OXFORD

Transportation Safety and Environment, 2022, **4**: tdac026

DOI: 10.1093/tse/tdac026

Review

Computer vision for road imaging and pothole detection: a state-of-the-art review of systems and algorithms

Nachuan Ma [1], Jiahe Fan [1], Wenshuo Wang [2], Jin Wu [3], Yu Jiang [4], Lihua Xie [5] and Rui Fan [1,*]

[1]Department of Control Science and Engineering, Frontiers Science Center for Intelligent Autonomous Systems, and State Key Laboratory of Intelligent Autonomous Systems, Tongji University, Shanghai 201804, P. R. China;
[2]Department of Civil Engineering, McGill University, Montréal, QC H3A 0C3, Canada;
[3]Department of Electronics and Computer Engineering, the Hong Kong University of Science and Technology, Hong Kong SAR 999077, P. R. China;
[4]CTO Office, ClearMotion Inc., Billerica, MA 01821, USA;
[5]School of Electrical and Electronic Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore.
*Corresponding author. E-mail: rui.fan@ieee.org

Abstract

Computer vision algorithms have been utilized for 3-D road imaging and pothole detection for over two decades. Nonetheless, there is a lack of systematic survey articles on state-of-the-art (SoTA) computer vision techniques, especially deep learning models, developed to tackle these problems. This article first introduces the sensing systems employed for 2-D and 3-D road data acquisition, including camera(s), laser scanners and Microsoft Kinect. It then comprehensively reviews the SoTA computer vision algorithms, including (1) classical 2-D image processing, (2) 3-D point cloud modelling and segmentation and (3) machine/deep learning, developed for road pothole detection. The article also discusses the existing challenges and future development trends of computer vision-based road pothole detection approaches: classical 2-D image processing-based and 3-D point cloud modelling and segmentation-based approaches have already become history; and convolutional neural networks (CNNs) have demonstrated compelling road pothole detection results and are promising to break the bottleneck with future advances in self/un-supervised learning for multi-modal semantic segmentation. We believe that this survey can serve as practical guidance for developing the next-generation road condition assessment systems.

Keywords: Computer vision, road imaging, pothole detection, deep learning, image processing, point cloud modelling, convolutional neural networks

1. Introduction

A pothole is a considerably sizeable structural road failure [1]. It is formed by the combined presence of water and traffic [2]. Water permeates the ground and weakens the soil under the road surface, and the traffic subsequently breaks the affected road surface, resulting in the removal of road surface chunks.

Road potholes are not just an inconvenience, they are also a significant threat to vehicle conditions and traffic safety [3]. For instance, the *Chicago Sun-Times* reported that drivers filed 11,706 complaints about road potholes in the first two months of 2018 [4]. According to *The Pothole Facts*, approximately one-third of 33,000 traffic fatalities in the United States involve poor road conditions. It is, therefore, necessary and crucial to frequently inspect roads and repair potholes [5].

Manual visual inspection is currently still the main form of road pothole detection [6]. Structural engineers and certified inspectors regularly detect road potholes and report their locations. This process is inefficient, expensive and dangerous. City councils in New Zealand, for example, spent millions of dollars in 2017 detecting and repairing potholes (Christchurch alone spent 525,000 USD) [7]. Additionally, it has been reported that more than 30,000 potholes are repaired in San Diego, the United States each year. San Diego residents were encouraged to report road potholes so as to relieve the burden of detection from the local road maintenance department [8]. Further, manual road pothole detection results produced by inspectors and engineers are always subjective, as the decisions depend entirely on an individual's experience and judgement [9]. For these reasons, researchers have been dedicated to developing automated road condition assessment systems that can reconstruct, recognize and localize road potholes efficiently, accurately and objectively [10]. Specifically, in recent years, road pothole detection has become more than just an infrastructure maintenance problem because it is also a function of advanced driver-assistance systems (ADAS) embedded into L3/L4 self-driving cars by many automotive companies, and emerging autonomous driving systems call for a higher road maintenance standard [11]. Jaguar Land Rover has experimented with data-driven technologies to inform drivers of pothole locations and issue warnings to slow down the car [12], while ClearMotion built an intelligent suspension system that uses a combination of hardware and software to anticipate, absorb and counteract the shocks and vibrations caused by road potholes [13].

Since the turn of the millennium, computer vision techniques have been extensively employed to acquire 3-D road data and/or detect road potholes. However, the latest survey on this research topic rarely discusses cutting-edge computer vision techniques, such as 3-D point cloud modelling and segmentation, machine/deep learning, etc. This article provides a comprehensive and thorough review of the state-of-the-art (SoTA) road imaging systems and computer vision-based pothole detection algorithms. An overview of the existing systems and algorithms

Received: December 22, 2021. **Revised:** April 1, 2022. **Accepted:** April 12, 2022

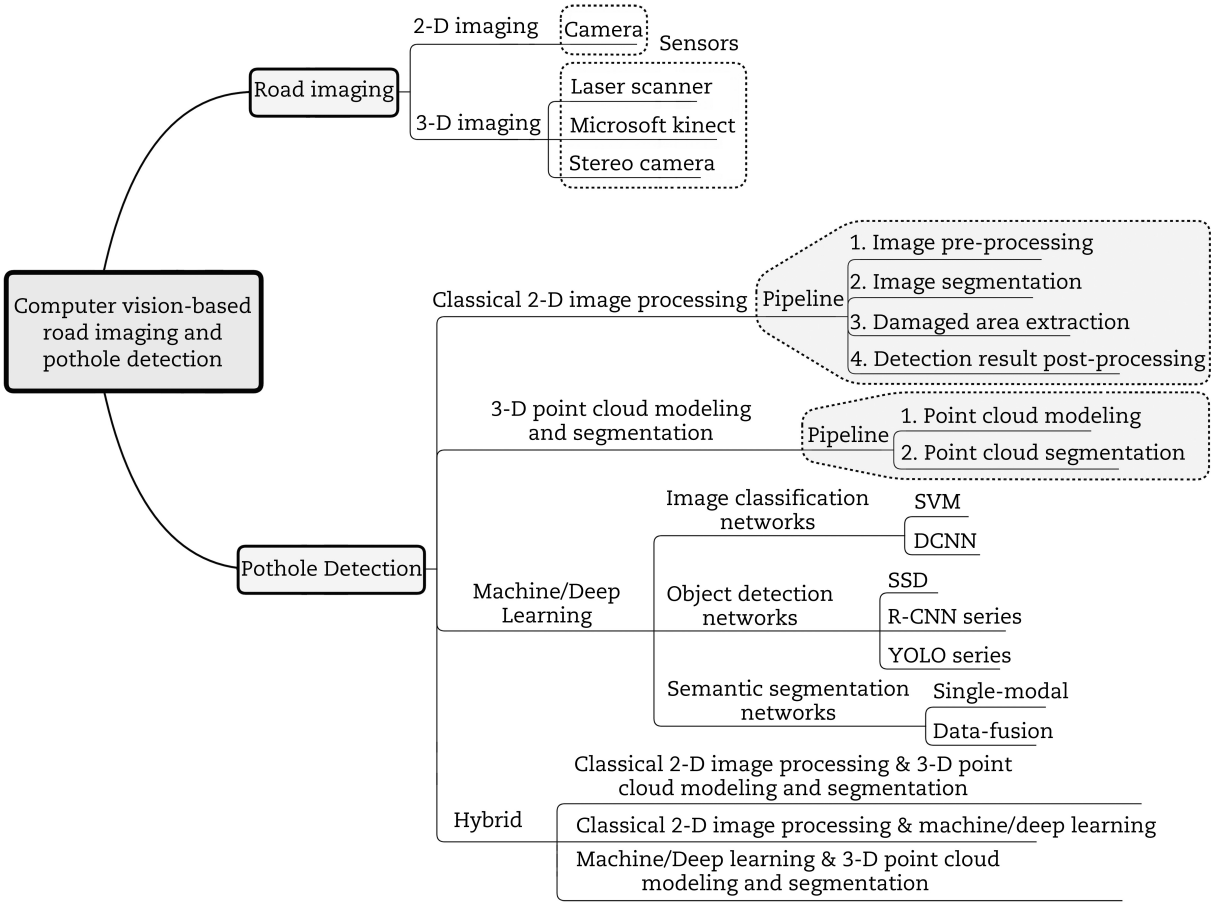

Fig. 1. An overview of road imaging systems and computer vision-based pothole detection algorithms.

is shown in Fig. 1. Laser scanners, Microsoft Kinect sensors and camera(s) are the three most prevalently used sensors for road data acquisition. The existing road pothole detection approaches are categorized into four groups: (1) classical 2-D image processing-based [14], (2) 3-D point cloud modelling and segmentation-based [15], (3) machine/deep learning-based [16] and (4) hybrid [3]. This article systematically reviews the prior arts (see Sections 2 and 3) and the open-access datasets (see Section 4), and discusses the existing challenges and their possible solutions (see Section 5). We believe that this article can provide readers with guidance when developing the next-generation 3-D road imaging and pothole detection algorithms.

2. Road imaging systems

Road imaging (or road data acquisition) is typically the first step of intelligent road inspection [10]. Cameras and range sensors have been extensively used to acquire visual road data. The use of 2-D imaging technology for this task began as early as 1991 [17]. However, the geometric structure of a road surface cannot be illustrated from unrelated 2-D road images (without overlapping areas) [18]. Additionally, the image segmentation algorithms performing on either grey-scale or colour road images can be severely affected by various environmental factors, most notably by poor illumination conditions [19]. Many researchers [5,18,20,21] have thus resorted to 3-D imaging technologies, which are more feasible for overcoming these two drawbacks. The most commonly used sensors for 3-D road data acquisition include laser scanners [22], Microsoft Kinect sensors [23] and stereo cameras [24], as shown in Fig. 2.

