve the performance of image-based road defect detection. Two possible directions for future research are (1) enhancing the real-time performance of the algorithms and (2) designing an invariant feature extractor that can adapt to different environments. By addressing these challenges, image-based road defect detection can be further improved and applied in real-world scenarios.

5.2 Point cloud-based methods

In recent years, various approaches have been proposed for depth-based methods, such as point clouds or depth maps, utilising explicit depth information. These methods can be classified into three groups: statistical, machine learning-based, and deep-learning-based. In this section, we will provide detailed descriptions of these methods. Table 5 gives a comprehensive overview of point cloud-based methods.

5.2.1 Statistical

In this section, we will discuss traditional methods for road defect detection using point clouds. A variety of point cloud processing methods have been used for this purpose. Furthermore, some of the hand-crafted features that have been successfully applied in visual crack detection can also be used in point cloud-based road defect detection.

In recent years, 3D sensors have enabled a plethora of crack detection studies using 3D point cloud data [30, 121, 135–137]. For instance, in [135], a planar triangulation modelling approach is employed to construct a triangular irregular network dataset and extract cracks based on the inverse distance weighting rasterisation method from wall points. Similarly, [136, 137] utilise the inverse distance weighting interpolation algorithm, the maximum entropy threshold algorithm, and tensor voting to identify crack curves. To enhance the accuracy of crack analysis and detect cracks from point cloud measurements, [121] employs the maximum gradient of signal-noise ratio distribution for Gaussian filtering. Bosurgi *et al.* [123] present a framework for detecting potholes automatically through the heuristic filtering of depth sensing results of pavement surfaces acquired using innovative high-performance depth sensors.

Table 5: List of point cloud-based methods.

| Cat | Subcat | Reference | Year | Based method | Output protocol | Details |
|-------------------|-----------------|-------------------------------|------|-------------------------|------------------------------|---|
| Signal processing | Statistical | Xu <i>et al.</i> [121] | 2019 | Gaussian filtering | Point-level (Binary) | Identify cracks from the point clouds measurement and optimise the accuracy of the crack analysis efficiently using Gaussian filtering |
| | | Yu <i>et al.</i> [30] | 2014 | Probabilistic modelling | Point-level (Binary) | Modelling the intensity information of point clouds and using the intensity model to identify pavement cracks and extract their skeletons. |
| | | Li <i>et al.</i> [122] | 2017 | Point clustering | Point-level (Binary) | laser-imaging techniques are employed to model the pavement surface with dense 3D points, and a sparse points grouping method is proposed to detect cracks from the 3D point clouds. |
| | | Bosturgi <i>et al.</i> [123] | 2023 | Thresholding | Point-level (Binary) | propose an approach for automated pothole detection through the processing of depth sensing results of pavement surfaces acquired using innovative high-performance equipment. |
| | | Du <i>et al.</i> [33] | 2020 | Plane fitting | Point-level (Binary) | 3D point clouds are first acquired by using binocular stereo vision, and then a pavement plane fitting method is used to eliminate the 3D point clouds of the road plane and roughly extract the pothole. |
| Machine learning | Supervised | Wu <i>et al.</i> [124] | 2016 | Clustering | Point-level (Binary) | Extract 3D crack skeletons from 3D point clouds acquired by a mobile Light Detection and Ranging (LiDAR) system. The algorithm leverages the clustering method to distinguish whether a point is a crack or a normal road surface. |
| | | Yan <i>et al.</i> [125] | 2017 | RANSAC | Point-level (Binary) | The proposed method assumes that the majority of scanned points from a road surface can be modelled well using RANSAC, and points extracted from road defects can be considered outliers by using the RANSAC algorithm. |
| | | Yu <i>et al.</i> [30] | 2014 | K-means clustering | Point-level (Binary) | Conduct binary clustering to identify 3D points extracted from normal road surfaces and defects. The clustering is conducted using depth value. |
| Deep learning | Supervised | Dizaji <i>et al.</i> [126] | 2019 | 3D InspectionNet | Voxel-level (Binary) | Learn 3D defect features from an artificially generated 3D dataset. Distinguish defect features extracted from the voxels. Voxels are defined from point clouds. |
| | | Zhang <i>et al.</i> [127] | 2017 | CrackNet | Point-level (Binary) | A deep learning-based method that detects cracks based on 3D asphalt surface data. Provide point-wise binary class scores for all individual points. This job is extended to CrackNet II [128] and CrackNet-V [129]. |
| | | Nasrollahi <i>et al.</i> [35] | 2019 | PointNet | Point-level (Binary) | Directly use point clouds as input without pre-processing, such as voxelisation or making a histogram. The proposed method provides defect detection on a concrete surface. |
| | | Feng <i>et al.</i> [130] | 2021 | Graph Convolutional | Point-level (Binary) | A graph-widen module has been proposed to construct a reasonable graph structure for point clouds on graph convolutional networks (GCN) for increasing the detection performance of road defects. |
| | | Li <i>et al.</i> [131] | 2023 | Mask R-CNN | Point-level (Binary) | A pipeline for accurately segmenting and extracting rural road surface objects in 3D LiDAR point-cloud is proposed. As well as a method to extract geometric parameters belonging to tar seal. |
| | | Li <i>et al.</i> [132] | 2021 | PointNet++ | Point-level (Multi-classes) | Propose PointNet++-base road surface defect classification method. The proposed method takes raw point-cloud without modification. The proposed method adaptively combines features from multiple scales. |
| | | Chen <i>et al.</i> [133] | 2023 | GCN, PointNet++ | Point-level (Binary-class) | A deep learning-based gradation extraction method using 3D point cloud data is proposed. A multi-feature fusion network was developed, which uses an extraction network as the backbone and additional auxiliary information. |
| | | Gui <i>et al.</i> [134] | 2023 | Deep Neural Network | Point-level (binary-classes) | Inspired by the heterogeneity of rapid-increasing 3D pavement data and the generalisation ability of transfer learning, a robust and generalised framework for cross-scene 3D pavement-crack detection and attribute extraction was proposed. |
| | Semi-supervised | Feng <i>et al.</i> [130] | 2021 | Graph Convolutional | Point-level (Binary) | A graph-widen module has been proposed to construct a reasonable graph structure for point clouds on graph convolutional networks (GCN) for increasing the detection performance of road defects. |

Several studies have proposed various methods to detect road defects using 3D point cloud data. In [30], the intensity information of point clouds was used to identify pavement cracks and extract their skeletons. Li *et al.* [122] have presented a clustering-based road defect detection method using 3D point captured by a 3D laser scanner. In [32], a novel method was proposed to identify and extract the cracks, which applied both dilation and the Canny algorithm to the point cloud data of terrestrial laser scanning measurement. In [31], a watershed method was proposed to extract road potholes. Plane fitting is also a widely used technique in point cloud-based road defect detection. For example, in [33], 3D point clouds were first acquired using binocular stereo vision, and then a pavement plane fitting method was used to eliminate the 3D point clouds of the road plane and roughly extract the pothole. Finally, k-means clustering and region-growing algorithms were adopted to extract the potholes. However, these methods still face challenges in extracting fine and low-connectivity cracks, and the automation of these methods needs to be improved.

5.2.2 Machine learning-based methods

Compared with the statistical methods and deep-learning-based methods, conventional machine learning-based methods for road surface defect detection are relatively less common. This is because the presence or absence of a defect can be determined through the distance information acquired by the laser scanner without needing a separate complex machine-learning algorithm. However, several machine learning-based algorithms have been proposed to develop a more accurate road defect detection model [30, 124, 125].

