Digital Transformation for Intelligent Road Condition Assessment

Sicen Guo, Yue Bai, Mohammud Junaid Bocus, and Rui Fan

Abstract Recently, governments have been resorting to cutting-edge artificial intelligence technologies to facilitate the digital transformation of smart cities. Remarkable progress has been made to strengthen smart city governance and sustainability, especially in road condition assessment. Road data acquisition and defect detection, two major processes of intelligent road condition assessment, play an important role in ensuring road maintainability while providing maximum traffic security and driving comfort. Traditional manual visual inspection is inefficient and lacks objectivity. Therefore, intelligent road condition assessment systems developed based on data-driven techniques have received increasing attention. This chapter presents the state-of-the-art intelligent road condition assessment systems, the existing challenges, and future development trends.

1 Introduction

We have ushered in the digital age since the mid-20th century [1]. In the context of the digital revolution, society will undergo a comprehensive upgrade and transformation

S. Guo and Y. Bai—Joint first authors of this chapter.

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

Y. Bai
e-mail: 1851853@tongji.edu.cn

R. Fan
e-mail: rfan@tongji.edu.cn

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

C. Kahraman and E. Haktanır (eds.), *Intelligent Systems in Digital Transformation*,
Lecture Notes in Networks and Systems 549,
https://doi.org/10.1007/978-3-031-16598-6_22

through modern digital technologies [2]. The increasing digitalization of various processes and manufacturing is resulting in increasing sensor data and researchers' participation [3]. The growth of data networks and the rise of digital processes have eventually led to the creation and development of a digital environment [4].

1.1 Digital Transformation

Digital transformation (DX) aims to drive faster product innovation and improve service delivery [5]. The realization of DX requires not only the support of data science technology [6], but also the unification of city management strategy and cooperation between departments [7].

DX has permeated many aspects of our daily life, such as medical care, industrial allocation, traffic control, and urban management. For instance, it has been playing a significant role in healthcare, providing services such as electronic health records, digital imaging, and e-prescribing services [9]. This not only improves the efficiency of healthcare operations but also helps to ensure the high quality of patient care [10]. To improve the technical efficiency of flow monitoring, many airports have introduced new digital solutions in their operations [11]. In the aspect of urban governance, city managers collect data from daily activities, such as parking lots, to generate data networks. After analyzing them in real-time, the generated predictive models will be used for future planning [12]. Similarly, road data are collected for its condition assessment, which helps eliminate potential safety hazards and ensure driving safety [13]. Figure 1 shows the basic concept of the future smart road technology, consisting of a combination of information, electrical, and transportation networks [14].

DX has also transformed the way many industries operate today. It helps ensure a business's competitive advantage, productivity, and execution efficiency [15]. The technologies used in DX can improve production efficiency, enhance industrial value, and build innovative models of business operations [16]. Data in agriculture, transportation, and other industries can be obtained more accurately and efficiently, mak-

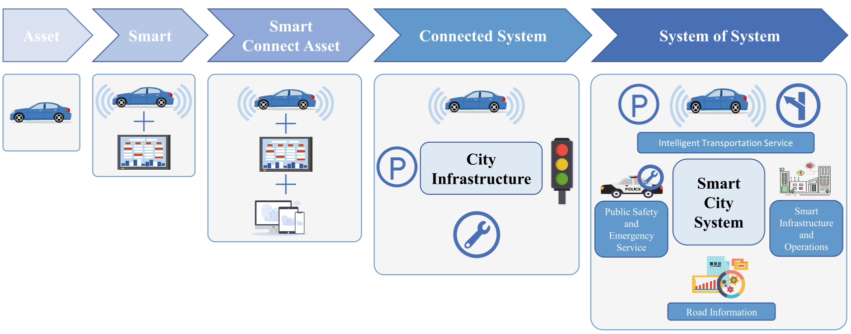


Fig. 1 An illustration of future smart road technology [8]