Laser scanning is a well-established imaging technology for accurate 3-D road data acquisition [1]. This technology is developed based on trigonometric triangulation [25]. The sensor (receiver) is

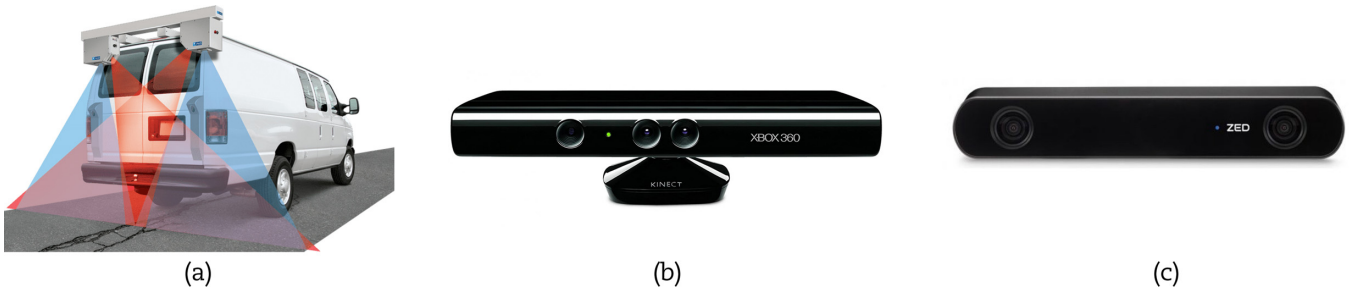

Fig. 2. Commonly used sensors for 3-D road data acquisition: (a) laser scanner [22]; (b) Microsoft Kinect [23]; (c) stereo camera [24].

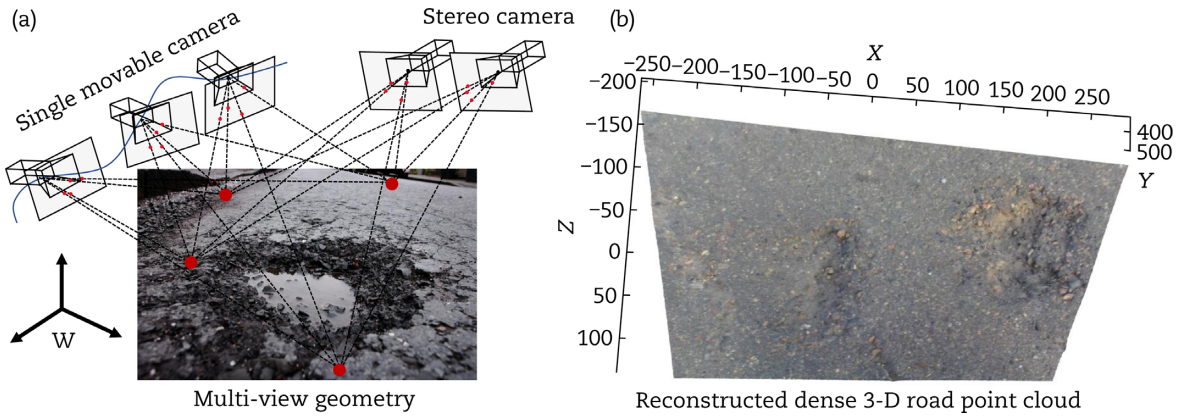

Fig. 3. 3-D road imaging with camera(s): (a) multi-view geometry; (b) reconstructed dense 3-D road point cloud.

located at a position with a known distance from the laser illumination source [26]. Accurate point measurements can, therefore, be made by calculating the reflection angle of the laser light. However, laser scanners have to be mounted on specific road inspection vehicles (see Fig. 2(a)) [27] for 3-D road data acquisition. Such vehicles are not widely used because of high equipment purchase and long-term maintenance costs.

Microsoft Kinect sensors [19] were initially designed for the Xbox-360 motion-sensing games, and are typically equipped with an RGB camera, an infrared sensor/camera, an infrared emitter, microphones, accelerometers and a tilt motor for motion tracking [1]. There have been three reported attempts [19,27,28] at 3-D road data acquisition using Microsoft Kinect sensors. Although such sensors are cost-effective and convenient to use, they greatly suffer from infra-red saturation in direct sunlight, and the 3-D road surface reconstruction accuracy is unsatisfactory [3].

3-D road data can also be obtained using multiple 2-D road images captured from different views, e.g. using either a single movable camera [29] or an array of synchronized cameras [20], as illustrated in Fig. 3. The theory behind this technique is gener-

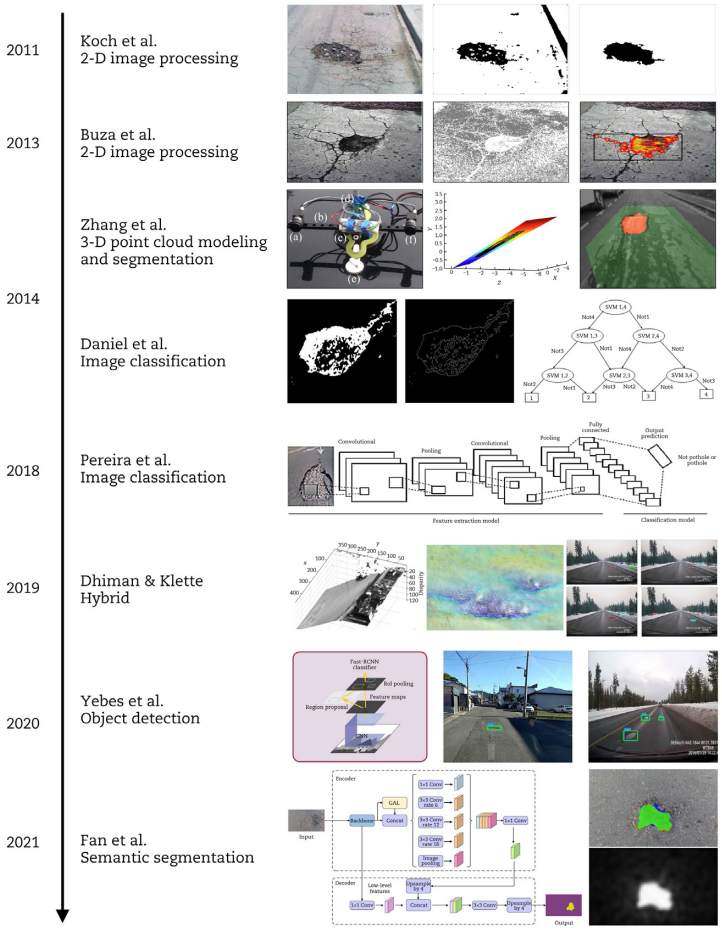

Fig. 4. The most representative road pothole detection algorithms developed between 2011 and 2021.

ally known as multi-view geometry [30]. The essential task of 3-D geometry reconstruction from multiple views is sparse or dense correspondence matching. A typical monocular sparse road surface 3-D reconstruction approach, as shown in Ref. [31], where the camera poses and sparse 3-D road point clouds are obtained using a structure from motion (SfM) [32] algorithm and are refined using a bundle adjustment (BA) [33] algorithm.

The use of stereo cameras for dense 3-D road point cloud acquisition was pioneered by researchers from Bristol Visual Information Laboratory [18,34,35]. In this case, depth information is acquired by finding the horizontal positional differences (disparities) of the visual feature correspondence pairs between two synchronously captured road images [36]. This process is commonly referred to as disparity estimation or stereo matching, which mimics human binocular vision. Ref. [34] proposed a seed-and-grow disparity estimation algorithm to acquire 3-D road data efficiently. Ref. [35] introduced a more adaptive disparity search range propagation strategy to improve the accuracy of the estimated road disparities. Refs. [5,18] utilized a perspective transformation algorithm to transform the target image into the reference view, which significantly minimizes the trade-off between stereo matching speed and disparity accuracy. Additionally, the bottleneck problems existing in Refs. [34,35] were also tackled with the use of efficient and adaptive cost volume processing algorithms. It is reported in Refs. [5,18] that the accuracy of the reconstructed 3-D road geometry models is over 3 mm (an example is given in Fig. 3). Compared to laser scanners and Microsoft Kinect sensors, stereo cameras are cheaper and more reliable for 3-D road imaging. With the recent advances in deep learning, convolutional neural networks (CNNs) have demonstrated greater disparity estimation results than traditional explicit programming methods. Their limitations and future development trends will be discussed in Section 5.

3. Road pothole detection approaches

The taxonomy of SoTA computer vision-based road pothole detection algorithms is illustrated in Fig. 1. The classical 2-D image processing-based algorithms process (e.g. enhance, compress, transform, segment) road RGB or disparity/depth images with explicit programming [9]. Machine/deep learning-based algorithms address the road pothole detection problem using image classification, object recognition or semantic segmentation algorithms, solvable with SoTA CNNs [37]. 3-D road point cloud modelling and segmentation-based algorithms fit a specific geometry model

Table 1. Representative classical 2-D image processing-based approaches.

Reference	Input	Details
Koch and Brilakis [14] (2011)	Colour image	A road image is segmented into damaged and undamaged road regions using a histogram-based thresholding method. The damaged road areas are processed with morphological operations and elliptic regression. The road potholes are detected by comparing the road textures inside and outside the ellipse.
Buza et al. [38] (2013)	Colour image	Otsu's thresholding method is adopted to segment road images. Spectral clustering is utilized to extract damaged road areas (potholes).
Ryu et al. [39] (2015)	Colour image	Road images are processed with morphological filters and segmented using a histogram-based thresholding method. A potential road pothole contour is extracted based on geometric properties. An ordered histogram intersection method is used to determine whether the extracted area contains a road pothole.
Schiopu et al. [40] (2016)	Colour image	A histogram-based thresholding method is utilized to generate a set of road pothole candidates. The candidates with specific geometric properties are determined to be road potholes.
Jakštys et al. [41] (2016)	Colour image	Triangle thresholding and adaptive thresholding methods are used to segment road images. A heuristic edge detection approach is designed for road pothole contour extraction.
Akagic et al. [42] (2017)	Colour image	Road pothole regions of interest (RoIs) are detected by (1) manipulating the B component in the RGB colour space and (2) performing two-level dynamic road pixel selection. The search for road potholes is conducted only in the RoIs. The road potholes are detected by comparing two cropped road images based on the method proposed in Ref. [38].
Wang et al. [43] (2017)	Grey-scale image	The wavelet energy field of a road image is constructed to highlight road potholes. Damaged road areas are processed with morphological filters. A Markov random fields-based image segmentation method is used to segment the damaged road areas for pothole detection. Morphological filters are used again to refine the road pothole detection results.
Chung and Khan [44] (2019)	Grey-scale image	Otsu's thresholding method is used to segment road images. The segmented images are processed with morphological filters before performing distance transform. The watershed algorithm is applied to the distance transform images for road pothole detection.
Moazzam et al. [28] (2013)	Depth images	The road potholes are detected by analysing road depth distribution with respect to different azimuth and elevation angles. The approximate volume of each road pothole is calculated using the trapezoidal rule with unit spacing on the area-depth curves.
Fan et al. [6] (2019)	Transformed disparity image	A dense road disparity image is transformed to better distinguish the damaged and undamaged road areas. The transformed disparity image is segmented using Otus's thresholding method for road pothole detection.
Fan et al. [5] (2021)	Transformed disparity image	SLIC is utilized to group the transformed disparities into a collection of superpixels. The road potholes are then detected by finding the superpixels, whose values are lower than an adaptively determined threshold.