Wu *et al.* [124] proposed a novel algorithm for rapidly extracting 3D crack skeletons from 3D point clouds acquired by a mobile LiDAR system. The algorithm leverages a clustering method to distinguish between points belonging to cracks and those belonging to the normal road surface. Similarly, Yan *et al.* [125] use RANSAC to identify points corresponding to road defects. Their method assumes that most scanned points from a road surface can be well-modelled using RANSAC, and the points extracted from road defects can be considered outliers using the RANSAC algorithm. Yu *et al.* [30] also employs a clustering method to gather the 3D point set related to the normal road surface.

5.2.3 Deep learning-based methods

The initial studies on point cloud analysis, such as [138, 139], typically involved transforming point clouds into regular 3D voxel grids or image grids to enable kernel optimisations and weight sharing. More recently, in [126], a 3D InspectionNet was proposed to learn 3D defect features from an artificially generated 3D dataset. This approach can identify defect features from voxels with high accuracy, achieving a validation accuracy of 98.85%.

The effectiveness of deep learning in image processing has motivated researchers to apply it to 3D data processing. CrackNet [127], a deep learning-based method, was proposed to detect cracks in 3D asphalt surface data without

using pooling layers to downsample the original data. This approach outputs predicted class scores for individual pixels and has demonstrated high pixel-level accuracy. However, its learning capability is limited by the fixed and non-learnable feature generator. Moreover, the processing speed of CrackNet is slow due to many parameters and significant data depth at hidden layers. Several works have proposed improvements to CrackNet’s architecture to address these limitations. CrackNet II [128] abandons the feature generator and constructs a deeper architecture, while CrackNet-V [129] uses small filters to reduce the number of parameters and achieve efficient feature extraction. Additionally, CrackNet-V proposes a new activation unit for detecting shallow cracks. Both CrackNet II and CrackNet-V demonstrate better efficiency and accuracy than CrackNet. However, their improvements are limited and heavily dependent on annotated data. As CrackNet-like approaches are pixel-level techniques, they are still unable to fully exploit 3D information to determine the position of identified cracks accurately.

Conversely, representing the 3D data using the centre of voxels can lead to high memory usage and loss of orientation cues. To address this, some research has used point clouds directly as input, respecting the permutation invariance of points. Feng *et al.* [130] applied the Graph Convolutional Network (GCN) for classifying the points. They proposed a graph-widen module to construct a reasonable graph structure for point clouds on GCN to increase the detection performance of road defects from a different perspective. Li *et al.* [131] and Gui *et al.* [134] modified CNN to process point clouds. Li *et al.* built a pipeline for accurate segmentation and extraction of road surface objects in 3D LiDAR point clouds. Gui *et al.* created a framework for cross-scene 3D pavement-crack detection and attribute extraction. Their method was inspired by the heterogeneity of rapid-increasing 3D pavement data and the generalisation ability of transfer learning. PointNet [140] was the first to work on the raw point clouds directly, and has been successfully applied in automated defect detection in [35] and automated bridge component recognition [34].

Although effective for some tasks, PointNet is limited in capturing local structures and recognising fine-grained patterns due to its inability to account for the metric space the points inhabit. PointNet++ was introduced as an extension of PointNet [141] to overcome this limitation. This approach uses a nested partitioning of the input point cloud to learn deep point set features efficiently and robustly. By recursively applying PointNet on these partitions and adaptively combining features from multiple scales, PointNet++ can capture local structures in the point cloud, resulting in improved performance compared to PointNet. For instance, in a pavement disease classification task, PointNet++ outperformed PointNet [132]. Chen *et al.* [133] proposed a deep learning-based gradation extraction method using 3D point cloud data. A multi-feature fusion network was developed, which uses an extraction network as the backbone with additional auxiliary information.

Remarks. Point cloud-based methods for road defect detection offer several distinct advantages. These methods provide high-precision 3D data, enabling detailed and accurate mapping of road surfaces and defects. The depth infor-

mation captured in point clouds allows for a more comprehensive analysis of road conditions, going beyond what is visible on the surface. This depth aspect is particularly useful in identifying subtle or emerging issues that might not be apparent in 2D images. Additionally, using advanced sensors and algorithms, including machine learning and deep learning techniques, enhances the detection and classification of various road defects. These methods can automate the detection process, increasing efficiency and reducing the time and labour involved in manual inspections. Moreover, point cloud-based methods are adaptable to various environments and can work effectively under different lighting and weather conditions, overcoming limitations faced by traditional image-based methods in poor visibility.

However, challenges accompany these benefits. Acquiring and maintaining specialized equipment, such as 3D laser scanners, incurs a high cost. Additionally, processing large volumes of 3D data demands significant computational resources and storage. The resolution of point clouds may pose limitations in detecting very fine or low-connectivity cracks. Automation improvement is needed, especially in handling complex or ambiguous scenarios, where algorithms may struggle, potentially leading to identification or classification errors necessitating manual review or oversight.

Numerous point cloud-based road defect detection methods have been proposed in recent years, utilising both hand-crafted features and deep learning approaches, similar to visual defect perception. However, the current point cloud networks cannot be effectively applied to large scenes or large point clouds due to the high computational load required to extract the latent representation of the point cloud. Moreover, the scarcity of public road defect datasets, particularly point cloud data, hinders the development of research on point cloud-based road defect detection. Therefore, the collection of comprehensive 3D road defect benchmark datasets would be beneficial for future studies on road assessment.

5.3 RGB+D data fusion-based methods

As discussed in the above sections, image-based methods rely on visual cues such as texture, colour, or brightness to identify road defects. However, they may struggle to detect defects that are not clearly visible in the image, such as those covered by shadows or occluded by other objects. On the other hand, explicit depth information allows for identifying road defects that are not visible in the image alone. Various methods have been proposed to combine image and depth information. This survey paper categorises these methods into machine learning-based and deep-learning-based methods. This section presents detailed descriptions of these methods using image and depth information, listed in Table 6.

5.3.1 Machine learning-based methods

RGB images and depth maps have their own advantages and limitations when it comes to road defect detection. RGB images can be helpful for capturing in-

Table 6: List of data fusion-based methods.

