Computer-Aided Road Inspection: Systems and Algorithms

Rui Fan, Sicen Guo, Li Wang, and Mohammud Junaid Bocus

Abstract Road damage is an inconvenience and a safety hazard, severely affecting vehicle condition, driving comfort, and traffic safety. The traditional manual visual road inspection process is pricey, dangerous, exhausting, and cumbersome. Also, manual road inspection results are qualitative and subjective, as they depend entirely on the inspector's personal experience. Therefore, there is an ever-increasing need for automated road inspection systems. This chapter first compares the five most common road damage types. Then, 2-D/3-D road imaging systems are discussed. Finally, state-of-the-art machine vision and intelligence-based road damage detection algorithms are introduced.

1 Introduction

The condition assessment of concrete and asphalt civil infrastructures (e.g., tunnels, bridges, and pavements) is essential to ensure their serviceability while still providing maximum safety for the users [1]. It also allows the government to allocate limited resources for infrastructure maintenance and appraise long-term investment schemes [2]. The detection and reparation of road damages (e.g., potholes, and cracks) is a crucial civil infrastructure maintenance task because they are not only an inconve-

R. Fan (✉) · S. Guo
Tongji University, Shanghai, China
e-mail: rfan@tongji.edu.cn

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

L. Wang
Continental Automotive Systems Shanghai Co. Ltd., Shanghai, China
e-mail: li.15.wang@continental-corporation.com

M. Junaid Bocus
University of Bristol, Bristol, England
e-mail: junaid.bocus@bristol.ac.uk

K. M. Hosny and A. Salah (eds.), *Recent Advances in Computer Vision Applications Using Parallel Processing*, Studies in Computational Intelligence 1073,
https://doi.org/10.1007/978-3-031-18735-3_2

nience but also a safety hazard, severely affecting vehicle condition, driving comfort, and traffic safety [3].

Traditionally, road damages are regularly inspected (i.e., detected and localized) by certified inspectors or structural engineers [4]. However, this manual visual inspection process is time-consuming, costly, exhausting, and dangerous [5]. Moreover, manual visual inspection results are qualitative and subjective, as they depend entirely on the individual's personal experience [6]. Therefore, there is an ever-increasing need for automated road inspection systems, developed based on cutting-edge machine vision and intelligence techniques [7].

A computer-aided road inspection system typically consists of two major procedures [8]: (1) road data acquisition, and (2) road damage detection (i.e., recognition/segmentation and localization). The former typically employs active or passive sensors (e.g., laser scanners [9], infrared cameras [10], and digital cameras [11]) to acquire road texture and/or spatial geometry, while the latter commonly uses 2-D image analysis/understanding algorithms, such as image classification [12], semantic segmentation [13], object detection [14], and/or 3-D road surface modeling algorithms [3] to detect the damaged road areas.

This chapter first compares the most common types of road damages, including crack, spalling, pothole, rutting, and shoving. Then, various technologies employed to acquire 2-D/3-D road data are discussed. Finally, state-of-the-art (SOTA) machine vision and intelligence approaches, including 2-D image analysis/understanding algorithms and 3-D road surface modeling algorithms, developed to detect road damages are introduced.

2 Road Damage Types

Crack, spalling, pothole, rutting, and shoving are five common types of road damages [6]. Their structures are illustrated in Fig. 1. Road crack has a much larger depth when compared to its dimensions on the road surface, presenting a unique challenge to the imaging systems; Road spalling has similar lateral and depth magnitudes, and thus, the imaging systems designed specifically to measure this type of road damages should perform similarly in both lateral and depth directions; Road pothole is a considerably large structural road surface failure, measurable by an imaging setup with a large field of view; Road rutting is extremely shallow along its depth, requiring a measurement system with high accuracy in the depth direction; Road shoving refers to a small bump on the road surface, which makes it difficult to profile with some imaging systems.

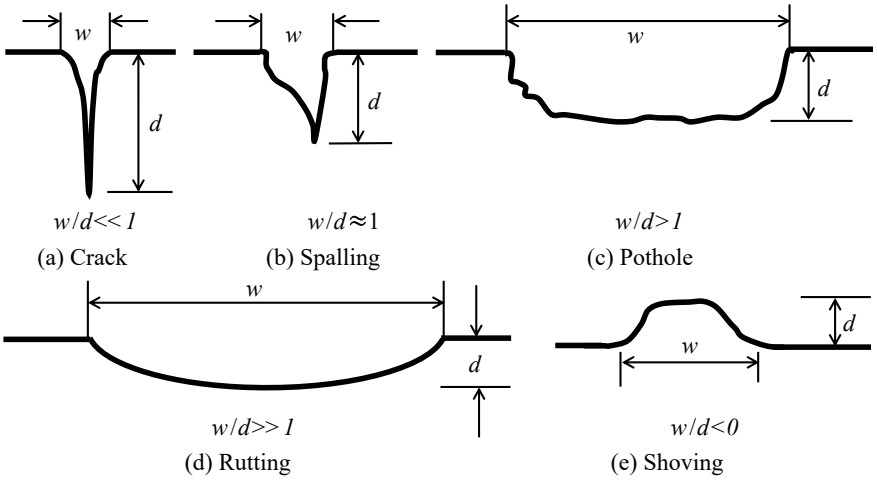

Fig. 1 Five common types of road damages, where w and d represent the lateral and depth magnitudes, respectively

3 Road Data Acquisition

3.1 Sensors

The use of 2-D imaging technologies (i.e., digital imaging or digital image acquisition) for road data collection started as early as 1991 [15, 16]. However, the spatial structure cannot always be explicitly illustrated in 2-D road images [1]. Moreover, the image segmentation algorithms performing on either gray-scale or color road images can be severely affected by various factors, most notably by poor illumination conditions [10]. Therefore, many researchers [1, 7, 11] have resorted to 3-D imaging technologies, which are more feasible to overcome the disadvantages mentioned above and simultaneously provide an enhancement of road damage detection accuracy.

The first reported effort [17] on leveraging 3-D imaging technology for road data collection can be dated back to 1997. This section discusses the existing 3-D imaging technologies applied for road data acquisition.

Laser scanning is a well-established imaging technology for accurate 3-D road geometry information acquisition [6]. Laser scanners are developed based on trigonometric triangulation, as illustrated in Fig. 2. The sensor is located at a known distance from the laser's source, and therefore, accurate point measurements can be made by calculating the reflection angle of the laser light. Auto-synchronized triangulation [17] is a popular variation of classic trigonometric triangulation and has been widely utilized in laser scanners to capture the 3-D geometry information of near-flat road surfaces [6]. However, laser scanners have to be mounted on dedicated road inspec-

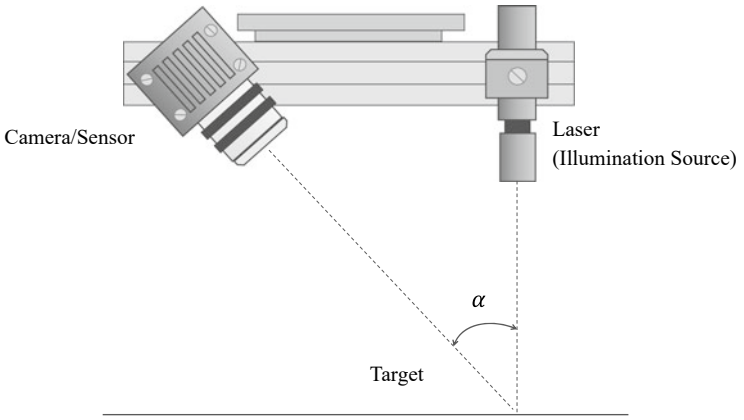

Fig. 2 Laser triangulation. The camera/sensor is located at a known distance from the laser's illumination source. Therefore, accurate point measurements can be made by calculating the reflection angle of the laser light

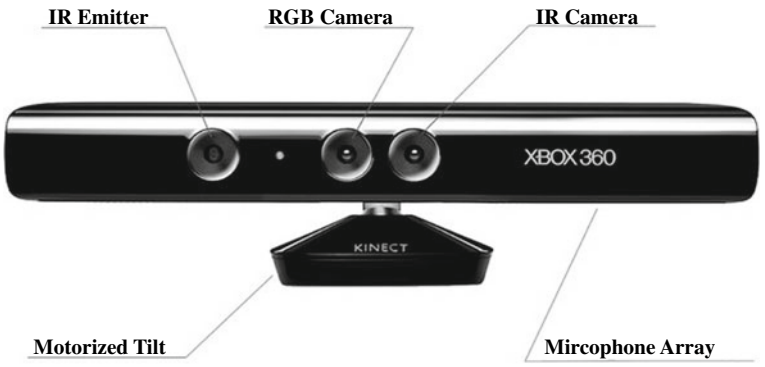

Fig. 3 Microsoft Kinect sensor [20]

tion vehicles, such as the Georgia Institute of Technology Sensing Vehicle [9], for 3-D road data collection, and such vehicles are not widely used because they involve high-end equipments and their long-term maintenance can be costly as well [7].

Microsoft Kinect sensors [10], as illustrated in Fig. 3, were initially designed for the Xbox-360 motion sensing games. Microsoft Kinect sensors are typically equipped with an RGB camera, an infrared (IR) sensor/camera, an IR emitter, microphones, accelerometers, and a tilt motor for motion tracking facility [6]. The working range of Microsoft Kinect sensors is 800–4000 mm, making it suitable for road imaging when mounted on a vehicle. There are three reported efforts on 3-D road data acquisition and damage detection using Microsoft Kinect sensors [10, 18, 19]. However, it is also reported that Microsoft Kinect sensors greatly suffer from IR saturation in direct sunlight [3].