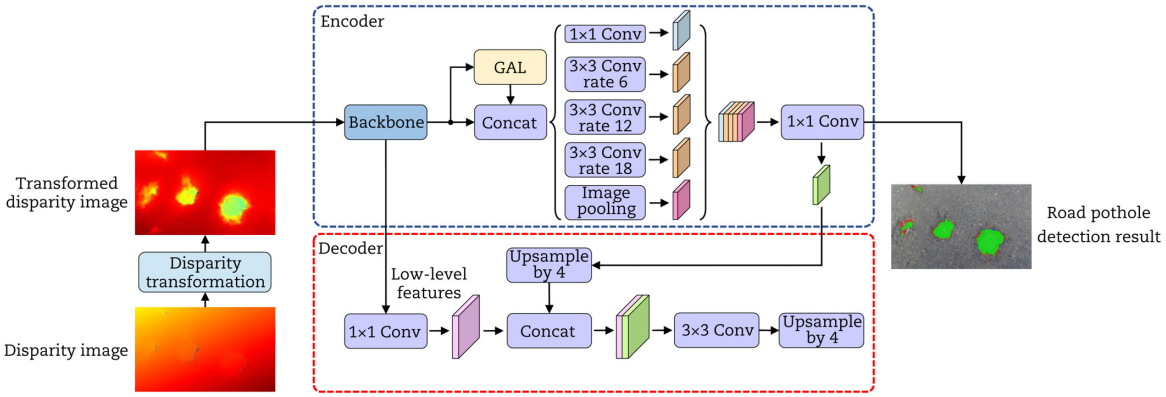

Fig. 5. Semantic segmentation for road pothole detection [37]. (The disparity transformation algorithm was introduced in Refs. [3,6].)

(typically a planar or quadratic surface) to the observed road point cloud and segment the road point cloud by comparing the observed and fitted surfaces [3]. Hybrid methods combine two or more categories of algorithms mentioned above to improve the overall road pothole detection performance. The most representative road pothole detection algorithms (from classical 2-D image processing-based to deep learning-based) developed between 2011 and 2021 are shown in Fig. 4.

3.1 Classical 2-D image processing

Classical 2-D image processing-based road pothole detection is a well-researched topic. As shown in Fig. 1, such approaches generally have a four-stage pipeline: (1) image pre-processing, (2) image segmentation, (3) damaged area extraction and (4) detection result post-processing [9]. The representative prior arts are summarized in Table 1.

Image pre-processing algorithms, such as median filtering [42], Gaussian filtering [45], bilateral filtering [46] and morphological filtering [47], are first utilized to reduce redundant information and highlight the damaged road areas. For instance, an adaptive histogram equalization algorithm is used in Ref. [45] to adjust the image brightness before binarizing the road images, and a Leung-Malik filter [48] and Schmid filter [49] are used in Ref. [14] to emphasize structural texture characteristics in colour road images. Recently, many researchers have resorted to 2-D spatial visual information (typically road depth/disparity images) for pothole detection [3,5,6,28,50]. For example, Refs. [3,50] transformed a road disparity image with a stereo rig roll angle and road disparity projection model, which were estimated by minimizing a global energy function using golden section search [51] and dynamic programming algorithms [52]. Disparity transformation makes the damaged road areas highly distinguishable, as illustrated in Fig. 5. Ref. [6] yields the closed-form solution of the above energy minimization problem, and therefore avoids the intensive computations in the iterative optimization process. As depth/disparity images can depict the geometric structure of road surfaces, they are more informative for pothole detection.

The pre-processed road images are then segmented to separate foreground (damaged road areas) and background (undamaged road areas). Most prior arts [38,41,46] employ histogram-based thresholding methods, such as Otsu's thresholding [53], triangle thresholding [14] and adaptive thresholding [40,46], to segment colour/grey-scale road images. As discussed in Ref. [38], Otsu's thresholding [53] method minimizes the intra-class variance and achieves better performance than the triangle thresholding [14] method in terms of segmenting road images. Ref. [41] employs an adaptive thresholding method to segment road images, and it also outperforms the commonly used triangle thresholding method. Recent works demonstrated that such image segmentation algorithms typically work more effectively and accurately on transformed disparity images, depicting the quasi bird's eye view of the road scene [3,5,6,50]. For example, Ref. [3] utilizes Otsu's thresholding [53] method to segment the transformed disparity images for road pothole detection, and in Ref. [5], a simple linear iterative clustering (SLIC) algorithm [54] is used to group the transformed disparities into a collection of superpixels. The road potholes are then detected by finding the superpixels, whose values are lower than an adaptively determined threshold.

The third and fourth stages are typically performed in a joint manner. The damaged road areas (potholes) are first extracted from the segmented foreground based on geometric and textural assumptions [5]:

1). Potholes are typically concave holes;
2). The pothole texture is typically grainier and coarser than that of the surrounding road surface;
3). The intensities of the pothole RoI pixels are typically lower than those of the surrounding road surface due to shadows.

For example, in Ref. [14], the contour of a potential pothole is modelled as an ellipse. The image texture within the ellipse is then compared with that of the undamaged road areas. If the elliptical RoI has a coarser and grainier texture than that of the surrounding region, the ellipse is identified as a road pothole. In Ref. [39], the contour of a potential pothole is extracted by analysing various geometric features, such as size, compactness, ellipticity and convex hull. An ordered histogram intersection method is then used to determine whether the extracted region contains a road pothole. Finally, the extracted damaged road areas are post-processed to further improve the road pothole detection results. This process is typically similar to the first stage.

Classical 2-D image processing-based road pothole detection approaches have been researched for almost two decades. These types of algorithm have been systematically studied [9] and we refer readers to this paper for more details. However, such approaches were developed based on early techniques and can be severely affected by various environmental factors. Fortunately, modern 3-D computer vision and machine learning algorithms have greatly overcome these shortcomings.

Table 2. Representative 3-D point cloud modelling and segmentation-based approaches.

Reference	Input	Key algorithm(s)	Details
Zhang and Elaksher [31] (2012)	3-D point cloud	SfM, BA, 3-D feature extraction	Sparse 3-D road geometry models are reconstructed with SfM and refined with BA. Road potholes are detected by finding distinguishable 3-D features.
Zhang [34] (2013)	3-D point cloud	Stereo vision, quadratic surface fitting, connected component labelling (CCL)	A quadratic surface is fitted to the observed 3-D road point cloud. The 3-D points under the fitted surface are considered to be part of road potholes. Different road potholes are labelled using CCL.
Li et al. [55] (2018)	3-D point cloud	Stereo vision, planar surface fitting, bi-square weighted robust least-squares approximation, CCL	An observed 3-D road point cloud is interpolated into a planar surface using a bi-square weighted robust least-squares approximation. The 3-D points under the fitted surface are considered to be part of road potholes. CCL is also used to label different road potholes.
Du et al. [56] (2020)	3-D point cloud	Stereo vision, planar surface fitting and segmentation, K-means clustering, region growing	The surface normal information is incorporated into the road surface modelling process. K-means clustering and region growing algorithms are used to extract road potholes.

Table 3. Image classification-based approaches.

Reference	Input	Key algorithm(s)	Details
Lin and Liu [16] (2010)	Grey-scale image	NL-SVM	Average grey level, contrast, consistency, entropy and three-order moments of grey-scale road images are computed to create hand-crafted visual features. An NL-SVM model is trained to learn these features for road image classification.
Daniel and Preeja [57] (2014)	Grey-scale image	SVM	Classical image processing algorithms are utilized to reduce road image noise and highlight informative visual features; CCL is then employed to obtain the connected components. The five most prominent components are selected as training samples to train an SVM model for road image classification.
Hadjidemetriou et al. [58] (2016)	Grey-scale image	SVM, DCT, GLCM	Road image patches are utilized to generate feature vectors using discrete cosine transform (DCT) [59] and grey-level co-occurrence matrix (GLCM) algorithms [60]. An SVM model is then trained with such feature vectors to realize binary road patch classification.
Hoang [61] (2018)	Grey-scale image	LS-SVM, ANN	Classical image processing algorithms are used to generate hand-crafted visual features. A least-squares SVM (LS-SVM) model and an artificial neural network (ANN) model are trained with such hand-crafted visual features to recognize road images containing potholes.
Pan et al. [62] (2018)	Colour image, Multi-spectral image	ANN, RF, SVM	Spectral, geometric and textural features are extracted. Three models—ANN, random forest (RF) and SVM—are trained to learn these features for road image classification.
Gao et al. [63] (2020)	Colour image	LIBSVM	Classical image processing algorithms, including binarization, morphology operations and integral projection, are used to generate hand-crafted visual features. A model based on the library for SVM (LIBSVM) is trained to detect road potholes and cracks.
Pereira et al. [64] (2018)	Colour image	Self-designed DCNN	A DCNN, consisting of four convolutional-pooling layers and one FC layer, is developed from scratch to classify road images.
An et al. [65] (2018)	Colour image, grey-scale image	Inception, ResNet, and MobileNet	Four existing DCNNs are trained to classify colour and grey-scale road image patches.
Ye et al. [66] (2019)	Colour image	Self-designed DCNN	A DCNN containing a pre-pooling layer (used to reduce the characteristics unrelated to road potholes) is designed from scratch to classify road images.
Bhatia et al. [67] (2019)	Thermal image	Self-designed DCNN	A DCNN model (with ResNet as the backbone network) is designed to classify thermal road images.

3.2 3-D point cloud modelling and segmentation

An example of the reconstructed dense 3-D road point clouds is given in Fig. 3. The approaches designed to process 3-D road point clouds generally have a two-stage pipeline [34,68]: (1) interpolating the observed 3-D road point cloud into an explicit geometric model (typically a planar or quadratic surface), and (2) segmenting the observed 3-D road point cloud by comparing it with the interpolated geometric model. The most representative algorithms are summarized in Table 2.

Taking Ref. [34] as an example, quadratic surfaces are fitted to dense 3-D road point clouds using least-squares fitting. By comparing the difference (elevation) between the observed and fitted 3-D road surfaces, the damaged road areas (potholes) can be effectively extracted. Different potholes are also labelled using a con-

nected component labelling (CCL) algorithm. Similarly, Ref. [56] interpolates the observed 3-D road point clouds into planar surfaces. The potential road potholes are roughly detected by finding the 3-D points under the fitted surface. K-means clustering [69] and region growing algorithms are subsequently used to refine the road pothole detection results.