| Cat | Subcat | Reference | Year | Based method | Output protocol | Details |
|------------------|--------------|-------------------------------|------|-----------------------------|----------------------------|--|
| Machine learning | Supervised | Salari <i>et al.</i> [142] | 2012 | SVM | Patch-level (Binary) | RGB pixel values and the depth information estimated by a stereo-matching algorithm are used as inputs for SVM. |
| | | Tsai <i>et al.</i> [143] | 2012 | Scoring method | Pixel-level (Binary) | A linear-buffered Hausdorff scoring method was used to quantitatively evaluate the crack segmentation performance by comparing the detected cracks with the manually established ground truth. |
| | | Moazzam <i>et al.</i> [144] | 2013 | Probabilistic modelling | Pixel-level (Binary) | Detect road potholes by analyzing road depth distribution estimated by using an estimated depth map calculated by a stereo matching algorithm. |
| | Unsupervised | Fan <i>et al.</i> [145] | 2021 | linear iterative clustering | Pixel-level (Binary) | Estimate a depth map using a stereo vision camera and apply the depth maps and RGB images to derive a model for a normal road surface. |
| | | Chen <i>et al.</i> [146] | 2021 | Otsu’s algorithm | Pixel-level (Binary) | Proposing an automatic crack-detection method that fuses 3D point clouds and 2D images based on an improved Otsu algorithm. |
| | | Kim <i>et al.</i> [147] | 2021 | RANSAC | Pixel-level (Binary) | A crack identification strategy is proposed, utilising a combination of RGB-D and high-resolution digital cameras to accurately measure cracks irrespective of the viewing angle. The camera system incorporates a custom sensor fusion algorithm dedicated to crack identification, enabling precise measurement resolution and robust depth estimation while addressing the challenge of skewed angles. |
| Deep learning | Supervised | Wang <i>et al.</i> [148] | 2019 | CNN+PointNet | Pixel-level (Binary) | Image data and point cloud are processed separately, and the processing results are applied to a novel dense fusion network to extract pixel-wise dense feature embedding. |
| | | Huang <i>et al.</i> [149] | 2014 | CNN+DNN | Pixel-level (Binary) | A method based on Dempster-Shafer (D-S) theory was proposed to combine the 2D grey-scale image and 3D laser scanning data as a mass function and generate fused detection results for each input image. |
| | | Wu <i>et al.</i> [150] | 2014 | CNN | Pixel-level (Binary) | A new method is proposed for detecting road potholes using mobile point cloud and images. The algorithm has three steps: extracting potential potholes from images using deep learning, extracting candidate potholes from the point cloud, and determining potholes based on depth analysis. Images of potholes and patch distress are used to train and test the deep learning system, as their texture features differ from those of a normal road. |
| | | Yang <i>et al.</i> [151] | 2020 | CNN+DNN | Patch-level (Binary) | RGB images were first used to in-paint per depth frame, and then a depth adaptive sliding-window size selection method was used to adjust the bounding box size. |
| | | Bolourian <i>et al.</i> [152] | 2023 | CNN,PointNet++ | Point-level (Binary-class) | developed a point cloud-based semantic segmentation DL method (SNEPointNet++) to detect different types of concrete surface defects. Use data fusion between the point cloud and images transferred from the point cloud data. |

formation about the texture and shape of defects on the road surface. However, depth maps may not be as effective in conveying precise information about surface texture. Nevertheless, depth maps are valuable in providing general defect detection performance by offering depth information, which can be critical in identifying defects.

Recent studies have proposed various machine learning methods that leverage the strengths and weaknesses of different data sources [31, 142, 144, 145]. Tsai *et al.* [31] applied a simple filtering technique on the colour histogram to find road defects on asphalt roads. Chen *et al.* [146] leveraged a thresholding algorithm to detect road cracks using point clouds and RGB images. Salari *et al.* [142] applied SVM to classify road defects in small-size image patches. Their approach utilised RGB pixel values and depth information obtained from stereo-matching algorithms as inputs to the SVM. Moazzam *et al.* [144] focused on detecting road potholes by analysing the depth distribution of roads, which was estimated by using stereo matching algorithms. Kim *et al.* [147] leveraged the RANSAC algorithm to identify road cracks based on depth information, and they used image information to assist in finding valid locations of the cracks.

Fan *et al.* [145] used a stereo vision camera to estimate depth maps and combined them to develop a model for a normal road surface. They applied a simple linear iterative clustering algorithm to derive the model.

5.3.2 Deep learning-based methods

The fusion of vision and depth data has emerged as a prominent topic in the field of computer vision and robotics. This approach has been employed in numerous applications, including 2D/3D object detection [153, 154], object segmentation [155], and object pose estimation [148, 156]. By leveraging both visual and depth information, these applications aim to improve the accuracy and robustness of their outputs, leading to more effective and reliable results.

In the past, most studies treated depth information as an image Grid and used the same type of feature learner for visual and depth data. However, this approach fails to fully leverage the potential of depth information, even when the depth information is represented with pseudo colours [153] or geometrically codes [155]. This limited use of depth data can result in suboptimal performance, particularly when dealing with complex scenes or objects with complex geometries.

Recent studies have highlighted the advantages of using heterogeneous network architectures to fuse visual and depth information to avoid the lossy input pre-processing such as projection or quantisation [148, 154]. For example, in [154], a CNN and PointNet were used independently to process the image data and raw point cloud data, respectively. The resulting image and point cloud representations were then fused to predict 3D boxes. Similarly, in [148], the image data and point cloud were also processed individually, and a novel dense fusion network was introduced to extract pixel-wise dense feature embedding. Furthermore, in identifying road defects, the synergy of visual and depth perception can notably boost detection accuracy. Attributes from visual

data, like colour, can expedite recognising potential defect-prone areas. On the other hand, depth data adds resilience to the detection process, is minimally influenced by factors like illumination and disturbances, and adeptly maps the authentic size and shape of defects. By harnessing the combined potential of CNN and PointNet, it is evident that diverse network structures can amplify the consistency and precision of fusion-centric techniques.

Despite the potential benefits of fusing vision and depth data for defect detection, there are only a few methods that have explored this approach in a straightforward manner. For example, Huang *et al.* [149] proposed a method that uses the Dempster-Shafer theory to combine 2D grayscale images and 3D laser scanning data as a mass function, and generates fused detection results at the decision-making level. Wu *et al.* [150] proposed a pothole detection method based on a deep learning-based feature ensemble for the two features extracted from the point cloud and RGB image.

In a more recent study, Yang *et al.* [151] used RGB images to paint per depth frame, and then applied a depth adaptive sliding-window size selection method to adjust the bounding box size. However, there is a growing interest in end-to-end heterogeneous network architectures, such as DenseFusion [148], which can offer a more effective and efficient means of integrating vision and depth data for defect detection. By leveraging the strengths of these advanced fusion-based methods, researchers can make significant progress towards developing more robust and accurate defect detection systems.

Remarks. RGB+D data fusion-based methods for road defect detection leverage the strengths of RGB imagery and depth data, providing a range of advantages. The primary benefit is the comprehensive analysis capability: RGB images provide detailed information about surface texture and colour, aiding in recognising defects like cracks or potholes. Depth data complements this by providing crucial three-dimensional information about the road surface, particularly useful for identifying defects that are not easily discernible in two-dimensional images. This synergy enhances the overall accuracy and robustness of defect detection, making the system less susceptible to issues like poor lighting or visual obstructions. Additionally, these methods can better understand defects' actual size and shape, an aspect critical for effective road maintenance. The fusion approach also adapts to various environmental conditions, offering consistent performance under different lighting and weather scenarios.

However, there are notable drawbacks to RGB+D data fusion methods in road defect detection. One of the primary challenges is the complexity and computational intensity of processing and fusing two distinct types of data, which requires advanced algorithms and significant processing power. This complexity may result in higher hardware and software costs and the requirement for specialised expertise in developing and maintaining these systems.

5.4 Evaluation metrics for Defect Detection Methods

In road defect detection, the selection of evaluation metrics depends on the methodologies to identify the road defects that are tailored to the two types

of classification tasks: one-class classification (*a.k.a.* anomaly detection) [94, 95, 101, 105–107, 157] and multi-class classification [21, 132]. The evaluation metrics for one-class classification measure methods’ performances to predict whether new data is normal or an anomaly, including unsupervised or semi-supervised approaches, which usually train their models with normal samples only (*i.e.*, one-class). On the other hand, multi-class evaluation metrics are applied to evaluate the performance of multi-class road defect detection [113, 115] or instance segmentation [158, 159].

For one-class classification tasks commonly seen in road defect detection based on unsupervised or semi-supervised methods, the focus predominantly distinguishes road defects from normal road surfaces. The metrics commonly used include the False Positive Rate (FPR), which indicates the rate at which normal instances are incorrectly classified as anomalies, and the True Positive Rate (TPR), which measures the accuracy in identifying actual anomalies [120, 160]. The Receiver Operating Characteristic (ROC) Curve and the Area Under the Curve (AUC) are also crucial in these scenarios [101, 105, 106, 157], as they provide insights into the performance across various threshold settings. The Precision-Recall Curve gains importance in datasets with a severe imbalance between normal and anomalous instances [95, 107].