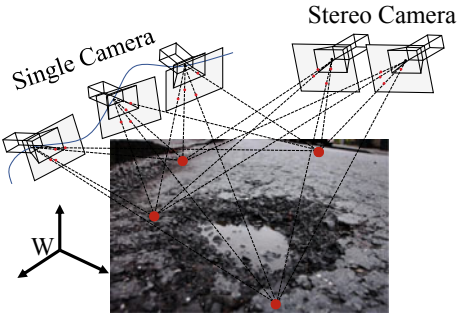

Fig. 4 3-D road imaging with camera(s) [27]

The 3-D geometry of a road surface can also be reconstructed using multiple images captured from different views [1]. The theory behind this 3-D imaging technique is typically known as *multi-view geometry* [21]. These images can be captured using a single movable camera [22] or an array of synchronized cameras [11]. In both cases, dense correspondence matching (DCM) between two consecutive road video frames or between two synchronized stereo road images is the key problem to be solved, as illustrated in Fig. 4. Structure from motion (SfM) [23] and optical flow (OF) estimation [24] are the two most commonly used techniques for monocular DCM, while stereo vision (also known as binocular vision, stereo matching, or disparity estimation) [1, 7] is typically employed for binocular DCM. SfM methods estimate both camera poses and the 3-D points of interest from road images captured from multiple views, linked by a collection of visual features. They also leverage bundle adjustment (BA) [25] algorithm to refine the estimated camera poses and 3-D point locations by minimizing a cost function known as total re-projection error [26]. OF describes the motion of pixels between consecutive frames of a video sequence. Stereo vision acquires depth information by finding the horizontal positional differences (disparities) of the visual feature correspondence pairs between two synchronously captured road images [1]. The traditional stereo vision algorithms formulate disparity estimation as either a local block matching problem or a global/semi-global energy minimization problem (solvable with various Markov random field-based optimization techniques) [28, 29], while data-driven algorithms typically solve stereo matching with convolutional neural networks (CNNs) [30]. Despite the low cost of digital cameras, DCM accuracy is always affected by various factors, most notably by poor illumination conditions [3]. Furthermore, it is always tricky for monocular DCM algorithms to recover the absolute scale of a 3-D road geometry model without considering the complicated environmental hypotheses [8]. On the other hand, the feasibility of binocular DCM algorithms relies on the well-conducted stereo rig calibration [1], and therefore, some systems incorporated stereo rig self-calibration functionalities to ensure that their captured stereo road images are always well-rectified.

In addition to the aforementioned three common types of 3-D imaging technologies, shape from focus (SFF) [31–33], shape from defocus (SFDF) [34–36], shape from shading (SFS) [37], photometric stereo [38, 39], interferometry [40, 41], structured light imaging [42, 43], and time-of-flight (ToF) [44, 45] are other alternatives for 3-D geometry reconstruction. However, they were not widely used for 3-D road data acquisition. The interested reader is referred to [6] for a detailed description of these technologies (both theoretical and practical aspects are covered).

3.2 Public Datasets

Road damage detection datasets are mainly created for crack or pothole detection. This section details the commonly used crack and pothole detection datasets.

3.2.1 Crack Detection Datasets

The Middle East Technical University (METU) dataset [46] was created using the method proposed in [12] for image classification. The original road images (resolution: 4032×3024 pixels) were taken at the METU. It is divided into two classes: negative and positive. Each class has 20,000 cropped road images (resolution: 227×227 pixels). This dataset is publicly available at https://data.mendeley.com/datasets/5y9wdsg2zt/2.

SDNET2018 [47] is another large-scale road crack detection dataset containing over 56,000 concrete road images (resolution: 256×256 pixels) of bridge decks, walls, and pavements, with crack length ranging from 0.06 to 25 mm. Because of its diversity in crack size and width, it is commonly used for network evaluation. This dataset can be accessed at https://digitalcommons.usu.edu/alldatasets/48.

3.2.2 Pothole Detection Datasets

Bristol pothole detection dataset [48] was collected using two PointGrey Flea 3 cameras, mounted in parallel onto the optical rail, with an Arduino synchronization board in the middle. This produces 20 Hz clock signal to synchronize the cameras. The lenses on the cameras are wide-angle lenses with 2.8 mm focal length and F/1.2 aperture. They have a 93.2°×70.7° field of view, with the forward point setup. A threshold of 0.04 m was set to detect the pothole point cloud.

[3] provides the world-first multi-modal road pothole detection dataset containing 67 groups of RGB images, disparity images, transformed disparity images, pixel-level annotations, and 3-D point clouds. This dataset can be used to design and evaluate semantic segmentation networks. This dataset is publicly available at https://github.com/ruirangerfan/stereo_pothole_datasets.

The Pothole-600 dataset [5] was created using a ZED stereo camera. It contains 600 collections of RGB images, transformed disparity images, and pixel-level annotations. This dataset can be employed to train both single-modal and data-fusion semantic segmentation networks, as illustrated in Fig. 10. It is available at https://sites.google.com/view/pothole-600/.

Cranfield PotDataset [49] was created using a DFK 33UX264 colour camera, collecting road images (resolution: 1024×768) at 30 fps. This dataset contains 1610 pothole images, 1664 manhole images, 2703 asphalt pavement images, 2365 road marking images, and 2533 shadow images. This dataset can be used to train and evaluate image classification networks. This dataset is publicly available at https://cord.cranfield.ac.uk/articles/dataset/PotDataset/5999699/1.

4 Road Damage Detection

The SOTA road damage detection approaches are developed based on either 2-D image analysis/understanding or 3-D road surface modeling. The former methods typically utilize traditional image processing algorithms [10, 50] or modern CNNs [5, 14] to detect road damage, by performing either pixel-level image segmentation or instance-level object recognition on RGB or depth/disparity images. The latter approaches typically formulate the 3-D road point cloud as a planar/quadratic surface [3, 48, 51], whose coefficients can be obtained by performing robust surface modeling. The damaged road areas can be detected by comparing the difference between the actual and modeled road surfaces.

4.1 2-D Image Analysis/Understanding-Based Approaches

4.1.1 Traditional Image Analysis-Based Approaches

An overview of the traditional image analysis-based road damage detection approaches is illustrated in Fig. 5. These approaches typically consist of four main procedures: (1) image pre-processing, (2) image segmentation, (3) shape extraction, and (4) object recognition [52]. [53] is a classic example of this algorithm type. It first uses a histogram-based thresholding algorithm [54] to segment (binarize) the input gray-scale road images. The segmented road images are then processed with a median filter and a morphology operation to further reduce the redundant noise. Finally, the damaged road areas are extracted by analyzing a pixel intensity histogram. Additionally, [54] solved road pothole detection with similar techniques. It first applies the triangle algorithm [55] to segment gray-scale road images. An ellipse then models the boundary of an extracted potential road pothole area. Finally, the image texture within the ellipse is compared with the undamaged road area texture. If the former is coarser than the latter, the ellipse is considered to be a pothole.

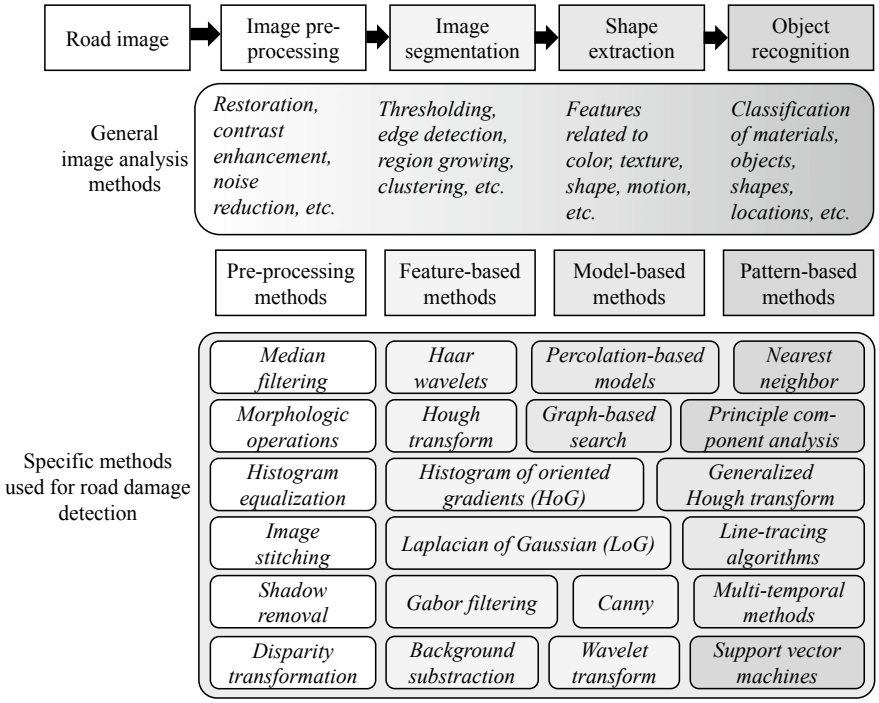

Fig. 5 An overview of the traditional image analysis-based road damage detection approaches

However, both color and gray-scale image segmentation techniques are severely affected by various factors, most notably by poor illumination conditions [3]. Therefore, some researchers performed segmentation on depth/disparity images, which has shown to achieve better performance in terms of separating damaged and undamaged road areas [9, 10, 50]. For instance, [10] acquires the depth images of pavements using a Microsoft Kinect sensor. The depth images are then segmented using wavelet transform algorithm [56]. For effective operation, the Microsoft Kinect sensor has to be mounted perpendicularly to the pavements. In 2018, [9] proposed to detect road potholes from depth images acquired by a highly accurate laser scanner mounted on a Georgia Institute of Technology sensing vehicle. The depth images are first processed with a high-pass filter so that the depth values of the undamaged road pixels become similar. The processed depth image is then segmented using watershed method [57] for road pothole detection.