Least-squares fitting, however, can be severely affected by outliers, often making the modelled road surface inaccurate [3]. Therefore, Ref. [55] employs the bi-square weighted robust least-squares approximation for road point cloud modelling. Ref. [50] utilized the random sample consensus (RANSAC) algorithm [70] to improve the robustness of quadratic surface fitting. Refs. [3,35] incorporated surface normal information into the process of quadratic surface fitting, which greatly enhances the performance of freespace and road pothole detection.

In addition to the aforementioned camera-based approaches, Ref. [71] employs high-speed 3-D transverse scanning technology for road shoving (abrupt waves across the road surface) and pothole detection. A subpixel line extraction method (including point cloud filtering, edge detection and spline interpolation) is performed on the laser stripe data. The road transverse profile is then generated from the laser stripe curve and is approximated by line segments. The second-order derivatives of the segment endpoints are used to identify the feature points of possible shoving and potholes. Recently, Ref. [72] introduced a LiDAR-based road pothole detection system, where the 3-D road points are classified as damaged and undamaged by comparing their distances to the best-fitting planar 3-D road surface. Unfortunately, Ref. [72] lacks the algorithm details and necessary quantitative experimental road damage detection results.

3-D point cloud modelling and segmentation-based methods are relatively rare compared to other approaches. Nevertheless, actual roads are always uneven, making such approaches occasionally unfeasible. Furthermore, acquiring 3-D road point clouds might not be necessary if the objective is only to recognize and localize road potholes instead of obtaining their geometric details. With the combination of 2-D image processing algorithms, the 3-D point cloud modelling performance can be significantly boosted [3].

3.3 Machine/deep learning

With recent advances in machine/deep learning, deep CNNs (DCNNs) have become the mainstream techniques for road pothole detection. Instead of setting explicit parameters to segment road images or point clouds for pothole detection, DCNNs are typically trained through back-propagation with a large amount of human-annotated road data [73]. Data-driven road pothole detection approaches are generally developed based on three types of techniques [26]: (1) image classification networks, (2) object detection networks and (3) semantic segmentation networks. Image classification networks are trained to classify positive (pothole) and negative (non-pothole) road images, object detection networks are trained to recognize road potholes at the instance level and semantic segmentation networks are trained to segment road (colour or disparity/depth) images for pixel-level (or semantic-level) road pothole detection. The remainder of this section details each type of these algorithms.

3.3.1 *Image classification-based methods*

Before deep learning technology exploded, researchers typically used classical image processing algorithms to generate hand-crafted visual features and trained a support vector machine (SVM) [74] model to classify road image patches. The most representative SVM-based approaches [16,57,58,61–63,75] are summarized in Table 3. As such algorithms are already outdated, we do not present readers with too many details here.

With the revolution of computational resources and the increase in training data sample size, DCNNs have been extensively used for road pothole detection. Compared to the traditional SVM-based approaches, DCNNs are capable of learning more abstract (hierarchical) visual features, and they have significantly improved the road pothole detection performance [46]. The most typical DCNN-based approaches [64–67] are summarized in Table 3. Refs. [64,66] designed DCNNs from scratch. The DCNN presented in Ref. [64] consists of four convolutional-pooling layers and one fully connected (FC) layer. Extensive experiments on the road data collected in Timor-Leste demonstrated the effectiveness of such a DCNN in terms of classifying pothole and non-pothole images. The DCNN introduced in Ref. [66] consists of a pre-pooling layer, three convolutional-pooling layers, a sigmoid layer and two FC layers. The pre-pooling layer was designed to reduce the characteristics unrelated to road potholes. The experimental results suggest that the proposed pre-pooling layer can greatly improve the performance of road image classification, and that the designed DCNN can effectively detect road potholes under different illumination conditions.

Refs. [65,67] developed road image classification networks based on existing DCNNs. Ref. [67] developed a DCNN based on the popular residual network [76]. Extensive experiments demonstrated that the proposed model can effectively classify thermal road images collected in the night and/or foggy weather, and it also outperforms the prior arts [61,65,77]. In Ref. [65], four well-developed DCNNs—(1) Inception-v4 [78], (2) Inception with ResNet-v2 [78], (3) ResNet-v2 [79] and (4) MobileNet-v1 [80]—are trained to classify road images. The experimental results suggest that these models performed similarly on the test set. Recently, Ref. [81] compared 30 SoTA image classification DCNNs in terms of detecting road cracks and found that road crack detection is a relatively easy task compared to the image classification tasks in other application domains. Road pothole detection is an easier task compared to road crack detection. Therefore, we believe that road pothole detection with image classification networks is a well-solved problem.

3.3.2 *Object detection-based methods*

As illustrated in Fig. 1, object detection-based road pothole detection approaches can be grouped into three types: (1) single shot multi-box detector (SSD)-based, (2) region-based CNN (R-CNN) series-based and (3) you only look once (YOLO) series-based. The most representative object detection-based approaches are summarized in Table 4.

An SSD has two components [82], namely a backbone model and an SSD head. The former is a deep image classification network for visual feature extraction, while the latter is one or more convolutional layers added to the backbone so that the outputs can be bounding boxes with object classes. Researchers in this field have mainly incorporated different image classification networks into the SSD for road pothole detection. For example, Inception-v2 [83] and MobileNet [80] were used as the backbone networks in Ref. [84], while ResNet-34 [76] and RetinaNet [85] were used as the backbone networks in Ref. [86].

Table 4. Object detection-based approaches.

Reference	Input	Key algorithm(s)	Details
Suong et al. [87] (2018)	Colour image	YOLO	Two object detection networks—F2-Anchor and Den-F2-Anchor—developed based on YOLOv2, are trained to detect potholes in the colour road images.
Maeda et al. [84] (2018)	Colour image	SSD	Two SSD-based DCNNs (with Inception-v2 and MobileNet as the backbone networks, respectively) are trained to detect potholes in colour road images.
Wang et al. [88] (2018)	Colour image	Faster R-CNN	Two Faster R-CNNs (with ResNet-101 and ResNet-152 as the backbone networks, separately) are trained to detect road potholes.
Ukhwah et al. [89] (2019)	Grey-scale image	YOLO	YOLOv3, YOLOv3 Tiny and YOLOv3 SPP are trained to detect potholes in grey-scale road images. YOLOv3 SPP achieves the best overall performance.
Dharneeshkar et al. [90] (2020)	Colour image	YOLO	YOLOv2, YOLOv3 and YOLOv3 Tiny are trained to detect road potholes. YOLOv3 Tiny achieves the highest mAP, precision and recall.
Baek and Chung [91] (2020)	Colour image	YOLO	Two YOLOv1 models are trained to detect cars (in the background) and road potholes (in the foreground).
Kortmann et al. [92] (2020)	Colour image	Faster R-CNN	A classifier is first trained to infer the country where the road image was taken. A Faster R-CNN is then trained with respect to each country for road crack and pothole detection.
Yebes et al. [93] (2020)	Colour image	Faster R-CNN, SSD	Three Faster R-CNNs (with Inception-ResNet-v2, Inception-v2 and ResNet-101 as the backbone networks, separately) and one SSD (with MobileNet-v2 as the backbone network) are trained to detect road potholes. Faster R-CNN (with ResNet-101 as the backbone network) achieves the best performance.
Gupta et al. [86] (2020)	Thermal image	SSD	Two SSDs (with ResNet-34 and ResNet-50 as the backbone networks, separately) are trained to detect potholes in thermal road images. The latter significantly outperforms the former.
Javed et al. [94] (2021)	Colour image	R-CNN, SSD	R-CNN and SSD are compared on the road data collected in Bangladesh. They achieve similar road pothole detection performances.

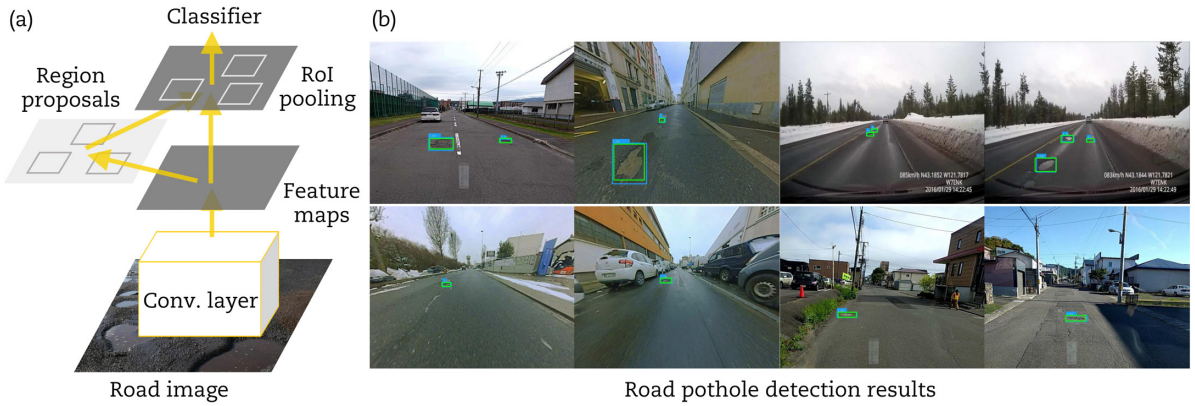

Fig. 6. Faster R-CNN architecture [96] and road pothole detection results [93]: (a) road image; (b) road pothole detection results.

Compared to SSD, R-CNN and YOLO series are more widely used for road pothole detection. In Ref. [94], R-CNN was demonstrated to achieve a similar performance to SSD for road pothole detection. In Ref. [93], four road pothole detection networks were developed: (1) Faster R-CNN (with Inception-v2 [83] as the backbone network), (2) Faster R-CNN (with ResNet-101 as the backbone network [76]), (3) Faster R-CNN (with Inception-ResNet-v2 as the backbone network [78]) and (4) SSD (with MobileNet-v2 [95] as the backbone network). Extensive experiments demonstrated that Faster R-CNN (with ResNet-101 as the backbone network) achieved the best overall performance [93,96]. Their experimental results are shown in Fig. 6. Ref. [88] compares the performance of two Faster R-CNNs (with ResNet-101 and ResNet-152 as the backbone networks, separately) for road damage detection on the dataset introduced in Ref. [84] with respect to three evaluation metrics: F1-Score, the harmonic mean of the precision and the harmonic mean of the recall. The experimental results indicated that Faster R-CNN (with ResNet-152 as the backbone network) outperforms Faster R-CNN (with ResNet-101 as the backbone network). This is probably because a deeper backbone can learn more abstract representations. Ref. [92] utilizes a Faster R-CNN to detect both cracks and potholes in the road images captured in Japan, India and the Czech Republic. A classifier is first trained to infer the country where the road image was captured. A Faster R-CNN is then trained with respect to each country (in order to reduce the effects caused by the regional difference) for road crack and pothole detection.