In the case of multi-class road defect detection and classification, considering instance segmentation, metrics like Precision, Recall (or Sensitivity), and the F1-Score become more significant [21, 132], because the metrics should provide the performance evaluation for detection results on multiple road defect classes. Precision measures the proportion of true positive detections among all positive detections for each defect class, ensuring the accuracy of defect identification. Recall assesses how many defects of a particular type are correctly identified, which is crucial for not missing any critical defects. The F1-Score balances precision and recall, which is especially valuable when the dataset is imbalanced. In addition, the Mean Absolute Error (MAE) and Intersection over Union (IoU) are employed in scenarios where the exact localisation of each defect type is significant [21]. Accuracy, while a common metric, can be less informative in multi-class scenarios with imbalanced classes [132].

These metrics, tailored to the specific nature of the classification task, are fundamental in evaluating and enhancing the performance of road defect detection systems. They provide a comprehensive understanding of how well these systems can distinguish between different classes of defects or identify anomalies in various operational conditions.

6 Conclusion

This survey paper comprehensively explores the current landscape in road defect detection, from traditional image-based methods to the emerging realms of point cloud and data fusion techniques. While image-based methods have seen significant advancements and remain a primary focus, our review highlights a burgeoning interest in point cloud-based methods and data fusion approaches.

These emerging methodologies hold immense potential for revolutionising road defect detection, although they still need more exploration.

The point cloud and data fusion-based methods open a new avenue for future research. Point cloud data has different geometrical properties than image domains, and it may not be suitable to process the data using the existing image-based methods. Data fusion between the two different domains is also quite a complicated issue. Hence, developing more sophisticated algorithms capable of processing vast and complex datasets will be crucial in advancing these technologies.

Applying these advanced methods in road defect detection can significantly contribute to smarter public transportation infrastructure management. With the integration of IoT and smart city initiatives, these technologies can provide real-time, accurate road condition data, facilitating proactive maintenance and enhancing public safety. Moreover, the environmental impact of road maintenance can be optimised by employing these precise detection methods, aligning with sustainability goals.

In conclusion, while the field has made remarkable strides in image-based road defect detection, the untapped potential of point cloud and data fusion methods presents exciting prospects. These technologies are not just evolutionary steps in road assessment but could be pivotal in shaping future intelligent transportation systems and sustainable infrastructure management. The road ahead, paved with these advanced technologies, beckons further exploration and innovation.

References

- [1] Department of Transport (United Kingdom), “Road safety statistics,” May 25, 2022.
- [2] R. S. Lim, H. M. La, Z. Shan, and W. Sheng, “Developing a crack inspection robot for bridge maintenance,” in *2011 IEEE International Conference on Robotics and Automation*, pp. 6288–6293, IEEE, 2011.
- [3] P. Prasanna, K. J. Dana, N. Gucunski, B. B. Basily, H. M. La, R. S. Lim, and H. Parvardeh, “Automated crack detection on concrete bridges,” *IEEE Transactions on automation science and engineering*, vol. 13, no. 2, pp. 591–599, 2014.
- [4] I. Al-Saadi, H. Wang, X. Chen, P. Lu, and A. Jasim, “Multi-objective optimization of pavement preservation strategy considering agency cost and environmental impact,” *International Journal of Sustainable Transportation*, vol. 15, no. 11, pp. 826–836, 2021.
- [5] H. Wang, I. Al-Saadi, P. Lu, and A. Jasim, “Quantifying greenhouse gas emission of asphalt pavement preservation at construction and use stages using life-cycle assessment,” *International Journal of Sustainable Transportation*, vol. 14, no. 1, pp. 25–34, 2020.

- [6] S. Chambon and J.-M. Moliard, "Automatic road pavement assessment with image processing: review and comparison," *International Journal of Geophysics*, vol. 2011, 2011.
- [7] H. Zakeri, F. M. Nejad, and A. Fahimifar, "Image based techniques for crack detection, classification and quantification in asphalt pavement: a review," *Archives of Computational Methods in Engineering*, vol. 24, no. 4, pp. 935–977, 2017.
- [8] P. Scholar, "Review and analysis of crack detection and classification techniques based on crack types," *International Journal of Applied Engineering Research*, vol. 13, no. 8, pp. 6056–6062, 2018.
- [9] W. Cao, Q. Liu, and Z. He, "Review of pavement defect detection methods," *IEEE Access*, vol. 8, pp. 14531–14544, 2020.
- [10] Y.-A. Hsieh and Y. J. Tsai, "Machine learning for crack detection: review and model performance comparison," *Journal of Computing in Civil Engineering*, vol. 34, no. 5, p. 04020038, 2020.
- [11] H. Peel, S. Luo, A. Cohn, and R. Fuentes, "Localisation of a mobile robot for bridge bearing inspection," *Automation in Construction*, vol. 94, pp. 244–256, 2018.
- [12] E. Salari and G. Bao, "Pavement distress detection and severity analysis," in *Image Processing: Machine Vision Applications IV*, vol. 7877, p. 78770C, International Society for Optics and Photonics, 2011.
- [13] E. Salari and X. Yu, "Pavement distress detection and classification using a genetic algorithm," in *2011 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)*, pp. 1–5, IEEE, 2011.
- [14] L. Peng, W. Chao, L. Shuangmiao, and F. Baocai, "Research on crack detection method of airport runway based on twice-threshold segmentation," in *2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control (IMCCC)*, pp. 1716–1720, IEEE, 2015.
- [15] S. Wang and W. Tang, "Pavement crack segmentation algorithm based on local optimal threshold of cracks density distribution," in *International Conference on Intelligent Computing*, pp. 298–302, Springer, 2011.
- [16] H. Zhao, G. Qin, and X. Wang, "Improvement of canny algorithm based on pavement edge detection," in *2010 3rd International Congress on Image and Signal Processing*, vol. 2, pp. 964–967, IEEE, 2010.
- [17] A. Ayenu-Prah and N. Attoh-Okine, "Evaluating pavement cracks with bidimensional empirical mode decomposition," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp. 1–7, 2008.

- [18] E. Douka, S. Loutridis, and A. Trochidis, “Crack identification in beams using wavelet analysis,” *International Journal of solids and structures*, vol. 40, no. 13-14, pp. 3557–3569, 2003.
- [19] F.-C. Chen and M. R. Jahanshahi, “Nb-cnn: Deep learning-based crack detection using convolutional neural network and naïve bayes data fusion,” *IEEE Transactions on Industrial Electronics*, vol. 65, no. 5, pp. 4392–4400, 2017.
- [20] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *Advances in neural information processing systems*, vol. 28, pp. 91–99, 2015.
- [21] F. Fang, L. Li, Y. Gu, H. Zhu, and J.-H. Lim, “A novel hybrid approach for crack detection,” *Pattern Recognition*, p. 107474, 2020.
- [22] V. Mandal, L. Uong, and Y. Adu-Gyamfi, “Automated road crack detection using deep convolutional neural networks,” in *2018 IEEE International Conference on Big Data (Big Data)*, pp. 5212–5215, IEEE, 2018.
- [23] M. D. Jenkins, T. A. Carr, M. I. Iglesias, T. Buggy, and G. Morison, “A deep convolutional neural network for semantic pixel-wise segmentation of road and pavement surface cracks,” in *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 2120–2124, IEEE, 2018.
- [24] J. König, M. D. Jenkins, P. Barrie, M. Mannion, and G. Morison, “A convolutional neural network for pavement surface crack segmentation using residual connections and attention gating,” in *2019 IEEE International Conference on Image Processing (ICIP)*, pp. 1460–1464, IEEE, 2019.
- [25] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, and S. Wang, “Deepcrack: Learning hierarchical convolutional features for crack detection,” *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1498–1512, 2018.
- [26] A. Ji, X. Xue, Y. Wang, X. Luo, and W. Xue, “An integrated approach to automatic pixel-level crack detection and quantification of asphalt pavement,” *Automation in Construction*, vol. 114, p. 103176, 2020.
- [27] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-decoder with atrous separable convolution for semantic image segmentation,” in *Proceedings of the European conference on computer vision (ECCV)*, pp. 801–818, 2018.
- [28] K. Zhang, Y. Zhang, and H.-D. Cheng, “Crackgan: Pavement crack detection using partially accurate ground truths based on generative adversarial learning,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 2, pp. 1306–1319, 2020.