Since the concept of "v-disparity map" was presented in [58], disparity image segmentation has become a common algorithm used for object detection [50]. In 2018, [3] extends the traditional v-disparity representation to a more general form: the projections of the on-road disparity (or inverse depth, because depth is in inverse proportion to disparity) pixels in the v-disparity map can be represented by a nonlinear model as follows [7]:

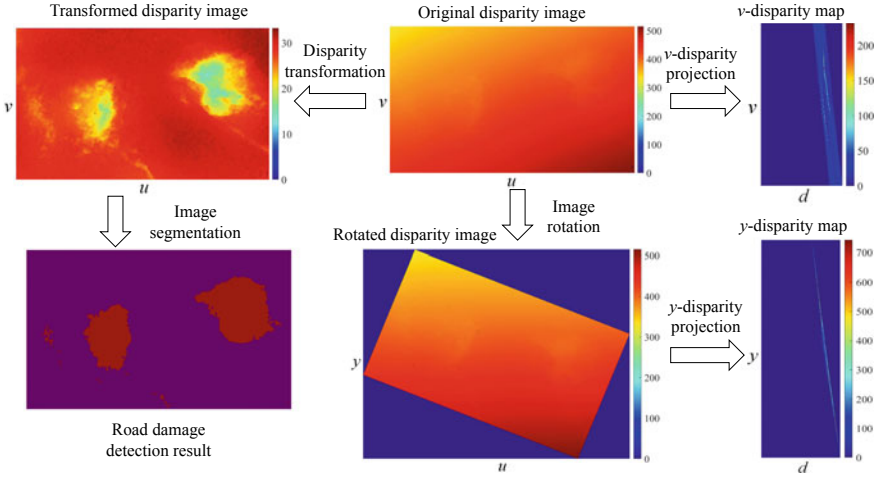

Fig. 6 Disparity transformation. The stereo rig roll angle Φ is non-zero, and therefore, the disparity values in the original disparity image change gradually in the horizontal direction. By rotating the original disparity image by Φ, the disparity values on each row become more uniform. Using Φ and $\mathbf{K}$, the original disparity image can be transformed, where the damaged road areas become highly distinguishable. Finally, the transformed disparity image can be segmented using common image segmentation algorithms for pixel-level road damage detection

$$\tilde{\mathbf{q}} = \mathbf{M}\tilde{\mathbf{p}} = \varkappa \begin{bmatrix} -\sin\Phi & \cos\Phi & \kappa \\ 0 & 1/\varkappa & 0 \\ 0 & 0 & 1/\varkappa \end{bmatrix} \tilde{\mathbf{p}}, \tag{1}$$

where $\tilde{\mathbf{p}} = (u; v; 1)$ is the homogeneous coordinates of a pixel in the disparity (or inverse depth) image, $\tilde{\mathbf{q}} = (g; v; 1)$ is the homogeneous coordinates of its projection in the v-disparity map, Φ is the stereo rig roll angle, and $\mathbf{K} = (\kappa; \varkappa)$ stores two road disparity projection model coefficients. Additionally, [3] also introduces a novel disparity image processing algorithm, referred to as disparity transformation. This algorithm can not only estimate Φ and $\mathbf{K}$ but also correct the disparity projections on the v-disparity image. As shown in Fig. 6, the original disparity image can then be transformed to better distinguish between damaged and undamaged road areas, making road damage detection a much easier task. The disparity transformation is achieved using Φ and $\mathbf{K}$, which was estimated using golden section search [59] and dynamic programming [60, 61], respectively. Later on, in [11], the authors introduced a more efficient solution to Φ, as follows:

$$\arg\min_{\Phi} \mathbf{g}^\top\mathbf{g} - \mathbf{g}^\top\mathbf{T}(\Phi)\big(\mathbf{T}(\Phi)^\top\mathbf{T}(\Phi)\big)^{-1}\mathbf{T}(\Phi)^\top\mathbf{g}, \tag{2}$$

where $\mathbf{g}$ is a k-entry vector of disparity (or inverse depth) values, $\mathbf{1}_k$ is a k-entry vector of ones, $\mathbf{u}$ and $\mathbf{v}$ are two k-entry vectors storing the horizontal and vertical coordinates of the observed pixels, respectively, and $\mathbf{T}(\Phi) = [\mathbf{1}_k, \cos\Phi\mathbf{v} - \sin\Phi\mathbf{u}]$.

Equation (2) was solved using gradient descent algorithm [11]. In 2019, [50] proves that (2) has a closed-form solution[1] as follows [50]:

$$\Phi = \arctan \frac{\omega_4\omega_0 - \omega_3\omega_1 + q\sqrt{\Delta}}{\omega_3\omega_2 + \omega_5\omega_1 - \omega_5\omega_0 - \omega_4\omega_2} \quad \text{s.t. } q \in \{-1, 1\}, \tag{3}$$

where

$$\Delta = (\omega_4\omega_0 - \omega_3\omega_1)^2 + (\omega_3\omega_2 - \omega_5\omega_0)^2 - (\omega_4\omega_2 - \omega_5\omega_1)^2. \tag{4}$$

The expressions of ω_0-ω_5 are given in [50]. κ and $\varkappa$ can then be obtained using:

$$\mathbf{x} = \varkappa \begin{bmatrix} \kappa \\ 1 \end{bmatrix} = \big(\mathbf{T}(\Phi)^\top \mathbf{T}(\Phi)\big)^{-1} \mathbf{T}(\Phi)^\top \mathbf{g}. \tag{5}$$

Disparity transformation can therefore be realized using [50]:

$$\mathbf{G}'(\mathbf{p}) = \mathbf{G}(\mathbf{p}) - \varkappa\big(\cos\Phi v - \sin\Phi u\big) - \varkappa\kappa + \Lambda, \tag{6}$$

where Λ is a constant used to ensure that the values in the transformed disparity (or inverse depth) image $\mathbf{G}'$ are non-negative. As illustrated in Fig. 6, the road damages become highly distinguishable after disparity transformation, and they can be detected with common image segmentation techniques. For instance, [7] applies superpixel clustering [62] to group the transformed disparities into a set of perceptually meaningful regions, which are then utilized to replace the rigid pixel grid structure. The damaged road areas can then be segmented by selecting the superpixels whose transformed disparity values are lower than those of the healthy road areas.

4.1.2 Convolutional Neural Network-Based Approaches

On the other hand, the CNN-based approaches typically address road damage detection in an end-to-end manner, with image classification [63], semantic segmentation, or object detection networks.

Image classification-based approaches

Since 2012, various CNNs have been proposed to solve image classification problems. AlexNet [64] is one of the modern CNNs that pushed image classification accuracy significantly. VGG architectures [65] improved over AlexNet [64] by increasing the network depth, which enabled them to learn more complicated image features.

[1] Source code is publicly available at https://github.com/ruirangerfan/unsupervised_disparity_map_segmentation.

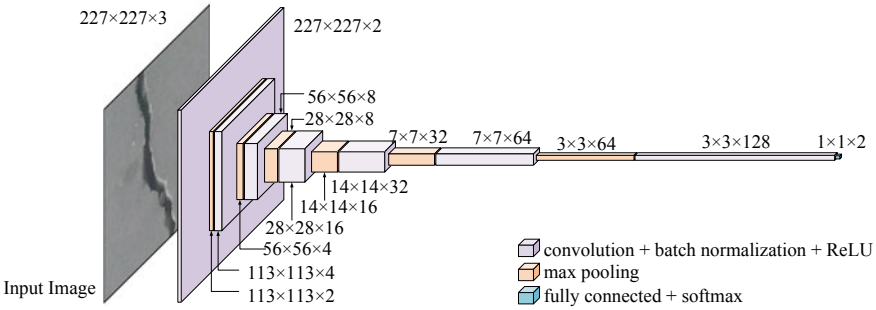

Fig. 7 The architecture of the road crack detection network presented in [73]

However, VGG architectures consist of hundreds of millions of parameters, making them very memory-consuming. GoogLeNet [66] (also known as Inception-v1) and Inception-v3 [67] go deeper in parallel paths with different receptive field sizes, so that the Inception module can act as a multi-level image feature extractor. Compared to VGG [65] architectures, GoogLeNet [66] and Inception-v3 [67] have lower computational complexity. However, with the increase of network depth, accuracy gets saturated and then degrades rapidly [68], due to vanishing gradients. To tackle this problem, [68] introduces residual neural network (ResNet). It is comprised of a stack of building blocks, each of which utilizes identity mapping (skip connection) to add the output from the previous layer to the layer ahead. This helps gradients propagate. In recent years, researchers have turned their focus to light-weight CNNs, which can be embedded in mobile devices to perform real-time image classification. MobileNet-v2 [69], ShuffleNet-v2 [70], and MNASNet [71] are three of the most popular CNNs of this kind. MobileNet-v2 [69] mainly introduces the inverted residual and linear bottleneck layer, which allows high accuracy and efficiency in mobile vision applications.

The above-mentioned image classification networks have been widely used to detect, i.e., recognize and localize, road cracks, as the visual features learned by CNNs can replace the traditional hand-crafted features [72]. For example, in 2016, [12] proposes a robust road crack detection network. An RGB road image is fed into a CNN consisting of a collection of convolutional layers. The learned visual feature is then connected with a fully connected (FC) layer to produce a scalar indicating the probability that the image contains road cracks. Similarly, in 2019, [73] proposes a deep CNN, as illustrated in Fig. 7, which is capable of classifying road images as either positive or negative. The road images are also segmented using an image thresholding method for pixel-level road crack detection.