Unlike the R-CNN series, which uses region proposals to localize road potholes within the image, the YOLO series generally splits the road image into a collection of grids, and within each grid a collection of bounding boxes is selected. The network outputs a class probability and the offset values for each bound-

Table 5. Semantic segmentation-based approaches.

Reference	Input	Key algorithm(s)	Details
Pereira et al. [97] (2019)	Colour image	U-Net	A conventional U-Net is trained to segment colour road images for pothole detection.
Chun and Ryu [98] (2019)	Colour image	FCN	An FCN is trained to segment colour road images; a semi-supervised learning strategy is also employed to produce additional pseudo labels for network fine-tuning.
Fan et al. [11] (2020)	Colour image, transformed disparity image	AA, GAN	An AA framework and a training set augmentation technique are developed to enhance both single-modal and data-fusion semantic segmentation networks. The developed networks outperform all other SoTA networks.
Masihullah et al. [99] (2021)	Colour image	DeepLabv3+	An attention-based feature refinement module is incorporated into DeepLabv3+ for road pothole detection; the effectiveness of few-shot learning for road pothole detection is also validated.
Fan et al. [100] (2021)	Colour image, transformed disparity image	DeepLabv3+	An MSFFM is proposed to refine the learning representations in single-modal semantic segmentation networks for road pothole detection.
Fan et al. [37] (2021)	Colour image, disparity image, transformed disparity image	DCNNs with GAL	A GNN-inspired GAL is designed; GAL-DeepLabv3+ achieves the best road pothole detection performance over all other SoTA single-modal DCNNs on colour images, disparity images and transformed disparity images.

ing box. The bounding boxes with the class probability above a threshold value are used to locate the road potholes within the image. Thanks to its accuracy and efficiency, the YOLO series has become the first choice for object detection-based road pothole detection. For example, in Ref. [87], two object detection DCNNs, referred to as F2-Anchor and Den-F2-Anchor, were developed for road pothole detection. F2-Anchor, a variant of YOLOv2, is capable of generating five new anchor boxes (obtained using the K-means clustering algorithm [69]). The experimental results suggest that F2-Anchor outperforms the original YOLOv2 in detecting road potholes with different shapes and sizes. Compared to F2-Anchor, Den-F2-Anchor densifies the grid and achieves better road pothole detection performance than YOLOv2 and F2-Anchor. Additionally, Ref. [90] trained three YOLO architectures—YOLOv3 [101], YOLOv2 [102] and YOLOv3 tiny [101]—for road pothole detection. YOLOv3 tiny achieved the best overall road pothole detection accuracy. Similarly, Ref. [89] compared three different YOLOv3 architectures—YOLOv3 [101], YOLOv3 Tiny [101] and YOLOv3 SPP [101]—for road pothole detection. YOLOv3 SPP demonstrated the highest road pothole detection accuracy. Recently, Ref. [91] designed a hierarchical road pothole detection approach with two YOLOv1 networks [103]. One pre-trained YOLOv1 model was used to detect cars (background), while the other YOLOv1 was used to detect road potholes from the foreground. Nevertheless, the aforementioned object detection approaches can only recognize road potholes at the instance level, and they are unfeasible when pixel-level road pothole detection results are desired.

3.3.3 *Semantic segmentation-based methods*

As shown in Fig. 1, the SoTA semantic segmentation networks are grouped into two major categories: (1) single-modal and (2) data-fusion. Single-modal networks generally segment RGB images with encoder-decoder architectures [100]. Data-fusion networks typically learn visual features from two different types of vision sensor data (colour images and depth maps were used in FuseNet [104], colour images and surface normal maps were used in SNE-RoadSeg series [105,106] and colour images and transformed disparity images were used in AA-RTFNet [11]) and fuse the learned visual features to provide a better semantic understanding of the environment. The most representative prior arts are summarized in Table 5.

Ref. [98] proposes a road pothole detection approach based on fully convolutional network (FCN). To mitigate the difficulty in providing pixel-level annotations required by supervised learning, Ref. [98] exploits a semi-supervised learning technique to generate pseudo labels and fine-tune the pre-trained FCN automatically. Compared to supervised learning, semi-supervised learning can greatly improve the overall F-score. Additionally, Ref. [100] incorporates an attention-based multi-scale feature fusion module (MSFFM) into DeepLabv3+ [107] for road pothole detection. Similarly, Ref. [99] proposes an attention-based coupled framework for road pothole detection. This framework leverages an attention-based feature fusion module to improve the image segmentation performance. The work also demonstrates the effectiveness of few-shot learning for road pothole detection.

We have conducted extensive research in this field. Ref. [11] introduces an attention aggregation (AA) framework, which takes the advantages of three types of attention modules: (1) channel attention module (CAM), (2) position attention module (PAM) and (3) dual attention module (DAM). Additionally, Ref. [11] develops an effective training set augmentation technique based on generative adversarial network (GAN), where fake colour road

images and transformed road disparity images are generated to enhance the training of semantic segmentation networks. The experimental results demonstrated that (1) AA-UNet (single-modal network) outperforms all other SoTA single-modal for road pothole detection, (2) AA-RTFNet (data-fusion network) outperforms all other SoTA data-fusion networks for road pothole detection and (3) the training set augmentation technique not only improves the accuracy of the SoTA semantic segmentation networks but also accelerates their convergence during training. Recently, we developed a graph attention layer (GAL) based on graph neural network (GNN) to further optimize image feature representations for single-modal semantic segmentation [37]. As illustrated in Fig. 5, GAL-DeepLabv3+, the best performing implementation, outperforms all other SoTA single-modal semantic segmentation DCNNs for road pothole detection.

It should be noted here that road pothole detection can be jointly solved with other driving scene understanding problems, notably freespace and road anomaly detection [105,106,108–110]. Unfortunately, the SoTA semantic segmentation networks are strong data-driven algorithms that require a considerable amount of data. Therefore, road pothole detection with unsupervised or self-supervised learning is a popular area of research that requires more attention.

3.4 Hybrid methods

Hybrid road pothole detection approaches typically leverage at least two categories of algorithms mentioned above, as shown in Fig. 1. They have been extensively studied for over a decade. Such approaches, as summarized in Tables 6, have brought the SoTA results to this task.

A decade ago, Ref. [111] developed a hybrid road pothole detection approach based on classical 2-D image processing as well as 3-D point cloud modelling and segmentation. An image gradient filter was first performed on the road videos (collected by a high-speed camera) to select keyframes that were considered to contain road potholes. The keyframes' 3-D road point

Table 6. Hybrid approaches.

Type	Reference	Input	Details
Classical 2-D image processing & 3-D point cloud modelling and segmentation.	Joubert et al. [111] (2011)	3-D point cloud, colour image	The keyframes (potentially containing road potholes) are selected using 2-D image processing algorithms; road potholes in the keyframes are detected by comparing the observed and modelled 3-D road point clouds.
	Jog et al. [29] (2012)	3-D point cloud, colour image	The road videos are analysed with 2-D image processing algorithms to produce keyframes; the road videos are also used to reconstruct 3-D road geometry for road pothole detection.
	Jahanshahi et al. [19] (2013)	3-D point cloud, depth image	A planar surface is fitted to the depth image; a normalized depth-difference image, reflecting the difference between the observed the fitted depth images, is created; Otsu's thresholding method is used to segment the normalized depth-difference image for road pothole detection.
	Fan et al. [3] (2019)	3-D point cloud, transformed disparity image	A disparity image is transformed into a quasi bird's eye view. Otsu's thresholding method is utilized to segment the transformed disparity image to produce the undamaged road areas. The 3-D points in the undamaged road areas are used to interpolate a quadratic surface; road potholes are detected by comparing the observed and interpolated surfaces.
Classical 2-D image processing & machine/deep learning.	Azhar et al. [112] (2016)	Colour image	HOG features are extracted from road images; an NBC is trained with the HOG features to classify road images. The NGCS method is used to segment road images potentially containing potholes.
	Yousaf et al. [113] (2018)	Colour image	Road images are classified using the BoW algorithm; the GCS is used to segment the road images that potentially contain potholes.
	Anand et al. [114] (2018)	Colour image	A SegNet is trained to segment road images for freespace detection; the freespace regions are processed to generate road pothole/crack candidates; a SqueezeNet is trained to determine whether the generated candidates were road potholes or cracks.
Machine/deep learning & 3-D point cloud modelling and segmentation.	Dhiman and Klette [115] (2019)	3-D point cloud, colour image, disparity image	Four existing computer vision techniques are compared: (1) single-frame stereo vision-based method; (2) multi-frame vision sensor data fusion-based method; (3) Mask R-CNN trained with transfer learning; and (4) YOLOv2 trained with transfer learning.
	Wu et al. [116] (2019)	3-D point cloud, colour image	A semantic segmentation network is used to provide initial road pothole detection results; a 3-D point cloud modelling and segmentation algorithm is used to refine such results and calculate the road pothole volumes.

clouds (acquired by Microsoft Kinect) were simultaneously modelled as planar surfaces. Similar to Ref. [50], RANSAC was employed to enhance the robustness of 3-D road point cloud modelling. Road potholes were then detected by comparing the observed and modelled road surfaces. Thanks to the efficient 2-D image processing-based keyframe selection, the approach greatly reduces the redundant computations in 3-D point cloud modelling. Ref. [29] presents a similar hybrid approach. A road video collected by a high-definition camera is first processed to recognize the keyframes potentially containing road potholes. Simultaneously, this road video is also utilized for sparse-to-dense 3-D road geometry reconstruction. The road potholes are efficiently and accurately detected by analysing such multi-modal road data. Such a hybrid method significantly reduces the number of incorrectly detected road potholes. Ref. [19] introduces a similar hybrid road pothole detection approach based on RGB-D data (collected by a Microsoft Kinect) analysis. A planar surface is first fitted to the acquired depth image. Similar to Ref. [111], this process is optimized with the RANSAC. A normalized depth-difference image, reflecting the difference between the actual and fitted depth images, is subsequently created and normalized. Otsu's thresholding method is then performed on the normalized depth-difference image to detect road potholes. Recently, Ref. [3] introduced a hybrid road pothole detection algorithm based on 2-D road disparity image transformation and 3-D road point cloud segmentation. A dense subpixel disparity map is first transformed to better distinguish between damaged and undamaged road areas. Otsu's thresholding method is then used to extract potential undamaged road areas from the transformed disparity map. The disparities in the extracted regions are modelled as a quadratic surface using least-squares fitting (also improved with RANSAC). Surface normal information is also integrated into the point cloud modelling process to reduce outliers. Finally, the road potholes are effectively detected by comparing the actual and the modelled disparity maps.