- [29] Y. Liu, J. Yao, X. Lu, R. Xie, and L. Li, “Deepcrack: A deep hierarchical feature learning architecture for crack segmentation,” *Neurocomputing*, vol. 338, pp. 139–153, 2019.
- [30] Y. Yu, J. Li, H. Guan, and C. Wang, “3d crack skeleton extraction from mobile lidar point clouds,” in *2014 IEEE Geoscience and Remote Sensing Symposium*, pp. 914–917, IEEE, 2014.
- [31] Y.-C. Tsai and A. Chatterjee, “Pothole detection and classification using 3d technology and watershed method,” *Journal of Computing in Civil Engineering*, vol. 32, no. 2, p. 04017078, 2018.
- [32] H. Yang and X. Xu, “Intelligent crack extraction based on terrestrial laser scanning measurement,” *Measurement and Control*, vol. 53, no. 3-4, pp. 416–426, 2020.
- [33] Y. Du, Z. Zhou, Q. Wu, H. Huang, M. Xu, J. Cao, and G. Hu, “A pothole detection method based on 3d point cloud segmentation,” in *Twelfth International Conference on Digital Image Processing (ICDIP 2020)*, vol. 11519, p. 1151909, International Society for Optics and Photonics, 2020.
- [34] H. Kim, J. Yoon, and S.-H. Sim, “Automated bridge component recognition from point clouds using deep learning,” *Structural Control and Health Monitoring*, vol. 27, no. 9, p. e2591, 2020.
- [35] M. Nasrollahi, N. Bolourian, and A. Hammad, “Concrete surface defect detection using deep neural network based on lidar scanning,” in *Proceedings of the CSCE Annual Conference, Laval, Greater Montreal, QC, Canada*, pp. 12–15, 2019.
- [36] H. Afrin, “A review on different types soil stabilization techniques,” *International Journal of Transportation Engineering and Technology*, vol. 3, no. 2, pp. 19–24, 2017.
- [37] K. Skorseth, *Gravel roads: maintenance and design manual*. US Department of Transportation, Federal Highway Administration, 2000.
- [38] K. Rajkamal, T. D. Reddy, D. Rohith, V. Chowdary, and C. Prasad, “Performance evaluation of gravel road sections sealed with surface dressing,” *Transportation Research Procedia*, vol. 17, pp. 81–89, 2016.
- [39] Y. Pan, X. Zhang, G. Cervone, and L. Yang, “Detection of asphalt pavement potholes and cracks based on the unmanned aerial vehicle multi-spectral imagery,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 10, pp. 3701–3712, 2018.
- [40] M. Y. Shahin, *Pavement management for airports, roads, and parking lots*, vol. 501. Springer, 2005.

- [41] H. Statistics, “Federal highway administration (fhwa),” *Retrieved from*, 2016.
- [42] J. Wang, “Analysis the control techniques of reflection crack on asphalt overlay of old cement concrete pavement,” in *Applied Mechanics and Materials*, vol. 204, pp. 1945–1948, Trans Tech Publ, 2012.
- [43] A. A. Association *et al.*, “Pothole damage costs drivers \$3 billion annually nationwide,” *Retrieved February*, vol. 17, p. 2016, 2016.
- [44] H. Song, K. Baek, and Y. Byun, “Pothole detection using machine learning,” *Advanced Science and Technology*, pp. 151–155, 2018.
- [45] L. Zhang, F. Yang, Y. D. Zhang, and Y. J. Zhu, “Road crack detection using deep convolutional neural network,” in *2016 IEEE international conference on image processing (ICIP)*, pp. 3708–3712, IEEE, 2016.
- [46] Y. Shi, L. Cui, Z. Qi, F. Meng, and Z. Chen, “Automatic road crack detection using random structured forests,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 12, pp. 3434–3445, 2016.
- [47] M. Eisenbach, R. Stricker, D. Seichter, K. Amende, K. Debes, M. Sesselmann, D. Ebersbach, U. Stoeckert, and H.-M. Gross, “How to get pavement distress detection ready for deep learning? a systematic approach,” in *2017 international joint conference on neural networks (IJCNN)*, pp. 2039–2047, IEEE, 2017.
- [48] Q. Zou, Y. Cao, Q. Li, Q. Mao, and S. Wang, “Cracktree: Automatic crack detection from pavement images,” *Pattern Recognition Letters*, vol. 33, no. 3, pp. 227–238, 2012.
- [49] J. König, M. D. Jenkins, M. Mannion, P. Barrie, and G. Morison, “Optimized deep encoder-decoder methods for crack segmentation,” *Digital Signal Processing*, vol. 108, p. 102907, 2021.
- [50] Q. Mei and M. Gül, “A cost effective solution for pavement crack inspection using cameras and deep neural networks,” *Construction and Building Materials*, vol. 256, p. 119397, 2020.
- [51] G. X. Hu, B. L. Hu, Z. Yang, L. Huang, and P. Li, “Pavement crack detection method based on deep learning models,” *Wireless Communications and Mobile Computing*, vol. 2021, 2021.
- [52] N. Yusof, A. Ibrahim, M. Noor, N. Tahir, N. Yusof, N. Abidin, and M. Osman, “Deep convolution neural network for crack detection on asphalt pavement,” in *Journal of Physics: Conference Series*, vol. 1349, p. 012020, IOP Publishing, 2019.

- [53] R. Fan, Y. Liu, X. Yang, M. J. Bocus, N. Dahnoun, and S. Tancock, "Real-time stereo vision for road surface 3-d reconstruction," in *2018 IEEE International Conference on Imaging Systems and Techniques (IST)*, pp. 1–6, IEEE, 2018.
- [54] C. Zhu, Q. Bi, W. Li, R. Yang, and B. Li, "Measurement and construction of road surface elevation information based on kriging model," in *Journal of Physics: Conference Series*, vol. 1486, p. 072076, IOP Publishing, 2020.
- [55] Y. Li, C. Papachristou, and D. Weyer, "Road pothole detection system based on stereo vision," in *NAECON 2018-IEEE National Aerospace and Electronics Conference*, pp. 292–297, IEEE, 2018.
- [56] Z. Zhang, X. Ai, C. Chan, and N. Dahnoun, "An efficient algorithm for pothole detection using stereo vision," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 564–568, IEEE, 2014.
- [57] E. M. Thompson, A. Ranieri, S. Biasotti, M. Chicchon, I. Sipiran, M.-K. Pham, T.-L. Nguyen-Ho, H.-D. Nguyen, and M.-T. Tran, "Shrec 2022: Pothole and crack detection in the road pavement using images and rgb-d data," *Computers & Graphics*, vol. 107, pp. 161–171, 2022.
- [58] R. Fan, U. Ozgunalp, B. Hosking, M. Liu, and I. Pitas, "Pothole detection based on disparity transformation and road surface modeling," *IEEE Transactions on Image Processing*, vol. 29, pp. 897–908, 2019.
- [59] H. Sarbolandi, D. Lefloch, and A. Kolb, "Kinect range sensing: Structured-light versus time-of-flight kinect," *Computer vision and image understanding*, vol. 139, pp. 1–20, 2015.
- [60] M. R. Jahanshahi, F. Jazizadeh, S. F. Masri, and B. Becerik-Gerber, "Unsupervised approach for autonomous pavement-defect detection and quantification using an inexpensive depth sensor," *Journal of Computing in Civil Engineering*, vol. 27, no. 6, pp. 743–754, 2013.
- [61] P. Wang, Y. Hu, Y. Dai, and M. Tian, "Asphalt pavement pothole detection and segmentation based on wavelet energy field," *Mathematical Problems in Engineering*, vol. 2017, 2017.
- [62] J. Gu, Y. Zhang, Y. Yin, R. Wang, J. Deng, and B. Zhang, "Surface defect detection of cabbage based on curvature features of 3d point cloud," *Frontiers in Plant Science*, vol. 13, p. 942040, 2022.
- [63] J. M. Kumar, R. Mahajan, D. Prabhu, and D. Ghose, "Cost effective road accident prevention system," in *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, pp. 353–357, IEEE, 2016.