In 2019, [63] proposed a self-supervised monocular road damage detection algorithm, which can not only reconstruct the 3-D road geometry models with multi-view images captured by a single movable camera (based on the hypothesis that the road surface is nearly planar) but also classify RGB road images as either damaged or

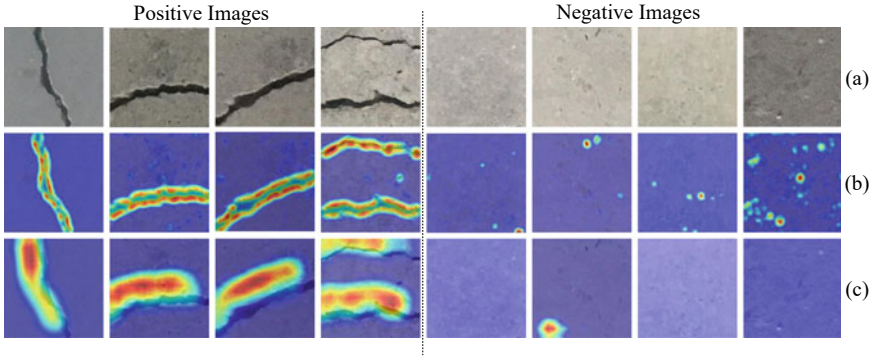

Fig. 8 Examples of Grad-CAM++ [76] results: **a** road crack images, **b** the class activation maps of ResNet-50 [68], **c** the class activation maps of SENet [77]. The warmer the color is, the more attention the CNN pays

undamaged with a classification CNN (the images used for training was automatically annotated via a 3-D road point cloud thresholding algorithm).

Recently, [74] conducted a comprehensive comparison among 30 SOTA image classification CNNs for road crack detection. The qualitative and quantitative experimental results suggest that road crack detection is not a challenging image classification task. The performances achieved by these deep CNNs are very similar [74]. Furthermore, learning road crack detection does not require a large amount of training data. It is demonstrated that 10,000 images are sufficient to train a well-performing CNN [74]. However, the pre-trained CNNs typically perform unsatisfactorily on additional test sets. Therefore, unsupervised domain adaptation (UDA) [75] capable of mapping two different domains becomes a hot research topic that requires more attention. Finally, explainable AI algorithms, such as Grad-CAM++ [76], become essential for image classification applications in terms of understanding CNN's decision-making, as shown in Fig. 8.

Object detection-based approaches

Region-based CNN (R-CNN) [78–80] series and you only look once (YOLO) [81–84] series are two representative groups of modern deep learning-based object detection algorithms. As discussed above, the SOTA image classification CNNs can classify an input image into a specific category. However, we sometimes hope to know the exact location of a particular object in the image, which requires the CNNs to learn bounding boxes around the objects of interest. A naive but straightforward way to achieve this objective is to classify a collection of patches (split from the original image) as positive (the object present in the patch) or negative (the object absent from the patch). However, the problem with this approach is that the objects of interest can have different spatial locations within the image and different aspect

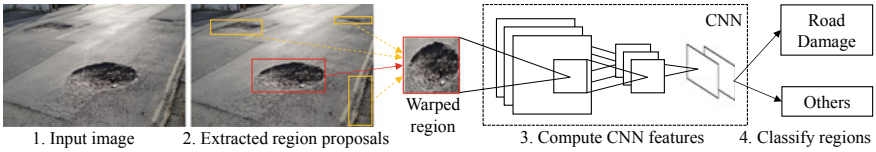

Fig. 9 R-CNN [78] architecture

ratios. Therefore, it is usually necessary but costly for this approach to select many regions of interests (RoIs).

To solve this dilemma, R-CNN [78] utilizes a selective search algorithm [85] to extract only 2000 RoIs (referred to as region proposals). Such RoIs are then fed into a classification CNN to extract visual features connected by a support vector machine (SVM) to produce their categories, as illustrated in Fig. 9. In 2015, Fast R-CNN [79] was introduced as a faster object detection algorithm and is based on R-CNN. Instead of feeding region proposals to the classification CNN, Fast R-CNN [79] directly feeds the original image to the CNN and outputs a convolutional feature map (CFM). The region proposals are identified from the CFM using selective search, which are then reshaped into a fixed size with an RoI pooling layer before being fed into a fully connected layer. A softmax layer is subsequently used to predict the region proposals' classes. However, both R-CNN [78] and Fast R-CNN [79] utilize a selective search algorithm to determine region proposals, which is slow and time-consuming. To overcome this disadvantage, [80] proposes Faster R-CNN, which employs a network to learn the region proposals. Faster R-CNN [80] is ∼250 times faster than R-CNN [78] and ∼11 times faster than Fast R-CNN [79].

On the other hand, the YOLO series are very different from the R-CNN series. YOLOv1 [81] formulates object detection as a regression problem. It first splits an image into an $S \times S$ grid, from which B bounding boxes are selected. The network then outputs class probabilities and offset values for these bounding boxes. Since YOLOv1 [81] makes a significant number of localization errors and its achieved recall is relatively low, YOLOv2 [82] was proposed to bring various improvements on YOLOv1 [81]: (1) it adds batch normalization on all the convolutional layers; (2) it fine-tunes the classification network at the full 448×448 resolution; (3) it removes the fully connected layers from YOLOv1 [81] and uses anchor boxes to predict bounding boxes. In 2018, [83] made several tweaks to further improve [82].

As for the application of the aforementioned object detectors in road damage detection, in 2018, [86] trained a YOLOv2 [82] object detector to recognize road potholes, while [87] utilized a Faster R-CNN [80] to achieve the same objective. In 2019, [88] utilized three different YOLOv3 [83] architectures to detect road potholes. In 2020, [89] leveraged YOLOv3 [83] to detect road potholes from RGB images captured by a smartphone mounted in a car. The algorithm was successfully implemented on a Raspberry Pi 2 Model B in TensorFlow. The reported best mean average precision (mAP) was 68.83% and inference time was 10 ms. Recently, [14] compared the

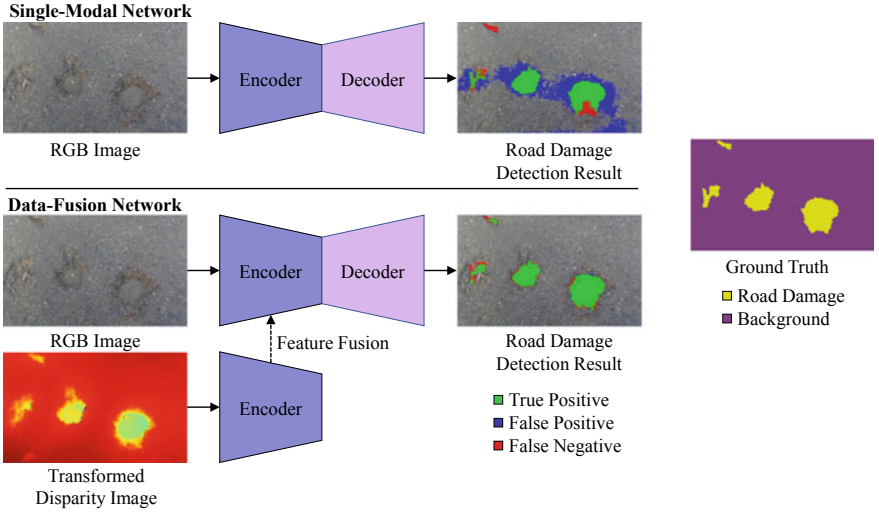

Fig. 10 Single-modal semantic image segmentation versus data-fusion semantic image segmentation for road pothole detection

performances of YOLOv2 [82] and Mask R-CNN [90] for road damage detection.[2] However, these approaches typically employ a well-developed object detection network to detect road damages, without modifications designed specifically for this task. Furthermore, these object detectors can only provide instance-level predictions instead of pixel-level predictions. Therefore, in recent years, semantic image segmentation has become a more desirable technique for road damage detection.

Semantic segmentation-based approaches

Semantic image segmentation aims at assigning each image pixel with a category [5]. It has been widely used for road damage detection in recent years. According to the number of input vision data types, the SOTA semantic image segmentation networks can be grouped into two categories: (1) single-modal [91] and (2) data-fusion [92, 93], as illustrated in Fig. 10. The former inputs only one type of vision data, such as RGB images or depth images, while the latter typically learns visual features from different types of vision data, such as RGB images and transformed disparity images [94] or RGB images and surface normal information [92].

Fully convolutional network (FCN) [95] was the first end-to-end single-modal CNN designed for semantic image segmentation. SegNet [96] was the first to use the encoder-decoder architecture, which is widely used in current networks. It consists of an encoder network, a corresponding decoder network, and a final pixel-wise

[2] A demo video can be found at https://vimeo.com/337886918.

classification layer. U-Net also [97] employs an encoder-decoder architecture. It adds skip connections between the encoder and decoder to smoothen the gradient flow and restore the object locations. Furthermore, PSPNet [98], DeepLabv3+ [99] and DenseASPP [100] leverage a pyramid pooling module to extract context information for better segmentation performance. Additionally, GSCNN [101] employs a two-branch framework consisting of (1) a shape branch and (2) a regular branch, which can effectively improve the semantic predictions on the boundaries.

On the other hand, data-fusion networks improve semantic segmentation accuracy by extracting and fusing the features from multi-modalities of visual information [102]. For instance, FuseNet [103] and depth-aware CNN [104] adopt the popular encoder-decoder architecture, but employ different operations to fuse the feature maps obtained from the RGB and depth branches. SNE-RoadSeg [92] and SNE-RoadSeg+ [93] extract and fuse the visual features from RGB images and surface normal information (translated from depth/disparity images in an end-to-end manner) for road segmentation.

Back to the road inspection applications, researchers have already employed both single-modal and data-fusion semantic image segmentation networks to detect road damages/anomalies (the approaches designed to detect road damages can also be utilized to detect road anomalies, as these two types of objects are below and above the road, respectively).

In 2020, [5] proposed a novel attention aggregation (AA) framework,[3] making the most of three different attention modules: (1) channel attention module (CAM), (2) position attention module (PAM), and (3) dual attention module (DAM). Furthermore, [5] introduces an effective training set augmentation technique based on adversarial domain adaptation (developed based on CycleGAN [105], as illustrated in Fig. 11), where the synthetic road RGB images and transformed road disparity (or inverse depth) images are generated to enhance the training of semantic segmentation networks. Extensive experiments conducted with five SOTA single-modal CNNs and three data-fusion CNNs demonstrated the effectiveness of the proposed AA framework and training set augmentation technique. Moreover, a large-scale multi-modal pothole detection dataset (containing RGB images and transformed disparity images), referred to as Pothole-600,[4] was published for research purposes.