In addition to the approaches discussed above, researchers developed hybrid approaches based on classical 2-D image processing algorithms and machine/deep learning models. Taking Ref. [112] as an example, a naive Bayes classifier (NBC) [117] is trained to learn histograms of oriented gradients (HOG) [118] features. Such HOG features are then utilized to train a road image classifier. Once an image is considered to contain road potholes, it is segmented using the normalized graph cut segmentation (NGCS) [119] algorithm to produce a pixel-level road pothole detection result. Furthermore, Ref. [113] proposes a two-stage road pothole detection approach. In the first stage, the bag of words (BoW) [120] algorithm is utilized to classify road images. This process has four steps: (1) scale-invariant feature transform (SIFT) [121] feature extraction and description, (2) visual vocabulary/codebook construction with K-means clustering, (3) histogram of words generation and (4) road image classification with SVM. In the second stage, the graph cut segmentation (GCS) [119] algorithm is used to segment road images for pixel-level road pothole detection. Recently, Ref. [114] proposed a hybrid road crack and pothole detection algorithm. A modified SegNet [122] is first trained to segment road images for freespace detection. The freespace regions are then processed with a Canny edge detector to generate road crack/pothole candidates. Finally, a SqueezeNet [123] is trained to determine whether the generated candidates are road cracks or potholes.

In recent years, road pothole detection approaches based on 3-D point cloud segmentation and machine/deep learning have also attracted much attention. Ref. [115] is a representative prior art in this field. Ref. [115] compares four existing computer vision techniques for road pothole detection: (1) SV1, a single-frame stereo vision-based method, based on v-disparity image analysis and 3-D plane fitting (in disparity space); (2) SV2, a multi-frame vision sensor data fusion-based method, developed based on the digital elevation model (DEM) and visual odometry; (3) LM1, Mask R-CNN [124] trained with transfer learning; and (4) LM2, YOLOv2 [102] trained with transfer learning. Furthermore, Ref. [116] introduced a hybrid road pothole detection method based on semantic road image segmentation and 3-D road point cloud segmentation. A DeepLabv3+ [107] model is first trained to produce initial pixel-level road pothole detection results. The 3-D points of the initially detected road potholes' edges are classified as exterior and interior ones. The exterior edges are used to fit local planes and calculate road pothole volumes, while the interior edges are used to reduce incorrectly detected potholes by analysing the road depth distribution.

4. Public datasets

This section briefly introduces the existing open-access road pothole detection datasets, which can provide researchers with indications of appropriate datasets when evaluating their developed road pothole detection algorithms.

Ref. [125] created a dataset for road image classification. It consists of a training set and a test set. The training set contains 367 colour images of healthy roads and 357 colour images of roads with potholes; the test set contains eight colour images of each category. This dataset is available at kaggle.com/virenbr11/pothole-and-plain-rode-images.

Ref. [126] presented a large-scale dataset for instance-level pothole detection. This dataset consists of a training set, a test set and an annotation CSV file. The training set contains 2,658 colour images of healthy roads and 1,119 colour images of roads with potholes. The test set contains 628 colour images. The images (resolution: 2,760 × 3,680 pixels) were captured using a GoPro Hero 3+ camera. This dataset can be accessed at kaggle.com/sovitrath/road-pothole-images-for-pothole-detection.

Ref. [127] created a dataset (image resolution: 720 × 1,280 pixels) of Indian roads, with semantic segmentation annotations (road, pothole, footpath, shallow path and background). This dataset contains a training set of 2,475 colour images and a test set of 752 colour images. This dataset is available at kaggle.com/eyantraiit/semantic-segmentation-datasets-of-indian-roads.

Ref. [128] created a dataset, referred to as CIMAT Challenging Sequences for Autonomous Driving (CCSAD). It was initially created to develop and test autonomous vehicle perception and navigation algorithms. The CCSAD dataset includes four scenarios: (1) colonial town streets, (2) urban streets, (3) avenues and small roads and (4) a tunnel network. This dataset contains 500 GB of high-resolution stereo images, complemented with inertial measurement unit (IMU) and GPS data. The CCSAD dataset is publicly available at aplicaciones.cimat.mx/Personal/jbhayet/research.

Ref. [84] presented a large-scale road damage dataset, including 9,053 colour road images (resolution: 600 × 600 pixels) collected in Japan. The images (containing 15,435 road damages) were captured using a smartphone mounted on a car under different weather and illumination conditions. This dataset is publicly available at github.com/sekilab/RoadDamageDetector.

Ref. [129] created a dataset of 665 pairs of colour road images and pothole ground truth labels under different road conditions. This dataset can be used for automatic pothole detection and lo-

calization in urban streets. This dataset is publicly available at public.roboflow.com/object-detection/pothole.

Another road pothole detection dataset [130] was created for binary road image classification. It contains 352 undamaged road images and 329 pothole images. This dataset is small and can only be used to test image classification CNNs. It is available at kaggle.com/datasets/atulyakumar98/pothole-detection-dataset.

Ref. [3] published the world's first multi-modal road pothole detection dataset (image resolution: 800 × 1,312 pixels), containing 55 groups of (1) colour images, (2) subpixel disparity images, (3) transformed disparity images and (4) pixel-level pothole annotations. This dataset is publicly available at github.com/ruirangerfan/stereo_pothole_datasets.

Pothole-600 [11] was recently published by the same research group. It also provides two modalities of vision sensor data: (1) colour images and (2) transformed disparity images. The transformed disparity images were obtained by performing the disparity transformation algorithm [50] on dense subpixel disparity images estimated using the stereo matching algorithm introduced in Ref. [21]. The Pothole-600 dataset is available at sites.google.com/view/pothole-600.

5. Existing challenges and future trends

Before the deep learning boom in 2012, classical 2-D image processing-based approaches dominated this research field. Such explicit programming approaches are, however, usually computationally intensive and sensitive to various environmental factors, most notably illumination and weather conditions [19]. Furthermore, road potholes have irregular shapes, making the geometric assumptions made in such approaches occasionally infeasible. Therefore, since 2013, 3-D point cloud modelling and segmentation-based approaches have emerged to boost the road pothole detection accuracy [34]. Nevertheless, such approaches generally require a small field of view because of the assumption that a single-frame 3-D road point cloud is a planar or quadratic surface. Although significant efforts have been made to further improve the robustness of road point cloud modelling, such as using the RANSAC algorithm [3], extensive parameters are required to ensure the satisfactory performance of these approaches, making them highly challenging to adapt to a new scenario.

Over the past five years, DCNNs have been widely used to solve this problem. Image classification networks can only determine whether a road image contains potholes. Object detection networks can only provide instance-level road pothole detection results. Since the transportation departments are more concerned about potholes' geometric properties, such as width, depth, volume, etc., developing hybrid approaches that combine 3-D road geometry reconstruction and semantic segmentation is the future trend of this research.

Recent deep stereo matching networks have demonstrated superior performance. We believe that they can be easily applied to reconstruct 3-D road geometry models through transfer learning. However, such (supervised) approaches typically require a large amount of well-labelled training data to learn stereo matching, making them often hard to implement in practice [131]. Therefore, un/self-supervised stereo matching algorithms, specifically developed for road surface 3-D reconstruction, are also a popular research area that requires more attention. Furthermore, as stated in Refs. [105,106,108,109], data-fusion semantic segmentation is currently a hot topic in driving scene understanding. However, such networks are generally computationally complicated. After extensive literature investigation, we believe that network pruning and knowledge distillation can be feasible solutions to this problem. In practical experiments, we can also apply a well-trained image classification DCNN to select keyframes (the road images that potentially contain potholes), significantly avoiding the redundant computations of semantic segmentation. Road potholes are not necessarily ubiquitous, and it is challenging to prepare a large, well-annotated dataset to train semantic segmentation DCNNs. Therefore, developing few/low-shot semantic segmentation networks for road pothole detection is also a popular area of research that requires more attention.

6. Conclusions

This article comprehensively reviewed the SoTA road imaging techniques and computer vision algorithms developed for road pothole detection. Classical 2-D image processing-based and 3-D point cloud modelling and segmentation-based approaches have serious limitations. Hence, this article mainly discussed the well-performing SoTA DCNNs, developed for road pothole detection. Since transportation departments are more interested in the geometric properties of potholes, developing hybrid approaches, consisting of stereo matching-based road surface 3-D reconstruction and data-fusion semantic segmentation functionalities, is the future trend of this research. However, training stereo matching and semantic segmentation networks require large human-annotated datasets, and preparing such datasets is exceptionally labour-intensive. Therefore, we believe that un/self-supervised stereo matching algorithms, developed specifically for road surface 3-D, and few/low-shot learning for semantic road image segmentation, are popular areas of research that require more attention.

Acknowledgements

This work was supported by the National Key R&D Program of China (Grant No. 2020AAA0108100), the Fundamental Research Funds for the Central Universities (Grant Nos. 22120220184, 22120220214 and 2022-5-YB-08), and the Shanghai Municipal Science and Technology Major Project (Grant No. 2021SHZDZX0100).

Conflict of interest statement

None declared.