- [64] L. Chen, X. He, X. Zhao, H. Li, Y. Huang, B. Zhou, W. Chen, Y. Li, C. Wen, and C. Wang, “Gocomfort: Comfortable navigation for autonomous vehicles leveraging high-precision road damage crowdsensing,” *IEEE Transactions on Mobile Computing*, 2022.
- [65] B. Zhang and X. Liu, “Intelligent pavement damage monitoring research in china,” *IEEE Access*, vol. 7, pp. 45891–45897, 2019.
- [66] Z. Huang, W. Chen, A. Al-Tabbaa, and I. Brilakis, “Nha12d: A new pavement crack dataset and a comparison study of crack detection algorithms,” *arXiv preprint arXiv:2205.01198*, 2022.
- [67] Q. Mei, M. Gül, and M. R. Azim, “Densely connected deep neural network considering connectivity of pixels for automatic crack detection,” *Automation in Construction*, vol. 110, p. 103018, 2020.
- [68] W. Tang, S. Huang, Q. Zhao, R. Li, and L. Huangfu, “An iteratively optimized patch label inference network for automatic pavement distress detection,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 8652–8661, 2021.
- [69] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, “Road damage detection and classification using deep neural networks with smartphone images,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 33, no. 12, pp. 1127–1141, 2018.
- [70] H. Maeda, T. Kashiyama, Y. Sekimoto, T. Seto, and H. Omata, “Generative adversarial network for road damage detection,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 36, no. 1, pp. 47–60, 2021.
- [71] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, H. Omata, T. Kashiyama, and Y. Sekimoto, “Global road damage detection: State-of-the-art solutions,” in *2020 IEEE International Conference on Big Data (Big Data)*, pp. 5533–5539, IEEE, 2020.
- [72] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, and Y. Sekimoto, “Rdd2022: A multi-national image dataset for automatic road damage detection,” *arXiv preprint arXiv:2209.08538*, 2022.
- [73] R. Fan, X. Ai, and N. Dahnoun, “Road surface 3d reconstruction based on dense subpixel disparity map estimation,” *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 3025–3035, 2018.
- [74] F. Timm and E. Barth, “Non-parametric texture defect detection using weibull features,” in *Image Processing: Machine Vision Applications IV*, vol. 7877, pp. 150–161, SPIE, 2011.
- [75] A. Kumar, “Computer-vision-based fabric defect detection: A survey,” *IEEE transactions on industrial electronics*, vol. 55, no. 1, pp. 348–363, 2008.

- [76] S. Youkachen, M. Ruchanurucks, T. Phatrapomnant, and H. Kaneko, “Defect segmentation of hot-rolled steel strip surface by using convolutional auto-encoder and conventional image processing,” in *2019 10th International Conference of Information and Communication Technology for Embedded Systems (IC-ICTES)*, pp. 1–5, IEEE, 2019.
- [77] D. Djukic and S. Spuzic, “Statistical discriminator of surface defects on hot rolled steel,” *Image Vis. Comput*, pp. 158–163, 2007.
- [78] T. Ojala, M. Pietikäinen, and D. Harwood, “A comparative study of texture measures with classification based on featured distributions,” *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [79] M. Chu, R. Gong, S. Gao, and J. Zhao, “Steel surface defects recognition based on multi-type statistical features and enhanced twin support vector machine,” *Chemometrics and Intelligent Laboratory Systems*, vol. 171, pp. 140–150, 2017.
- [80] Y. Liu, K. Xu, and D. Wang, “Online surface defect identification of cold rolled strips based on local binary pattern and extreme learning machine,” *Metals*, vol. 8, no. 3, p. 197, 2018.
- [81] J. Zhao, Y. Peng, and Y. Yan, “Steel surface defect classification based on discriminant manifold regularized local descriptor,” *IEEE Access*, vol. 6, pp. 71719–71731, 2018.
- [82] J. Wang, Q. Li, J. Gan, H. Yu, and X. Yang, “Surface defect detection via entity sparsity pursuit with intrinsic priors,” *IEEE Transactions on Industrial Informatics*, vol. 16, no. 1, pp. 141–150, 2019.
- [83] S.-S. Yang, Y.-H. He, Z.-L. Wang, and W.-S. Zhao, “A method of steel strip image segmentation based on local gray information,” in *2008 IEEE International Conference on Industrial Technology*, pp. 1–4, IEEE, 2008.
- [84] J. Choi and C. Kim, “Unsupervised detection of surface defects: A two-step approach,” in *2012 19th IEEE International Conference on Image Processing*, pp. 1037–1040, IEEE, 2012.
- [85] Y.-h. Ai and X. Ke, “Surface detection of continuous casting slabs based on curvelet transform and kernel locality preserving projections,” *Journal of Iron and Steel Research, International*, vol. 20, no. 5, pp. 80–86, 2013.
- [86] D. Aiger and H. Talbot, “The phase only transform for unsupervised surface defect detection,” in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 295–302, IEEE, 2010.
- [87] S. Xie, S. Shan, X. Chen, and J. Chen, “Fusing local patterns of gabor magnitude and phase for face recognition,” *IEEE transactions on image processing*, vol. 19, no. 5, pp. 1349–1361, 2010.