Similar to [5], [106] also introduced a road pothole detection approach based on single-modal semantic image segmentation in 2021. Its network architecture, as shown in Fig. 12, is developed based on DeepLabv3 [107]. It first extracts visual features from an input (RGB or transformed disparity) road image using a backbone CNN. A channel attention module then reweighs the channel features to enhance the consistency of different feature maps. Then, an atrous spatial pyramid pooling (ASPP) module (comprising of atrous convolutions in series, with progressive rates of dilation) is employed to integrate the spatial context information. This greatly helps to better distinguish between potholes and undamaged road areas. Finally, the feature maps in the adjacent layers are fused using the proposed multi-scale feature fusion

[3] Source code is publicly available at: https://github.com/hlwang1124/AAFramework.

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

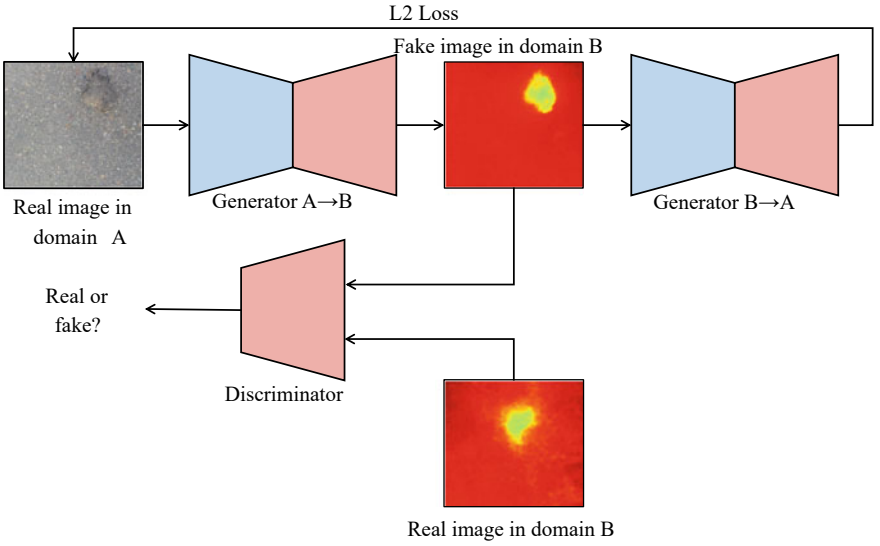

Fig. 11 CycleGAN [105] architecture

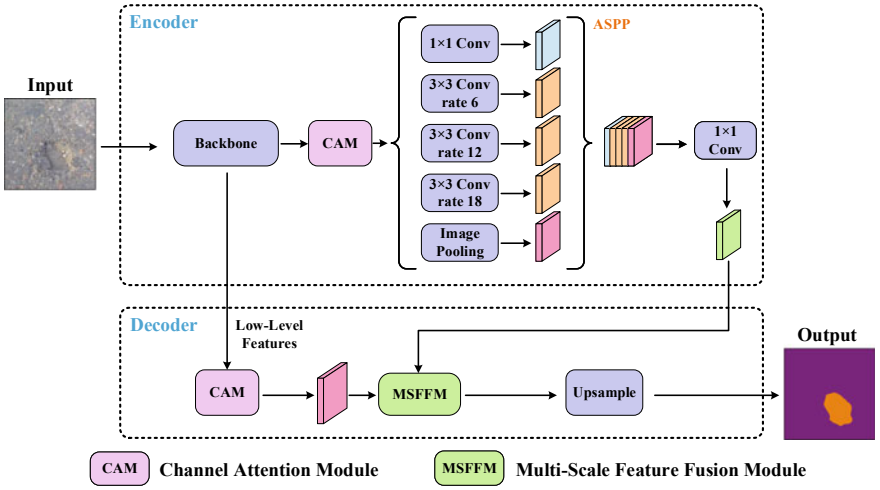

Fig. 12 The architecture of the road pothole detection network introduced in [106]

module. This further reduces the semantic gap between different feature channel layers. Extensive experimental results conducted on the Pothole-600 [5] dataset show that the pixel-level mIoU and mFsc are 72.75 and 84.22%, respectively.

Recently, [94] developed a data-fusion semantic segmentation CNN for road anomaly detection and the authors also conducted a comprehensive comparison of data fusion performance with respect to six different modalities of visual features,

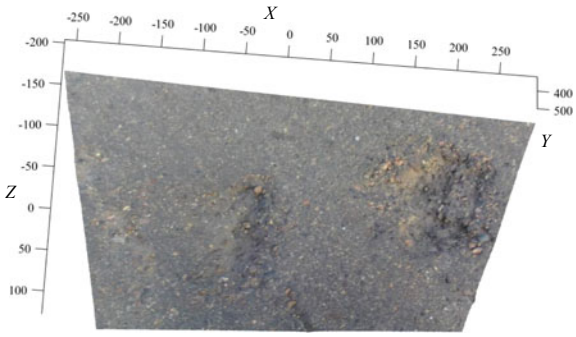

Fig. 13 An example of 3-D road pothole cloud acquired using the stereo vision technique presented in [1]. The accuracy of the 3-D road point cloud is 3 mm

including (1) RGB images, (2) disparity images, (3) surface normal images [108], (4) elevation images, (5) HHA images (having three channels: (a) disparity, (b) height of the pixels, and (3) angle between the normals and gravity vector based on the estimated ground [103]), and (6) transformed disparity images [3]. The experimental results demonstrated that the transformed disparity image is the most informative visual feature.

4.2 3-D Road Surface Modeling-Based Approaches

In 2020, [109] proposed a LiDAR-based road inspection 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, [109] lacks the algorithm details and necessary quantitative experimental road damage detection results.

In [48, 51], the 3-D road point clouds,[5] as shown in Fig. 13, are formulated as quadratic surfaces, as follows:

$$f(X, Z) = a_0 + a_1 X + a_2 Z + a_3 X^2 + a_4 Z^2 + a_5 X Z, \tag{7}$$

where $\mathbf{P} = (X; Y; Z)$ is a 3-D point on the road surface in the camera coordinate system and $\mathbf{a} = (a_0; a_1; a_2; a_3; a_4; a_5)$ stores the quadratic surface coefficients. $\mathbf{a}$ can be estimated by minimizing:

$$E = \sum_{q=1}^{K} \Big(Y_q - f(X_q, Z_q) \Big)^2. \tag{8}$$

The optimal $\mathbf{a}$ can be obtained when [51]:

[5] 3-D road point clouds are publicly available at https://github.com/ruirangerfan/stereo_pothole_datasets.

$$\frac{\partial E}{\partial a_0} = \frac{\partial E}{\partial a_1} = \frac{\partial E}{\partial a_2} = \frac{\partial E}{\partial a_3} = \frac{\partial E}{\partial a_4} = \frac{\partial E}{\partial a_5} = 0, \tag{9}$$

which results in:

$$\mathbf{Ma} = \mathbf{q}, \tag{10}$$

where

$$\mathbf{M} = \begin{bmatrix} K & S_X & S_Z & S_{X^2} & S_{Z^2} & S_{XZ} \\ S_X & S_{X^2} & S_{XZ} & S_{X^3} & S_{XZ^2} & S_{ZX^2} \\ S_Z & S_{XZ} & S_{Z^2} & S_{ZX^2} & S_{Z^3} & S_{XZ^2} \\ S_{X^2} & S_{X^3} & S_{ZX^2} & S_{X^4} & S_{X^2Z^2} & S_{ZX^3} \\ S_{Z^2} & S_{XZ^2} & S_{Z^3} & S_{X^2Z^2} & S_{Z^4} & S_{XZ^3} \\ S_{XZ} & S_{X^2Z} & S_{XZ^2} & S_{X^3Z} & S_{XZ^3} & S_{X^2Z^2} \end{bmatrix}, \tag{11}$$

and

$$\mathbf{q} = (S_Y; S_{XY}; S_{YZ}; S_{YX^2}; S_{YZ^2}; S_{XYZ}). \tag{12}$$

S represents sum operation. For example, $S_{XZ^2} = \sum_1^K X_q {Z_q}^2$ and $S_{XYZ} = \sum_1^K X_q Y_q Z_q$. Equations (11) and (12) can be rewritten as:

$$\mathbf{M} = \mathbf{W}^\top \mathbf{W}, \tag{13}$$

and

$$\mathbf{q} = \mathbf{W}^\top \mathbf{y}, \tag{14}$$

where

$$\mathbf{W} = \begin{bmatrix} 1 & X_1 & Z_1 & {X_1}^2 & {Z_1}^2 & X_1 Z_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & X_K & Z_K & {X_K}^2 & {Z_K}^2 & X_K Z_K \end{bmatrix}, \tag{15}$$

and

$$\mathbf{y} = (Y_1; Y_2; \cdots; Y_n). \tag{16}$$

Therefore, **a** has the following expression:

$$\mathbf{a} = \left(\mathbf{W}^\top \mathbf{W}\right)^{-1} \mathbf{W}^\top \mathbf{y}. \tag{17}$$

Plugging (17) into (7) and comparing the difference between the actual and modeled road surfaces, the damaged road areas can be extracted. Nevertheless, 3-D road surface modeling is very sensitive to noise. Hence, [51] incorporates the surface normal information into the road surface modeling process to eliminate outliers. Furthermore, [3] also employs random sample consensus (RANSAC) [110] to further improve its robustness.