References

1. Mathavan S, Kamal K, Rahman M. A review of three-dimensional imaging technologies for pavement distress detection and measurements. *IEEE Transactions on Intelligent Transportation Systems* 2015;**16** (5): 2353–2362.
2. Miller JS, Bellinger WY et al. Distress identification manual for the long-term pavement performance program. United States. Federal Highway Administration. Office of Infrastructure Research and Development, 2003.
3. Fan R, Ozgunalp U, Hosking B et al. Pothole detection based on disparity transformation and road surface modeling. *IEEE Transactions on Image Processing* 2019;**29**:897–908.
4. Heaton A. Potholes and more potholes: Is it just us? March 2018. URL: shorturl.at/mLN56
5. Fan R, et al. Rethinking Road Surface 3-D Reconstruction and Pothole Detection: From Perspective Transformation to Disparity Map Segmentation. *IEEE Transactions on Cybernetics* 2022;**52**:(7) 5799–5808 doi:10.1109/TCYB.2021.3060461

6. Fan R, Liu M. Road damage detection based on unsupervised disparity map segmentation. *IEEE Transactions on Intelligent Transportation Systems* 2019;**21**(11):4906–4911.
7. Guildford J. Christchurch the pothole capital of New Zealand. February 2018. URL: shorturl.at/ayDP5
8. Devine R. City of San Diego asking residents to report potholes. January 2017. URL: shorturl.at/gnLPV
9. Koch C, Georgieva K, Kasireddy V et al. A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Advanced Engineering Informatics* 2015;**29**(2):196–210.
10. Fan R, et al. Long-Awaited Next-Generation Road Damage Detection and Localization System is Finally Here. In: *2021 29th European Signal Processing Conference (EUSIPCO)*, 2021. p. 641–645. IEEE
11. Fan R, et al. We learn better road pothole detection: from attention aggregation to adversarial domain adaptation. In: *European Conference on Computer Vision Workshops (ECCVW)*. Springer, 2020. p. 285–300.
12. O'Donnell N, McConomy K. Jaquar land rover announces technology research project to detect, predict and share data on potholes'. June 2015. URL: shorturl.at/btKS2
13. Andy R. A billerica company is trying to make potholes less annoying. January 2019. URL: shorturl.at/ktEJM.
14. Koch C, Brilakis I. Pothole detection in asphalt pavement images. *Advanced Engineering Informatics* 2011;**25**(3): 507–515.
15. Chang KT, Chang J, Liu J. Detection of pavement distresses using 3D laser scanning technology. In: *Computing in civil engineering (2005)*. American Society of Civil Engineers (ASCE), 2012, p. 1–11.
16. Lin J, Liu Y. 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*, IEEE, 2010. p. 544–547.
17. Mahler DS, Kharoufa ZB, Wong EK et al. Pavement distress analysis using image processing techniques. *Computer-Aided Civil and Infrastructure Engineering* 1991;**6**(1):1–14.
18. Fan R, Ai X, Dahnoun N. Road surface 3D reconstruction based on dense subpixel disparity map estimation. *IEEE Transactions on Image Processing* 2018;**27**(6):3025–3035.
19. Jahanshahi MR, Jazizadeh F, Masri SF et al. Unsupervised approach for autonomous pavement-defect detection and quantification using an inexpensive depth sensor. *Journal of Computing in Civil Engineering* 2013;**27**(6):743–754.
20. Fan R, et al. Real-time dense stereo embedded in a UAV for road inspection. In: *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, June 2019. p. 535–543.
21. Fan R, et al. Real-time stereo vision for road surface 3-D reconstruction. In: *2018 IEEE International Conference on Imaging Systems and Techniques (IST)*, IEEE, 2018. p. 1–6.
22. Laurent J, Héert JF, Lefebvre D et al. Using 3D laser profiling sensors for the automated measurement of road surface conditions. In: *7th RILEM international conference on cracking in pavements*, Springer, 2012. p. 157–167.
23. Banerjee T, et al. Exploratory analysis of older adults' sedentary behavior in the primary living area using Kinect depth data. *Journal of Ambient Intelligence and Smart Environments* 2017;**9**(2):163–179.
24. StereoLabs ZED stereo camera. April 2022. URL: stereolabs.com/zed
25. Laurent J, et al. Road surface inspection using laser scanners adapted for the high precision 3D measurements of large flat surfaces. In: *International Conference on Recent Advances in 3-D Digital Imaging and Modeling*, 1997. p. 303–310. IEEE doi: 10.1109/IM.1997.603880
26. Fan R, et al. Computer-aided road inspection: Systems and algorithms. In: *Recent Advances in Computer Vision Applications Using Parallel Processing*, Springer 2022. (in press) arXiv:2203.02355.
27. Tsai YC, Chatterjee A. Pothole detection and classification using 3D technology and watershed method. *Journal of Computing in Civil Engineering* 2018;**32**(2):04017078.
28. Moazzam I, Kamal K, Mathavan S et al. Metrology and visualization of potholes using the microsoft kinect sensor. In: *16th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, IEEE 2013. p. 1284–1291.
29. Jog GM, Koch C, Golparvar-Fard M et al. Pothole properties measurement through visual 2D recognition and 3D reconstruction. In: *Computing in Civil Engineering (2012)*. American Society of Civil Engineers (ASCE), y, p. 553–560.
30. Hartley R, Zisserman A. Multiple view geometry in computer vision. Cambridge University Press, 2003.
31. Zhang C, Elaksher A. An unmanned aerial vehicle-based imaging system for 3D measurement of unpaved road surface distresses. *Computer-Aided Civil and Infrastructure Engineering* 2012;**27**(2):118–129.
32. Ullman S. The interpretation of structure from motion. *Proceedings of the Royal Society of London Series B Biological Sciences* 1979;**203**(1153):405–426.
33. Triggs B, et al. Bundle adjustment—a modern synthesis. In: *International workshop on vision algorithms Springer*; 1999. p. 298–372.
34. Zhang Z. Advanced stereo vision disparity calculation and obstacle analysis for intelligent vehicles. PhD thesis, University of Bristol, 2013.
35. Ozgunalp U. Vision based lane detection for intelligent vehicles. PhD thesis, University of Bristol, 2016.
36. Fan R, et al. Computer stereo vision for autonomous driving. *Recent Advances in Computer Vision Applications Using Parallel Processing* 2022; (in press) arXiv:2012.03194
37. Fan R, et al. Graph attention layer evolves semantic segmentation for road pothole detection: A benchmark and algorithms. *IEEE Transactions on Image Processing* 2021;**30**: 8144–8154.
38. Buza E, Omanovic S, Huseinovic A. Pothole detection with image processing and spectral clustering. In: *Proceedings of the 2nd International Conference on Information Technology and Computer Networks*, vol. **810**, 2013. p. 4853.
39. Ryu SK, Kim T, Kim YR. Feature-based pothole detection in two-dimensional images. *Transportation Research Record* 2015;**2528**(1):9–17.
40. Schiopu I, Saarinen JP, Kettunen L. Pothole detection and tracking in car video sequence. In: *2016 39th International Conference on Telecommunications and Signal Processing (TSP)*, IEEE, 2016. p. 701–706.
41. Jakštys V, Marcinkevicius V, Treigys P et al. Detection of the Road Pothole Contour in Raster Images. *Information Technology and Control* 2016;**45**(3):300–307.
42. Akagic A, Buza E, Omanovic S. Pothole detection: An efficient vision based method using rgb color space image segmentation. In: *2017 40th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, IEEE,2017. p. 1104–1109.

43. Wang P, Hu Y, Dai Y et al. Asphalt pavement pothole detection and segmentation based on wavelet energy field. *Mathematical Problems in Engineering* 2017;**2017**.
44. Chung TD, Khan MA. Watershed-based real-time image processing for multi-potholes detection on asphalt road. In: *2019 IEEE 9th International Conference on System Engineering and Technology (ICSET) IEEE*, 2019. p. 268–272.
45. Kim T, Ryu SK. System and method for detecting potholes based on video data. *Journal of Emerging Trends in Computing and Information Sciences* 2014;**5**(9):703–709.
46. Fan R et al. Road crack detection using deep convolutional neural network and adaptive thresholding. In: *2019 IEEE Intelligent Vehicles Symposium (IV)*, IEEE,2019. p. 474–479.
47. Pitas I. *Digital image processing algorithms and applications*. John Wiley & Sons, 2000.
48. Leung T, Malik J. Representing and recognizing the visual appearance of materials using three-dimensional textons. *International journal of computer vision* 2001;**43**(1):29–44.
49. Schmid C. Constructing models for content-based image retrieval. In: *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. **2** IEEE, 2001. p. II–II.
50. Fan R, et al. A novel disparity transformation algorithm for road segmentation. *Information Processing Letters* 2018;**140**: 18–24.
51. Pierre DA. *Optimization theory with applications*. Courier Corporation, 1986.
52. Ozgunalp U et al. Multiple lane detection algorithm based on novel dense vanishing point estimation. *IEEE Transactions on Intelligent Transportation Systems* 2016;**18**(3):621–632.
53. Otsu N. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics* 1979;**9**(1):62–66.
54. Achanta R, Shaji A, Smith K et al. SLIC superpixels compared to state-of-the-art superpixel methods. *IEEE transactions on pattern analysis and machine intelligence* 2012;**34**(11): 2274–2282.
55. Li Y, Papachristou C, Weyer D. Road pothole detection system based on stereo vision. In: *NAECON 2018-IEEE National Aerospace and Electronics Conference*, IEEE,2018. p. 292–297.
56. Du Y, et al. A pothole detection method based on 3D point cloud segmentation. In: Twelfth International Conference on Digital Image Processing (ICDIP 2020), **vol. 11519** International Society for Optics and Photonics; 2020. p. 1151909.
57. Daniel A, Preeja V. Automatic road distress detection and analysis. *International Journal of Computer Applications* 2014;**101** (10).
58. Hadjidemetriou GM, Christodoulou SE, Vela PA. Automated detection of pavement patches utilizing support vector machine classification. In: *2016 18th Mediterranean Electrotechnical Conference (MELECON)*, IEEE,2016. p. 1–5.
59. Ahmed N, Natarajan T, Rao KR. Discrete cosine transform. *IEEE transactions on Computers* 1974;**100**(1):90–93.
60. Haralick RM, Shanmugam K, Dinstein IH. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics* 1973;(**6**):610–621.
61. Hoang ND. An artificial intelligence method for asphalt pavement pothole detection using least squares support vector machine and neural network with steerable filter-based feature extraction. *Advances in Civil Engineering* 2018;**2018**.
62. Pan Y, Zhang X, Cervone G et al. Detection of asphalt pavement potholes and cracks based on the unmanned aerial vehicle multispectral imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2018;**11**(10): 3701–3712.
63. Gao M, Wang X, Zhu S et al. Detection and segmentation of cement concrete pavement pothole based on image processing technology. *Mathematical Problems in Engineering* 2020;**2020**.doi: 10.1155/2020/1360832
64. Pereira V, Tamura S, Hayamizu S et al. A deep learning-based approach for road pothole detection in timor leste. In: *2018 IEEE International Conference on Service Operations and Logistics, and Informatics (SOLI) IEEE*, 2018. p. 279–284.
65. An KE, et al. Detecting a pothole using deep convolutional neural network models for an adaptive shock observing in a vehicle driving. In: *2018 IEEE International Conference on Consumer Electronics (ICCE) IEEE*, 2018. p. 1–2.
66. Ye W, Jiang W, Tong Z et al. Convolutional neural network for pothole detection in asphalt pavement. *Road materials and pavement design* 2021;**22**(1):42–58.
67. Bhatia Y, Rai R, Gupta V et al. Convolutional neural networks based potholes detection using thermal imaging. *Journal of King Saud University-Computer and Information Sciences* 2019.
68. Wu R, et al. Scale-Adaptive Road Pothole Detection and Tracking from 3D Point Clouds. In: *2021 IEEE International Conference on Imaging Systems and Techniques (IST)*, IEEE,2021.
69. Hartigan JA, Wong MA. Algorithm AS 136: A k-means clustering algorithm. *Journal of the royal statistical society series c (applied statistics)* 1979;**28**(1):100–108.
70. Fischler MA, Bolles RC. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM* 1981;**24**(6):381–395.
71. Li Q, Yao M, Yao X et al. A real-time 3D scanning system for pavement distortion inspection. *Measurement Science and Technology* 2009;**21**(1):015702.
72. Ravi R, Bullock D, Habib A. Highway and Airport Runway Pavement Inspection Using Mobile LiDAR. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 2020;**43**:349–354.
73. LeCun Y, Bengio Y, Hinton G. Deep learning. *nature* 2015;**521**(7553):436–444.
74. Cortes C, Vapnik V. Support-vector networks. *Machine learning* 1995;**20**(3):273–297.
75. Pan Y, Zhang X, Sun M et al. Object-based and supervised detection of potholes and cracks from the pavement images acquired by UAV. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences* 2017; **42**.
76. He K, Zhang X, Ren S et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016. p. 770–778.
77. Ryu SK, Kim T, Kim YR. Image-based pothole detection system for ITS service and road management system. *Mathematical Problems in Engineering* 2015;**2015**.
78. Szegedy C, Ioffe S, Vanhoucke V et al. Inception-v4, inception-resnet and the impact of residual connections on learning. In: *Thirty-first AAAI conference on artificial intelligence*, 2017.
79. He K, Zhang X, Ren S et al. Identity mappings in deep residual networks. In: *European conference on computer vision Springer*, 2016. p. 630–645.
80. Howard AG, Zhu M, Chen B et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:170404861 2017.