- [88] J. Kim, S. Um, and D. Min, “Fast 2d complex gabor filter with kernel decomposition,” *IEEE Transactions on Image Processing*, vol. 27, no. 4, pp. 1713–1722, 2017.
- [89] A. Kumar, “Automated defect detection in textured materials,” *Dept. Elect Electr. Eng. The Univ. Hong Kong, Hong Kong*, 2001.
- [90] W. Liu and Y. Yan, “Automated surface defect detection for cold-rolled steel strip based on wavelet anisotropic diffusion method,” *International Journal of Industrial and Systems Engineering*, vol. 17, no. 2, pp. 224–239, 2014.
- [91] J. Lin and Y. Liu, “Potholes detection based on svm in the pavement distress image,” in *2010 ninth international symposium on distributed computing and applications to business, engineering and science*, pp. 544–547, IEEE, 2010.
- [92] T. Saar and O. Talvik, “Automatic asphalt pavement crack detection and classification using neural networks,” in *2010 12th Biennial Baltic Electronics Conference*, pp. 345–348, IEEE, 2010.
- [93] N. Yusof, M. K. Osman, Z. Hussain, M. H. M. Noor, A. Ibrahim, N. Tahir, and N. Abidin, “Automated asphalt pavement crack detection and classification using deep convolution neural network,” in *2019 9th IEEE International Conference on Control System, Computing and Engineering (ICCSCCE)*, pp. 215–220, IEEE, 2019.
- [94] H. Oliveira and P. L. Correia, “Automatic road crack detection and characterization,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 14, no. 1, pp. 155–168, 2012.
- [95] E. Buza, S. Omanovic, and A. Huseinovic, “Pothole detection with image processing and spectral clustering,” in *Proceedings of the 2nd International Conference on Information Technology and Computer Networks*, vol. 810, p. 4853, 2013.
- [96] J. Dong, N. Wang, H. Fang, R. Wu, C. Zheng, D. Ma, and H. Hu, “Automatic damage segmentation in pavement videos by fusing similar feature extraction siamese network (sfe-snet) and pavement damage segmentation capsule network (pds-capsnet),” *Automation in Construction*, vol. 143, p. 104537, 2022.
- [97] J. Chen, Y. Wen, Y. A. Nanekaran, D. Zhang, and A. Zeb, “Multi-scale attention networks for pavement defect detection,” *IEEE Transactions on Instrumentation and Measurement*, 2023.
- [98] Y. Wang, M. Niu, K. Song, P. Jiang, and Y. Yan, “Normal-knowledge-based pavement defect segmentation using relevance-aware and cross-reasoning mechanisms,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 4, pp. 4413–4427, 2023.

- [99] A. Mujeeb, W. Dai, M. Erdt, and A. Sourin, “One class based feature learning approach for defect detection using deep autoencoders,” *Advanced Engineering Informatics*, vol. 42, p. 100933, 2019.
- [100] J. Yu, D. Y. Kim, Y. Lee, and M. Jeon, “Unsupervised pixel-level road defect detection via adversarial image-to-frequency transform,” in *2020 IEEE Intelligent Vehicles Symposium (IV)*, pp. 1708–1713, IEEE, 2020.
- [101] J. Yu, H. Oh, S. Fichera, P. Paoletti, and S. Luo, “Multi-source domain adaptation for unsupervised road defect segmentation,” in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023.
- [102] G. Li, J. Wan, S. He, Q. Liu, and B. Ma, “Semi-supervised semantic segmentation using adversarial learning for pavement crack detection,” *IEEE Access*, vol. 8, pp. 51446–51459, 2020.
- [103] S. Shim, J. Kim, G.-C. Cho, and S.-W. Lee, “Multiscale and adversarial learning-based semi-supervised semantic segmentation approach for crack detection in concrete structures,” *IEEE Access*, vol. 8, pp. 170939–170950, 2020.
- [104] R. Ren, P. Shi, P. Jia, and X. Xu, “A semi-supervised learning approach for pixel-level pavement anomaly detection,” *IEEE Transactions on Intelligent Transportation Systems*, 2023.
- [105] I. Abdel-Qader, S. Pashaie-Rad, O. Abudayyeh, and S. Yehia, “Pca-based algorithm for unsupervised bridge crack detection,” *Advances in Engineering Software*, vol. 37, no. 12, pp. 771–778, 2006.
- [106] H. Oliveira, J. J. Caeiro, and P. L. Correia, “Improved road crack detection based on one-class parzen density estimation and entropy reduction,” in *2010 IEEE International Conference on Image Processing*, pp. 2201–2204, IEEE, 2010.
- [107] H. Oliveira and P. L. Correia, “Automatic road crack segmentation using entropy and image dynamic thresholding,” in *2009 17th European Signal Processing Conference*, pp. 622–626, IEEE, 2009.
- [108] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*, pp. 234–241, Springer, 2015.
- [109] L. Pauly, D. Hogg, R. Fuentes, and H. Peel, “Deeper networks for pavement crack detection,” in *Proceedings of the 34th ISARC*, pp. 479–485, IAARC, 2017.
- [110] J. Dong, N. Wang, H. Fang, Q. Hu, C. Zhang, B. Ma, D. Ma, and H. Hu, “Innovative method for pavement multiple damages segmentation and measurement by the road-seg-capsnet of feature fusion,” *Construction and Building Materials*, vol. 324, p. 126719, 2022.

- [111] D. Ma, H. Fang, N. Wang, C. Zhang, J. Dong, and H. Hu, “Automatic detection and counting system for pavement cracks based on pegan and yolo-mf,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 22166–22178, 2022.
- [112] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 779–788, 2016.
- [113] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, “Ssd: Single shot multibox detector,” in *European conference on computer vision*, pp. 21–37, Springer, 2016.
- [114] J. Redmon and A. Farhadi, “Yolo9000: better, faster, stronger,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7263–7271, 2017.
- [115] J. Redmon and A. Farhadi, “Yolov3: An incremental improvement,” *arXiv preprint arXiv:1804.02767*, 2018.
- [116] S. E. Park, S.-H. Eem, and H. Jeon, “Concrete crack detection and quantification using deep learning and structured light,” *Construction and Building Materials*, vol. 252, p. 119096, 2020.
- [117] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, “Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion,” *Journal of machine learning research*, vol. 11, no. Dec, pp. 3371–3408, 2010.
- [118] G. Kang, S. Gao, L. Yu, and D. Zhang, “Deep architecture for high-speed railway insulator surface defect detection: Denoising autoencoder with multitask learning,” *IEEE Transactions on Instrumentation and Measurement*, 2018.
- [119] P. Perera, R. Nallapati, and B. Xiang, “Ocgan: One-class novelty detection using gans with constrained latent representations,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2898–2906, 2019.
- [120] S. Pidhorskyi, R. Almhosen, and G. Doretto, “Generative probabilistic novelty detection with adversarial autoencoders,” in *Advances in Neural Information Processing Systems*, pp. 6822–6833, 2018.
- [121] X. Xu and H. Yang, “Intelligent crack extraction and analysis for tunnel structures with terrestrial laser scanning measurement,” *Advances in Mechanical Engineering*, vol. 11, no. 9, p. 1687814019872650, 2019.
- [122] Q. Li, D. Zhang, Q. Zou, and H. Lin, “3d laser imaging and sparse points grouping for pavement crack detection,” in *2017 25th European Signal Processing Conference (EUSIPCO)*, pp. 2036–2040, IEEE, 2017.