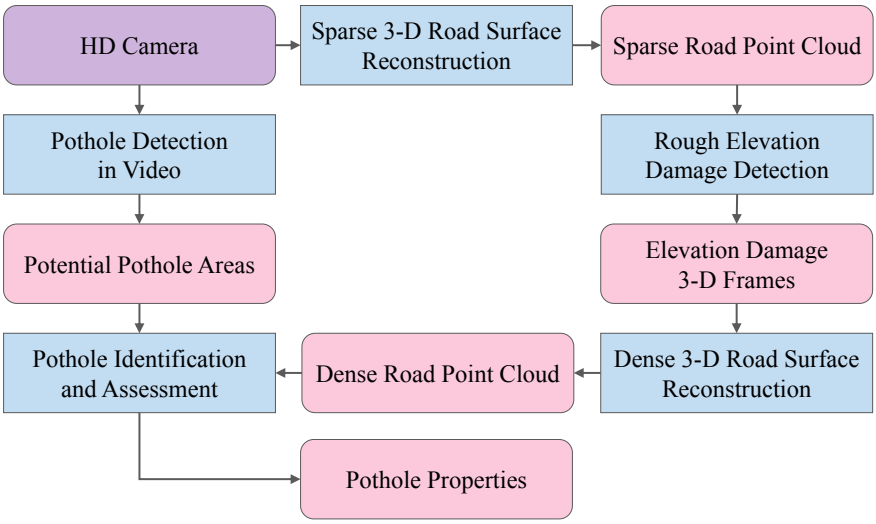

Fig. 14 The block diagram of the method proposed in [22]

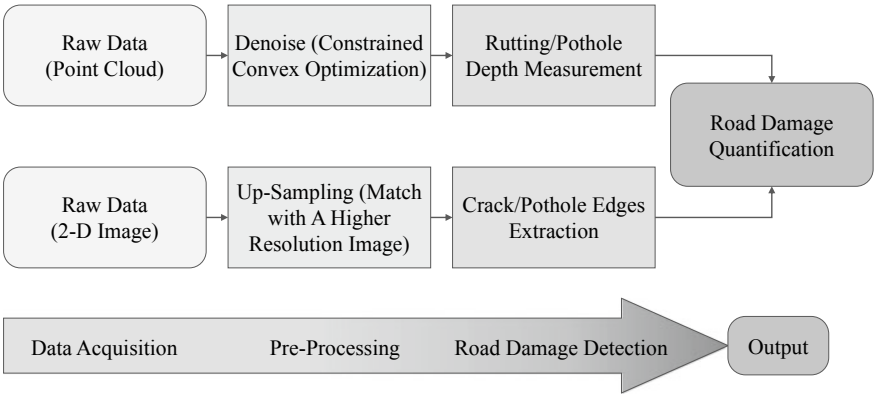

Fig. 15 The block diagram of the road damage detection method proposed in [111]

4.3 Hybrid Approaches

In 2012, [22] introduced a hybrid road pothole detection approach based on 2-D object recognition and 3-D road geometry reconstruction. As shown in Fig. 14, the road video frames acquired by a high definition (HD) camera were first utilized to detect (recognize) road potholes. Simultaneously, the same video was also employed for sparse 3-D road geometry reconstruction. By analyzing such multi-modal road inspection results, the potholes were accurately detected. Such a hybrid method dramatically reduces the incorrectly detected road potholes.

In 2014, [111] also introduced a hybrid road damage detection approach, which can effectively recognize and measure road damages. As illustrated in Fig. 15, this algorithm requires two modalities of road data: (1) 2-D gray-scale/RGB road images (2) 3-D road point clouds. The road images are up-sampled by matching them with higher resolution images. The road pothole/crack edges are then extracted from the up-sampled images. On the other hand, the road point clouds are denoised through constrained convex optimization to measure rutting/pothole depth. Finally, the extracted pothole/crack edges and measured rutting/pothole depth are combined for road damage quantification. Unfortunately, these two modalities of road data were processed independently, and the data fusion process was not discussed in this work.

In 2017, [112] proposed an automated road pothole detection system based on the analysis of 2-D LiDAR data and RGB road images. This hybrid system has the advantage of not being affected by electromagnetic waves and poor road conditions. The 2-D LiDAR data provide road profile information, while the RGB road images provide road texture information. To obtain a highly accurate and large road area, this system uses two LiDARs. Such a hybrid system can combine the advantages of different sources of vision data to improve the overall road pothole detection accuracy.

In 2019, [3] also introduced a hybrid framework for road pothole detection. The methodology is illustrated in Fig. 16. It first applies disparity transformation, discussed in Sect. 4.1.1, and Otsu's thresholding [113] methods to extract potential undamaged road areas. In the next step, a quadratic surface was fitted to the original

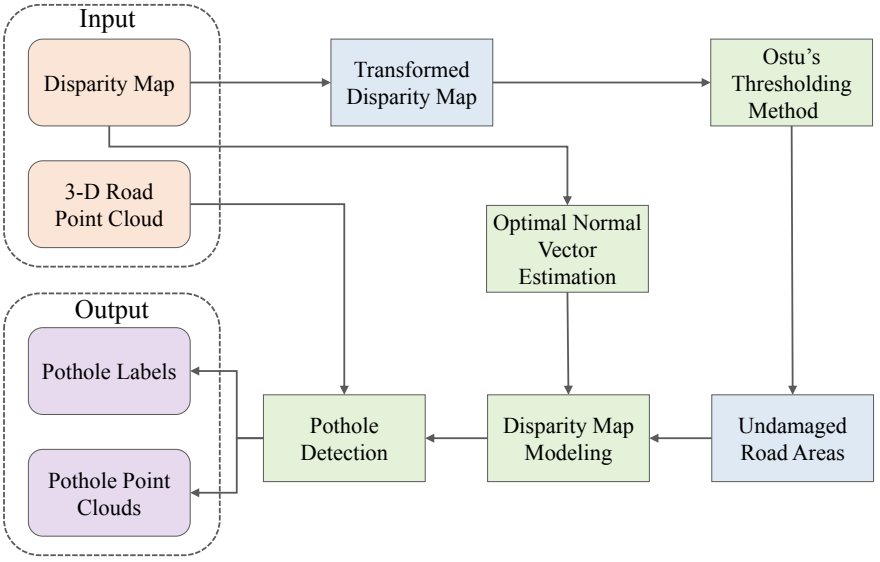

Fig. 16 The road pothole detection workflow proposed in [3]

disparity image, where the RANSAC algorithm was employed to enhance the surface fitting robustness. Potholes can be successfully detected by comparing the actual and fitted disparity images. This hybrid framework uses both 2-D image processing and 3-D road surface modeling algorithms, greatly improving the road pothole detection performance.

5 Parallel Computing Architecture

Graphics processing unit (GPU) was initially developed to accelerate graphics processing. It has now evolved as a specialized processing core, significantly speeding up computational processes. GPUs have unique parallel computing architectures and they can be used for a wide range of massively distributed computational processes, such as graphics rendering, supercomputing, weather forecasting, and autonomous driving.

The main advantage of GPUs is its massively parallel architecture. There are several different ways to classify parallel computers. One of the most widely used taxonomies, in use since 1966, is referred to as Flynn's taxonomy [114]. Flynn classifies parallel architectures based on the two independent dimensions of instruction and data. Each dimension can only be either single or multiple. There are four parallel computing architectures: (1) single instruction, single data (SISD), (2) single instruction, multiple data (SIMD), (3) multiple instructions, single data (MISD), and (4) multiple instructions, multiple data (MIMD).

As illustrated in Fig. 17, there are currently two main popular GPU architectures [115]: (1) NVIDIA's Tesla/Fermi GPU architecture and (2) AMD's Evergreen/Northern Island series GPU architecture. The major difference between these two types of GPU architectures is the processor core: the computing unit of NVIDIA GPUs is a streaming multiprocessor (SM) cluster, while the computing unit of AMD GPUs is a data-parallel processor (DPP) array.

For NVIDIA GPUs, different types of GPUs may have different SM numbers. Each SM has 8 to 48 CUDA cores composed of three parts: an integer/floating-point unit, a warp scheduler for instruction scheduling, and on-chip memory. For AMD GPU, its computing unit DPP has 16 thread processors, containing the execution pipeline of the very long instruction word (VLIW) architecture and the local data sharing part. Furthermore, NVIDIA uses a GigaThread engine while AMD uses a command processor/ultra-threaded dispatch processor to control compute units. In addition, the GPU also includes the interconnection part between the computing unit and the memory subsystem. They are all composed of cache, memory controller, and GPU device memory.

Parallelism is the basis of high-performance computing. With the continuous improvement in computing power and parallelism programmability, GPUs have been employed in an increasing number of deep learning applications, such as road condition assessment with semantic segmentation networks, to accelerate model training and inference.

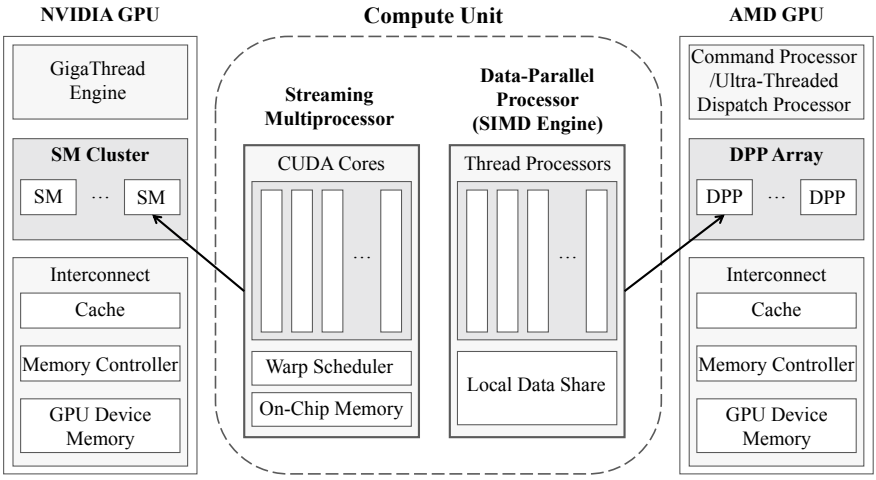

Fig. 17 NVIDIA and AMD GPU architectures presented in [115]

6 Summary

This chapter gave readers an overall picture of the SOTA computer-aided road inspection approaches. Five common types of road damages, including crack, spalling, pothole, rutting, and shoving, were discussed; The 2-D/3-D road imaging techniques, especially laser scanning, IR sensing, and multi-view geometry from digital images, were discussed. Finally, the existing machine vision and intelligence approaches, including 2-D image analysis/understanding algorithms, 3-D road surface modeling algorithms, and the hybrid methods developed to detect road damages were reviewed.