81. Fan J, et al. Deep convolutional neural networks for road crack detection: Qualitative and quantitative comparisons. In: *2021 IEEE International Conference on Imaging Systems and Techniques (IST) IEEE*, 2021. p. 1–6.
82. Liu W, Anguelov , Erhan D et al. Ssd: Single shot multibox detector. In: *European conference on computer vision Springer*, 2016. p. 21–37.
83. Szegedy C, Vanhoucke V, Ioffe S et al. Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016. p. 2818–2826.
84. Maeda H, Sekimoto Y, Seto T et al. Road damage detection using deep neural networks with images captured through a smartphone. *CoRR* 2018.
85. Lin TY, Goyal P, Girshick R et al. Focal loss for dense object detection. In: *Proceedings of the IEEE international conference on computer vision*, 2017. p. 2980–2988.
86. Gupta S, Sharma P, Sharma D et al. Detection and localization of potholes in thermal images using deep neural networks. *Multimedia Tools and Applications* 2020;**79**(35):26265–26284.
87. Suong LK, Kwon J. Detection of Potholes Using a Deep Convolutional Neural Network. *J Univers Comput Sci* 2018;**24**(9):1244–1257.
88. Wang W, Wu B, Yang S et al. Road damage detection and classification with faster r-cnn. In: *2018 IEEE international conference on big data (Big data)*, IEEE,2018. p. 5220–5223.
89. Ukhwah EN, Yuniarno EM, Suprapto YK. Asphalt pavement pothole detection using deep learning method based on YOLO neural network. In: *2019 International Seminar on Intelligent Technology and Its Applications (ISITIA) IEEE*, 2019. p. 35–40.
90. Dharneeshkar J, Aniruthan S, Karthika R et al. Deep Learning based Detection of potholes in Indian roads using YOLO. In: *2020 International Conference on Inventive Computation Technologies (ICICT) IEEE*, 2020. p. 381–385.
91. Baek JW, Chung K. Pothole classification model using edge detection in road image. *Applied Sciences* 2020;**10**(19):6662.
92. Kortmann F, Talits K, Fassmeyer P et al. Detecting Various Road Damage Types in Global Countries Utilizing Faster R-CNN. In: *2020 IEEE International Conference on Big Data (Big Data)*; IEEE,2020. p. 5563–5571.
93. Yebes JJ, Montero D, Arriola I. Learning to Automatically Catch Potholes in Worldwide Road Scene Images. *IEEE Intelligent Transportation Systems Magazine* 2020;**13**(3):192–205.
94. Javed A, Mahmud MS, Alam MT et al. Pothole Detection System Using Region-Based Convolutional Neural Network. In: *2021 IEEE 4th International Conference on Computer and Communication Engineering Technology (CCET)*, IEEE,2021. p. 6–11.
95. Sandler M, Howard A, Zhu M et al. Mobilenetv2: Inverted residuals and linear bottlenecks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018. p. 4510–4520.
96. Ren S, He K, Girshick R et al. Faster R-CNN: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* 2015;**28**:91–99.
97. Pereira V, Tamura S, Hayamizu S et al. Semantic segmentation of paved road and pothole image using U-net architecture. In: *2019 International Conference of Advanced Informatics: Concepts, Theory and Applications (ICAICTA) IEEE*, 2019. p. 1–4.
98. Chun C, Ryu SK. Road surface damage detection using fully convolutional neural networks and semi-supervised learning. *Sensors* 2019;**19**(24):5501.
99. Masihullah S, Garg R, Mukherjee P, et al. Attention Based Coupled Framework for Road and Pothole Segmentation. In: *2020 25th International Conference on Pattern Recognition (ICPR)*, IEEE,2021. p. 5812–5819.
100. Fan J, et al. Multi-Scale Feature Fusion: Learning Better Semantic Segmentation for Road Pothole Detection. In: *2021 IEEE International Conference on Autonomous Systems (ICAS)*, IEEE,2021. p. 1–5
101. Redmon J, Farhadi A. Yolov3: An incremental improvement. *CoRR* 2018.
102. Redmon J, Farhadi A. YOLO9000: better, faster, stronger. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, IEEE,2017. p. 7263–7271.
103. Redmon J, Divvala S, Girshick R et al. You only look once: Unified, real-time object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016. p. 779–788.
104. Hazirbas C, Ma L, Domokos C, et al. Fusenet: Incorporating depth into semantic segmentation via fusion-based cnn architecture. In: *Asian conference on computer vision Springer*, 2016. p. 213–228.
105. Fan R, Wang H, Cai P, et al. SNE-Roadseg: Incorporating surface normal information into semantic segmentation for accurate freespace detection. In: *European Conference on Computer Vision (ECCV)*, Springer,2020. p. 340–356.
106. Wang H, et al. SNE-RoadSeg+: Rethinking Depth-Normal Translation and Deep Supervision for Freespace Detection. In: *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE,2021. p. 1140–1145.
107. Chen LC, Zhu Y, Papandreou G et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. In: *Proceedings of the European conference on computer vision (ECCV)*; 2018. p. 801–818.
108. Wang H, Fan R, Sun Y et al. Applying surface normal information in drivable area and road anomaly detection for ground mobile robots. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2020. p. 2706–2711.
109. Wang H, et al. Dynamic fusion module evolves drivable area and road anomaly detection: A benchmark and algorithms. IEEE Transactions on Cybernetics, 2021;doi: 10.1109/TCYB.2021.3064089
110. Fan R, et al. Learning collision-free space detection from stereo images: Homography matrix brings better data augmentation. *IEEE/ASME Transactions on Mechatronics* 2022;**27**(1):225–233.
111. Joubert D, Tyatyantsi A, Mphahlehle J et al. *Pothole tagging system, the Robotics and Mechatronics Conference of South Africa, CSIR International Conference Centre, Pretoria*, 2011. p. 23–25.
112. Azhar K, Murtaza F, Yousaf MH et al. Computer vision based detection and localization of potholes in asphalt pavement images. In: *2016 IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)*, IEEE,2016. p. 1–5.
113. Yousaf MH, Azhar K, Murtaza F et al. Visual analysis of asphalt pavement for detection and localization of potholes. *Advanced Engineering Informatics* 2018;**38**:527–537.
114. Anand S, Gupta S, Darbari V et al. Crack-pot: Autonomous road crack and pothole detection. In: *2018 Digital Image Computing: Techniques and Applications (DICTA) IEEE*, 2018. p. 1–6.
115. Dhiman A, Klette R. Pothole detection using computer vision and learning. *IEEE Transactions on Intelligent Transportation Systems* 2019;**21**(8):3536–3550.

116. Wu H, Yao L, Xu Z et al. Road pothole extraction and safety evaluation by integration of point cloud and images derived from mobile mapping sensors. *Advanced Engineering Informatics* 2019;**42**:100936.
117. Rish I, et al. An empirical study of the naive Bayes classifier. In: *IJCAI 2001 workshop on empirical methods in artificial intelligence*, vol. **3**, 2001. p. 41–46.
118. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. **1** Ieee, 2005. p. 886–93.
119. Shi J, Malik J. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence* 2000;**22**(8):888–905.
120. Csurka G, Dance C, Fan L et al. *Visual categorization with bags of keypoints. In: Workshop on statistical learning in computer vision, ECCV, vol. 1 Prague*, 2004. p. 1–2.
121. Lowe DG. Distinctive image features from scale-invariant keypoints. *International journal of computer vision* 2004;**60**(2):91–110.
122. Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence* 2017;**39**(12):2481–2495.
123. Iandola FN, Han S, Moskewicz MW et al. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. *CoRR* 2016.
124. He K, Gkioxari G, Dollár P et al. Mask r-cnn. In: *Proceedings of the IEEE international conference on computer vision*, 2017. p. 2961–2969.
125. Viren, Pothole and Plain Road Images. December 2019. URL: shorturl.at/gqvKU.
126. Rath SR. Road Pothole Images for Pothole detection;. September 2020. URL: shorturl.at/sxKUX.
127. Bombay YI, Semantic Segmentation datasets of Indian roads;. November 2021. URL: shorturl.at/coyzB.
128. Guzmán R, Hayet JB, Klette R. Towards ubiquitous autonomous driving: The CCSAD dataset. In: *International Conference on Computer Analysis of Images and Patterns Springer*, 2015. p. 582–593.
129. Chitholian RA. *Pothole dataset*;. November 2020. URL: shorturl.at/uzY16.
130. Kumar A. *Pothole Detection Dataset*;. November 2019. URL: shorturl.at/blBJK.
131. Wang H, et al. PVStereo: Pyramid voting module for end-to-end self-supervised stereo matching. *IEEE Robotics and Automation Letters* 2021;**6**(3):4353–4360.