- [123] G. Bosurgi, M. Modica, O. Pellegrino, and G. Sollazzo, “An automatic pothole detection algorithm using pavement 3d data,” *International Journal of Pavement Engineering*, vol. 24, no. 2, p. 2057978, 2023.
- [124] F. Wu, C. Wen, Y. Guo, J. Wang, Y. Yu, C. Wang, and J. Li, “Rapid localization and extraction of street light poles in mobile lidar point clouds: A supervoxel-based approach,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 2, pp. 292–305, 2016.
- [125] L. Yan, Z. Li, H. Liu, J. Tan, S. Zhao, and C. Chen, “Detection and classification of pole-like road objects from mobile lidar data in motorway environment,” *Optics & Laser Technology*, vol. 97, pp. 272–283, 2017.
- [126] M. S. Dizaji and D. K. Harris, “3d inspectionnet: a deep 3d convolutional neural networks based approach for 3d defect detection on concrete columns,” in *Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XIII*, vol. 10971, p. 109710E, International Society for Optics and Photonics, 2019.
- [127] A. Zhang, K. C. Wang, B. Li, E. Yang, X. Dai, Y. Peng, Y. Fei, Y. Liu, J. Q. Li, and C. Chen, “Automated pixel-level pavement crack detection on 3d asphalt surfaces using a deep-learning network,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 32, no. 10, pp. 805–819, 2017.
- [128] A. Zhang, K. C. Wang, Y. Fei, Y. Liu, S. Tao, C. Chen, J. Q. Li, and B. Li, “Deep learning-based fully automated pavement crack detection on 3d asphalt surfaces with an improved cracknet,” *Journal of Computing in Civil Engineering*, vol. 32, no. 5, p. 04018041, 2018.
- [129] Y. Fei, K. C. Wang, A. Zhang, C. Chen, J. Q. Li, Y. Liu, G. Yang, and B. Li, “Pixel-level cracking detection on 3d asphalt pavement images through deep-learning-based cracknet-v,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 1, pp. 273–284, 2019.
- [130] H. Feng, W. Li, Z. Luo, Y. Chen, S. N. Fatholahi, M. Cheng, C. Wang, J. M. Junior, and J. Li, “Gcn-based pavement crack detection using mobile lidar point clouds,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 11052–11061, 2021.
- [131] H. T. Li, Z. Todd, N. Bielski, and F. Carroll, “3d lidar point-cloud projection operator and transfer machine learning for effective road surface features detection and segmentation,” *The Visual Computer*, vol. 38, no. 5, pp. 1759–1774, 2022.
- [132] Y. Li, S. Xue, Y. Mao, C. Zhao, X. Qin, and Y. Liu, “Classification of pavement disease 3d point cloud images based on deep learning network,” in *2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP)*, pp. 812–815, IEEE, 2021.

- [133] S. Chen, C. Chen, T. Ma, C. Han, H. Luo, S. Wang, Y. Gao, and Y. Yang, “Rapid extraction of pavement aggregate gradation based on point clouds using deep learning networks,” *Automation in Construction*, vol. 154, p. 105023, 2023.
- [134] R. Gui, Q. Sun, W. Wu, D. Zhang, and Q. Li, “Transfer learning for cross-scene 3d pavement crack detection based on enhanced deep edge features,” *Engineering Applications of Artificial Intelligence*, vol. 123, p. 106452, 2023.
- [135] H. Jiang, Q. Li, Q. Jiao, X. Wang, and L. Wu, “Extraction of wall cracks on earthquake-damaged buildings based on tfs point clouds,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 9, pp. 3088–3096, 2018.
- [136] H. Guan, J. Li, Y. Yu, M. Chapman, H. Wang, C. Wang, and R. Zhai, “Iterative tensor voting for pavement crack extraction using mobile laser scanning data,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 3, pp. 1527–1537, 2014.
- [137] H. Guan, J. Li, Y. Yu, M. Chapman, and C. Wang, “Automated road information extraction from mobile laser scanning data,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 194–205, 2014.
- [138] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, *et al.*, “Shapenet: An information-rich 3d model repository,” *arXiv preprint arXiv:1512.03012*, 2015.
- [139] D. Maturana and S. Scherer, “Voxnet: A 3d convolutional neural network for real-time object recognition,” in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pp. 922–928, IEEE, 2015.
- [140] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, “Pointnet: Deep learning on point sets for 3d classification and segmentation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 652–660, 2017.
- [141] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, “Pointnet++: Deep hierarchical feature learning on point sets in a metric space,” *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [142] E. Salari, E. Chou, J. J. Lynch, *et al.*, “Pavement distress evaluation using 3d depth information from stereo vision,” tech. rep., Michigan Ohio University Transportation Center, 2012.
- [143] Y.-C. J. Tsai and F. Li, “Critical assessment of detecting asphalt pavement cracks under different lighting and low-intensity contrast conditions using

- emerging 3d laser technology,” *Journal of Transportation Engineering*, vol. 138, no. 5, pp. 649–656, 2012.
- [144] I. Moazzam, K. Kamal, S. Mathavan, S. Usman, and M. Rahman, “Metrology and visualization of potholes using the microsoft kinect sensor,” in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, pp. 1284–1291, IEEE, 2013.
- [145] R. Fan, U. Ozgunalp, Y. Wang, M. Liu, and I. Pitas, “Rethinking road surface 3-d reconstruction and pothole detection: From perspective transformation to disparity map segmentation,” *IEEE Transactions on Cybernetics*, vol. 52, no. 7, pp. 5799–5808, 2021.
- [146] X. Chen, J. Li, S. Huang, H. Cui, P. Liu, and Q. Sun, “An automatic concrete crack-detection method fusing point clouds and images based on improved otsu’s algorithm,” *Sensors*, vol. 21, no. 5, p. 1581, 2021.
- [147] H. Kim, S. Lee, E. Ahn, M. Shin, and S.-H. Sim, “Crack identification method for concrete structures considering angle of view using rgb-d camera-based sensor fusion,” *Structural Health Monitoring*, vol. 20, no. 2, pp. 500–512, 2021.
- [148] C. Wang, D. Xu, Y. Zhu, R. Martín-Martín, C. Lu, L. Fei-Fei, and S. Savarese, “Densefusion: 6d object pose estimation by iterative dense fusion,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3343–3352, 2019.
- [149] J. Huang, W. Liu, and X. Sun, “A pavement crack detection method combining 2d with 3d information based on dempster-shafer theory,” *Computer-Aided Civil and Infrastructure Engineering*, vol. 29, no. 4, pp. 299–313, 2014.
- [150] H. Wu, L. Yao, Z. Xu, Y. Li, X. Ao, Q. Chen, Z. Li, and B. Meng, “Road pothole extraction and safety evaluation by integration of point cloud and images derived from mobile mapping sensors,” *Advanced Engineering Informatics*, vol. 42, p. 100936, 2019.
- [151] L. Yang, B. Li, W. Li, H. BrandH, B. Jiang, and J. Xiao, “Concrete defects inspection and 3d mapping using cityflyer quadrotor robot.,” *IEEE CAA J. Autom. Sinica*, vol. 7, no. 4, pp. 991–1002, 2020.
- [152] N. Bolourian, M. Nasrollahi, F. Bahreini, and A. Hammad, “Point cloud-based concrete surface defect semantic segmentation,” *Journal of Computing in Civil Engineering*, vol. 37, no. 2, p. 04022056, 2023.
- [153] J. Hoffman, S. Gupta, J. Leong, S. Guadarrama, and T. Darrell, “Cross-modal adaptation for rgb-d detection,” in *2016 IEEE international conference on robotics and automation (ICRA)*, pp. 5032–5039, IEEE, 2016.

- [154] D. Xu, D. Anguelov, and A. Jain, “Pointfusion: Deep sensor fusion for 3d bounding box estimation,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 244–253, 2018.
- [155] S. Gupta, R. Girshick, P. Arbeláez, and J. Malik, “Learning rich features from rgb-d images for object detection and segmentation,” in *European conference on computer vision*, pp. 345–360, Springer, 2014.
- [156] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, “Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes,” *arXiv preprint arXiv:1711.00199*, 2017.
- [157] J. Yu, K. C. Yow, and M. Jeon, “Joint representation learning of appearance and motion for abnormal event detection,” *Machine Vision and Applications*, vol. 29, no. 7, pp. 1157–1170, 2018.
- [158] X. Wang, R. Zhang, T. Kong, L. Li, and C. Shen, “Solov2: Dynamic and fast instance segmentation,” *Advances in Neural information processing systems*, vol. 33, pp. 17721–17732, 2020.
- [159] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, “Yolact: Real-time instance segmentation,” in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 9157–9166, 2019.
- [160] J. Yu, H. Oh, M. Kim, and J. Kim, “Normality-calibrated autoencoder for unsupervised anomaly detection on data contamination,” *arXiv preprint arXiv:2110.14825*, 2021.