Acknowledgements This work was supported by the National Key R&D Program of China, under grant No. 2020AAA0108100, awarded to Prof. Qijun Chen. This work was also supported by the Fundamental Research Funds for the Central Universities, under projects No. 22120220184, No. 22120220214, and No. 2022-5-YB-08, awarded to Prof. Rui Fan.

References

1. R. Fan et al., Road surface 3d reconstruction based on dense subpixel disparity map estimation. IEEE Trans. Image Process. **27**(6), 3025–3035 (2018)
2. C. Koch et al., A review on computer vision based defect detection and condition assessment of concrete and asphalt civil infrastructure. Adv. Eng. Inform. **29**(2), 196–210 (2015)
3. R. Fan et al., Pothole detection based on disparity transformation and road surface modeling. IEEE Trans. Image Process. **29**, 897–908 (2019)
4. T. Kim, S.-K. Ryu, Review and analysis of pothole detection methods. J. Emerg. Trends Comput. Inf. Sci. **5**(8), 603–608 (2014)

5. R. Fan et al., We learn better road pothole detection: from attention aggregation to adversarial domain adaptation, in *European Conference on Computer Vision Workshops* (Springer, Berlin, 2020), pp. 285–300
6. S. Mathavan et al., A review of three-dimensional imaging technologies for pavement distress detection and measurements. IEEE Trans. Intell. Transp. Syst. **16**(5), 2353–2362 (2015)
7. R. Fan et al., Rethinking road surface 3-d reconstruction and pothole detection: from perspective transformation to disparity map segmentation. IEEE Trans. Cybern. **52**(7), 5799–5808 (2022)
8. R. Fan, Long-awaited next-generation road damage detection and localization system is finally here, in *29th European Signal Processing Conference (EUSIPCO)* (IEEE, 2021), pp. 641–645
9. Y.-C. Tsai, A. Chatterjee, Pothole detection and classification using 3d technology and watershed method. J. Comput. Civ. Eng. **32**(2), 04017078 (2018)
10. M.R. Jahanshahi et al., Unsupervised approach for autonomous pavement-defect detection and quantification using an inexpensive depth sensor. J. Comput. Civ. Eng. **27**(6), 743–754 (2013)
11. R. Fan 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 Computer Society, 2019), pp. 535–543
12. L. Zhang, Road crack detection using deep convolutional neural network, in *IEEE International Conference on Image Processing (ICIP)* (IEEE, 2016), pp. 3708–3712
13. R. Fan et al., Graph attention layer evolves semantic segmentation for road pothole detection: a benchmark and algorithms. IEEE Trans. Image Process. **30**, 8144–8154 (2021)
14. A. Dhiman, R. Klette, Pothole detection using computer vision and learning. IEEE Trans. Intell. Transp. Syst. **21**(8), 3536–3550 (2019)
15. D.S. Mahler et al., Pavement distress analysis using image processing techniques. Comput.-Aided Civ. Infrastruct. Eng. **6**(1), 1–14 (1991)
16. H.N. Koutsopoulos, A. Downey, Primitive-based classification of pavement cracking images. J. Transp. Eng. **119**(3), 402–418 (1993)
17. J. Laurent et al., Road surface inspection using laser scanners adapted for the high precision 3d measurements of large flat surfaces, in *Proceedings. International Conference on Recent Advances in 3-D Digital Imaging and Modeling (Cat. No. 97TB100134)* (IEEE, 1997), pp. 303–310
18. D. Joubert, A. Tyatyantsi, J. Mphahlehle, V. Manchidi, Pothole tagging system, in *Robotics and Mechatronics Conference of South Africa, CSIR International Conference Centre, Pretoria* (2011), pp. 23–25
19. I. Moazzam, K. Kamal, S. Mathavan, S. Usman, M. Rahman, Metrology and visualization of potholes using the microsoft kinect sensor, in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)* (IEEE, 2013), pp. 1284–1291
20. M. Barbato, G. Orlandi, M. Panella, Real-time identification and tracking using kinect
21. A.M. Andrew, Multiple view geometry in computer vision, in *Kybernetes* (2001)
22. G. Jog, C. Koch, M. Golparvar-Fard, I. Brilakis, Pothole properties measurement through visual 2d recognition and 3d reconstruction. Comput. Civ. Eng. **2012**, 553–560 (2012)
23. S. Ullman, The interpretation of structure from motion. Proc. R. Soc. Lond. Ser. B. Biol. Sci. **203**(1153), 405–426 (1979)
24. H. Wang et al., CoT-AMFlow: adaptive modulation network with co-teaching strategy for unsupervised optical flow estimation, in *Conference on Robot Learning (CoRL)* (2020)
25. B. Triggs, P.F. McLauchlan, R.I. Hartley, A.W. Fitzgibbon, Bundle adjustment-a modern synthesis, in *International workshop on vision algorithms* (Springer, Berlin, 1999), pp. 298–372
26. M.U.M. Bhutta et al., Loop-box: multiagent direct slam triggered by single loop closure for large-scale mapping. IEEE Trans. Cybern. **52**(6), 5088–5097 (2022)
27. N. Ma et al., Computer vision for road imaging and pothole detection: a state-of-the-art review of systems and algorithms. Transp. Saf. Environ. (2022) (in press)

28. H. Hirschmuller, Stereo processing by semiglobal matching and mutual information. IEEE Trans. Pattern Anal. Mach. Intell. **30**(2), 328–341 (2007)
29. J. Sun, N.-N. Zheng, H.-Y. Shum, Stereo matching using belief propagation. IEEE Trans. Pattern Anal. Mach. Intell. **25**(7), 787–800 (2003)
30. H. Wang et al., Pvstereo: pyramid voting module for end-to-end self-supervised stereo matching. IEEE Robot. Autom. Lett. **6**(3), 4353–4360 (2021)
31. R. Danzl, F. Helmli, S. Scherer, Focus variation–a new technology for high resolution optical 3d surface metrology, in *The 10th International Conference of the Slovenian Society for Non-Destructive Testing* (Citeseer, 2009), pp. 484–491
32. Y. Sun, S. Duthaler, B.J. Nelson, Autofocusing algorithm selection in computer microscopy, in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems* (IEEE, 2005), pp. 70–76
33. S.K. Nayar, Y. Nakagawa, Shape from focus. IEEE Trans. Pattern Anal. Mach. Intell. **16**(8), 824–831 (1994)
34. C. Wöhler, Triangulation-based approaches to three-dimensional scene reconstruction, in *3D Computer Vision* (Springer, Berlin, 2013), pp. 3–87
35. A. Kuhl, C. Wöhler, L. Krüger, P. d'Angelo, H.-M. Groß, Monocular 3d scene reconstruction at absolute scales by combination of geometric and real-aperture methods, in *Joint Pattern Recognition Symposium* (Springer, Berlin, 2006), pp. 607–616
36. M. Subbarao, G. Surya, Depth from defocus: a spatial domain approach. Int. J. Comput. Vision **13**(3), 271–294 (1994)
37. S. Nayar, K. Ikeuchi, T. Kanade, Surface reflection: physical and geometrical perspectives, in *Proceedings: Image Understanding Workshop* (1990), pp. 185–212
38. R.J. Woodham, Photometric method for determining surface orientation from multiple images. Opt. Eng. **19**(1), 191139 (1980)
39. S. Barsky, M. Petrou, The 4-source photometric stereo technique for three-dimensional surfaces in the presence of highlights and shadows. IEEE Trans. Pattern Anal. Mach. Intell. **25**(10), 1239–1252 (2003)
40. J.G. Fujimoto, C. Pitris, S.A. Boppart, M.E. Brezinski, Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy. Neoplasia **2**(1–2), 9–25 (2000)
41. W.J. Walecki, P. Van, Determining thickness of slabs of materials by inventors, 3 Oct 2006, uS Patent 7,116,429
42. D. Scharstein, R. Szeliski, High-accuracy stereo depth maps using structured light, in *Proceedings 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003*, vol. 1 (IEEE, 2003), p. I
43. V. Muzet, C. Heinkele, Y. Guillard, J.M. Simonin, Surface deflection measurement using structured light, in *Testing in Civil Engineering, Nantes, France* (2009)
44. T. Oggier, M. Lehmann, R. Kaufmann, M. Schweizer, M. Richter, P. Metzler, G. Lang, F. Lustenberger, N. Blanc, An all-solid-state optical range camera for 3d real-time imaging with sub-centimeter depth resolution (swissranger), in *Optical Design and Engineering*, vol. 5249 (International Society for Optics and Photonics, 2004), pp. 534–545
45. D. Anderson, H. Herman, A. Kelly, Experimental characterization of commercial flash ladar devices, in *International Conference of Sensing and Technology*, vol. 2 (Citeseer, 2005), pp. 17–23
46. C.F. Özgenel, Concrete crack images for classification, in *Mendeley Data, V1* (2018)
47. M. Maguire, S. Dorafshan, R.J. Thomas, Sdnet2018: a concrete crack image dataset for machine learning applications (2018)
48. Z. Zhang, Advanced stereo vision disparity calculation and obstacle analysis for intelligent vehicles, Ph.D. dissertation, University of Bristol (2013)
49. A. Alzoubi, PotDataset **4** (2018). https://cord.cranfield.ac.uk/articles/dataset/PotDataset/5999699
50. R. Fan, M. Liu, Road damage detection based on unsupervised disparity map segmentation. IEEE Trans. Intell. Transp. Syst. **21**(11), 4906–4911 (2019)

51. U. Ozgunalp, Vision based lane detection for intelligent vehicles, Ph.D. dissertation, University of Bristol (2016)
52. E. Buza, S. Omanovic, 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 (2013), p. 4853
53. S.-K. Ryu, T. Kim, Y.-R. Kim, Image-based pothole detection system for its service and road management system. Math. Probl. Eng. **2015** (2015)
54. C. Koch, I. Brilakis, Pothole detection in asphalt pavement images. Adv. Eng. Inform. **25**(3), 507–515 (2011)
55. G.W. Zack, W.E. Rogers, S.A. Latt, Automatic measurement of sister chromatid exchange frequency. J. Histochem. Cytochem. **25**(7), 741–753 (1977)
56. G. Beylkin, R. Coifman, V. Rokhlin, Fast wavelet transforms and numerical algorithms, in *Fundamental Papers in Wavelet Theory* (Princeton University Press, 2009), pp. 741–783
57. L. Najman, M. Schmitt, Watershed of a continuous function. Signal Process. **38**(1), 99–112 (1994)
58. R. Labayrade, D. Aubert, J.-P. Tarel, Real time obstacle detection in stereovision on non flat road geometry through "v-disparity" representation, in *Intelligent Vehicle Symposium*, vol. 2 (IEEE, 2002), pp. 646–651
59. P. Pedregal, *Introduction to Optimization* (Springer Science & Business Media, 2006), vol. 46
60. R. Fan, N. Dahnoun, Real-time stereo vision-based lane detection system. Meas. Sci. Technol. **29**(7), 074005 (2018)
61. U. Ozgunalp et al., Multiple lane detection algorithm based on novel dense vanishing point estimation. IEEE Trans. Intell. Transp. Syst. **18**(3), 621–632 (2016)
62. R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, S. Süsstrunk, Slic superpixels compared to state-of-the-art superpixel methods. IEEE Trans. Pattern Anal. Mach. Intell. **34**(11), 2274–2282 (2012)
63. Y. Hu, T. Furukawa, A self-supervised learning technique for road defects detection based on monocular three-dimensional reconstruction, in *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, vol. 59216 (American Society of Mechanical Engineers, 2019), p. V003T01A021
64. A. Krizhevsky, One weird trick for parallelizing convolutional neural networks (2014). arXiv:1404.5997
65. K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in *International Conference on Learning Representations (ICLR)* (2015)
66. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (2015), pp. 1–9
67. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 2818–2826
68. K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 770–778
69. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.-C. Chen, Mobilenetv2: inverted residuals and linear bottlenecks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2018), pp. 4510–4520
70. N. Ma, X. Zhang, H.-T. Zheng, J. Sun, Shufflenet v2: practical guidelines for efficient CNN architecture design, in *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), pp. 116–131
71. M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard, Q.V. Le, Mnasnet: platform-aware neural architecture search for mobile, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2019), pp. 2820–2828
72. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. Nature **521**(7553), 436–444 (2015)
73. R. Fan, Road crack detection using deep convolutional neural network and adaptive thresholding, in *IEEE Intelligent Vehicles Symposium (IV)* (IEEE, 2019), pp. 474–479

74. J. Fan 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)
75. J. Hoffman, E. Tzeng, T. Park, J.-Y. Zhu, P. Isola, K. Saenko, A. Efros, T. Darrell, Cycada: cycle-consistent adversarial domain adaptation, in *International Conference on Machine Learning* (PMLR, 2018), pp. 1989–1998
76. A. Chattopadhay, A. Sarkar, P. Howlader, V.N. Balasubramanian, Grad-cam++: generalized gradient-based visual explanations for deep convolutional networks, in *IEEE Winter Conference on Applications of Computer Vision (WACV)* (IEEE, 2018), pp. 839–847
77. J. Hu, L. Shen, G. Sun, Squeeze-and-excitation networks, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 7132–7141
78. R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 580–587
79. R. Girshick, Fast R-CNN, in *Proceedings of the IEEE International Conference on Computer Vision* (2015), pp. 1440–1448
80. S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks. Adv. Neural. Inf. Process. Syst. **28**, 91–99 (2015)
81. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: Unified, real-time object detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 779–788
82. J. Redmon, A. Farhadi, Yolo9000: better, faster, stronger, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 7263–7271
83. J. Redmon, A. Farhadi, Yolov3: an incremental improvement. CoRR (2018)
84. A. Bochkovskiy, C.-Y. Wang, H.-Y. M. Liao, Yolov4: optimal speed and accuracy of object detection. CoRR (2020)
85. J.R. Uijlings, K.E. Van De Sande, T. Gevers, A.W. Smeulders, Selective search for object recognition. Int. J. Comput. Vision **104**(2), 154–171 (2013)
86. L.K. Suong, J. Kwon, Detection of potholes using a deep convolutional neural network. J. Univers. Comput. Sci. **24**(9), 1244–1257 (2018)
87. W. Wang, B. Wu, S. Yang, Z. Wang, Road damage detection and classification with faster R-CNN, in *IEEE International Conference on Big Data (Big Data)* (IEEE, 2018), pp. 5220–5223
88. E.N. Ukhwah, E.M. Yuniarno, Y.K. Suprapto, Asphalt pavement pothole detection using deep learning method based on yolo neural network, in *International Seminar on Intelligent Technology and Its Applications (ISITIA)* (IEEE, 2019), pp. 35–40
89. N. Camilleri, T. Gatt, Detecting road potholes using computer vision techniques, in *2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP)* (IEEE, 2020), pp. 343–350
90. K. He, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, in *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 2961–2969
91. R. Fan, H. Wang, P. Cai, J. Wu, M.J. Bocus, L. Qiao, M. Liu, Learning collision-free space detection from stereo images: homography matrix brings better data augmentation. IEEE/ASME Trans. Mechatron. **27**(1), 225–233 (2021)
92. R. Fan, H. Wang, P. Cai, M. Liu, Sne-roadseg: incorporating surface normal information into semantic segmentation for accurate freespace detection, in *European Conference on Computer Vision* (Springer, Berlin, 2020), pp. 340–356
93. H. Wang, R. Fan, P. Cai, M. Liu, 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. (to be published)
94. H. Wang, R. Fan, Y. Sun, M. Liu, Dynamic fusion module evolves drivable area and road anomaly detection: a benchmark and algorithms. IEEE Trans. Cybern. (2021). https://doi.org/10.1109/TCYB.2021.3064089

95. J. Long, E. Shelhamer, T. Darrell, Fully convolutional networks for semantic segmentation, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015), pp. 3431–3440
96. V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **39**(12), 2481–2495 (2017)
97. O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, Berlin, 2015), pp. 234–241
98. H. Zhao, J. Shi, X. Qi, X. Wang, J. Jia, Pyramid scene parsing network, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 2881–2890
99. L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), pp. 801–818
100. M. Yang, K. Yu, C. Zhang, Z. Li, K. Yang, Denseaspp for semantic segmentation in street scenes, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 3684–3692
101. T. Takikawa, D. Acuna, V. Jampani, S. Fidler, Gated-SCNN: gated shape CNNs for semantic segmentation, in *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2019), pp. 5229–5238
102. H. Wang, R. Fan, Y. Sun, M. Liu, 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), pp. 2706–2711
103. C. Hazirbas, L. Ma, C. Domokos, D. Cremers, Fusenet: incorporating depth into semantic segmentation via fusion-based CNN architecture, in *Asian Conference on Computer Vision* (Springer, 2016), pp. 213–228
104. W. Wang, U. Neumann, Depth-aware CNN for RGB-D segmentation, in *Proceedings of the European Conference on Computer Vision (ECCV)* (2018), pp. 135–150
105. J.-Y. Zhu, T. Park, P. Isola, A.A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 2223–2232
106. J. Fan, M.J. Bocus, B. Hosking, R. Wu, Y. Liu, S. Vityazev, R. Fan, Multi-scale feature fusion: learning better semantic segmentation for road pothole detection, in *2021 IEEE International Conference on Autonomous Systems (ICAS)* (IEEE, 2021), pp. 1–5
107. L.-C. Chen, G. Papandreou, F. Schroff, H. Adam, Rethinking atrous convolution for semantic image segmentation. CoRR (2017)
108. R. Fan, H. Wang, B. Xue, H. Huang, Y. Wang, M. Liu, I. Pitas, Three-filters-to-normal: an accurate and ultrafast surface normal estimator. IEEE Robot. Autom. Lett. **6**(3), 5405–5412 (2021)
109. R. Ravi, D. Bullock, A. Habib, Highway and airport runway pavement inspection using mobile lidar. Int. Arch. Photogramm., Remote. Sens. Spat. Inf. Sci. **43**, 349–354 (2020)
110. A. Hast, J. Nysjö, A. Marchetti, Optimal ransac-towards a repeatable algorithm for finding the optimal set (2013)
111. C. Yuan, H. Cai, Automatic detection of pavement surface defects using consumer depth camera, in *Construction Research Congress 2014: Construction in a Global Network* (2014), pp. 974–983
112. B.-h. Kang, S.-i. Choi, Pothole detection system using 2d lidar and camera, in *2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN)* (IEEE, 2017), pp. 744–746
113. N. Otsu, A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979)
114. M.J. Flynn, Very high-speed computing systems. Proc. IEEE **54**(12), 1901–1909 (1966)
115. S. Lee, W.W. Ro, Parallel GPU architecture simulation framework exploiting work allocation unit parallelism, in *IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)* (IEEE, 2013), pp. 107–117