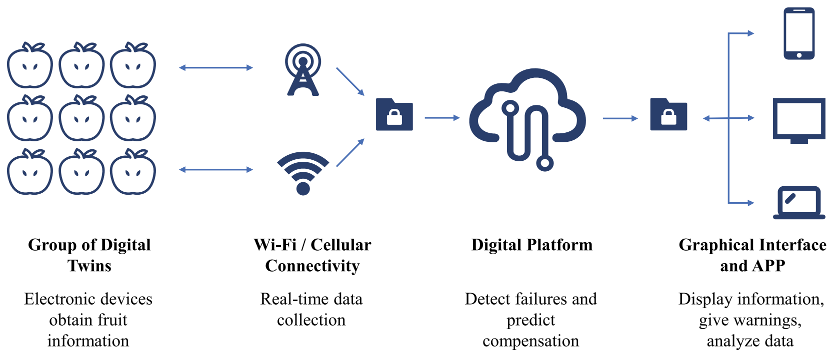


Fig. 2 An example of the DT application in fruits and vegetable transportation [21]

ing the systems automated and more intelligent [17]. Many industries have been affected by the pandemic of COVID-19 in recent years. Digitalized systems requiring minimal human intervention have, therefore, become more crucial than ever [18].

Digital twins (DT) technology, which is a crucial component of the digital transformation, uses innovative approximate methods to extract, analyze and upgrade relevant features [19]. DT technology has been extensively applied in many fields. It consists of (1) the physical information in real scenarios, (2) virtual replicas of the real scenarios, and (3) the relationship between these two states [20]. Learning and predicting the behavior of physical entities is typically implemented by artificial intelligence and other information technology [18]. Figure 2 illustrates an example of DT technology applied in the agricultural field, where smart electronic devices are transported alongside the fruits and they constantly monitor the state of the fruits in real-time through sensors. The collected information is then sent to the cloud platform for fault detection and defects analysis.

1.2 Smart City

As one of the most important applications of DX, smart city has been a feasible solution to modern infrastructure maintenance, such as intelligent road condition assessment. A smart city is typically a city in which information and communication technologies (ICTs) are combined with traditional infrastructures, allowing intelligent planning and management of its resources [22]. Smart city technology allows city managers to monitor the city's development in time, to know what is going on in the city, and to regulate the infrastructure [23]. The quality of urban services and the linkage between citizens and government have been enhanced with ICT [24]. At the same time, as an essential strategy, the smart city will also integrate the Internet of Things (IoTs) technologies to improve public infrastructure, services, and utilities.

IoTs can connect numerous aspects of life into an organic entirety. Incorporating data-driven applications can help provide safe and intelligent services in smart cities [25]. Smart cities are expected to improve liveability and make production management more efficient and accurate [26].

The key success of smart city mainly relies on improving fundamental public services and utilities, developing an intelligent city network platform and improving public transportation [27]. Among these, a reliable and high-performing telecommunication infrastructure is an essential component of the smart city that will help in monitoring resources and providing quality services such as medical treatment. Built upon fast optical fiber networks, wireless broadband technologies such as WiMAX provide comprehensive coverage, which will facilitate the management of city's resources and allow citizens to access services and information wherever they are [28]. The construction of smart cities also plays an influential role in the field of innovative medical care [29]. Hospitals can establish digital systems to manage medical supplies [30]. Regarding the management of resources, city managers can implement a green city platform to monitor resources. In conclusion, smart cities have a bright prospect [31].

Driven by the DX technology, extensive interdisciplinary research on smart cities using data science techniques is receiving increasing attention, especially in computer vision-based intelligent road condition assessment. Each city establishes an intelligent traffic management system according to its traffic conditions. In such a system, the use of sensor networks, IoTs, and other technologies can realize adaptive traffic signal control as part of urban construction, and ultimately facilitate the life of citizens [31].

1.3 Intelligent Road Condition Assessment

Road condition assessment is a requisite task to ensure maintainability of road networks while providing maximum security for users [32]. Manual visual inspection by professional inspectors is still the main form of road condition assessment [33]. However, this process is subjective, costly, tedious, and time-consuming [13]. Therefore, there is an increasing demand for intelligent road condition assessment systems developed based on digital twin techniques [34].

An intelligent road condition assessment system typically consists of two main components: (1) road data acquisition, and (2) road defect detection, as illustrated in Fig. 3. Remote sensing, vibration sensing, and computer vision are commonly used for intelligent road condition assessment [32]. The remote sensing methods which have been utilized in unmanned aerial vehicles, satellites, multi-purpose survey vehicles, or aeroplanes have indeed reduced the workload of road inspectors. Nevertheless, the traditional geotechnical methods can never be entirely replaced by the remote sensing approaches [32]. Vibration-based methods generally use accelerometers and GPS for multi-modal road data collection [35, 36]. However, such methods

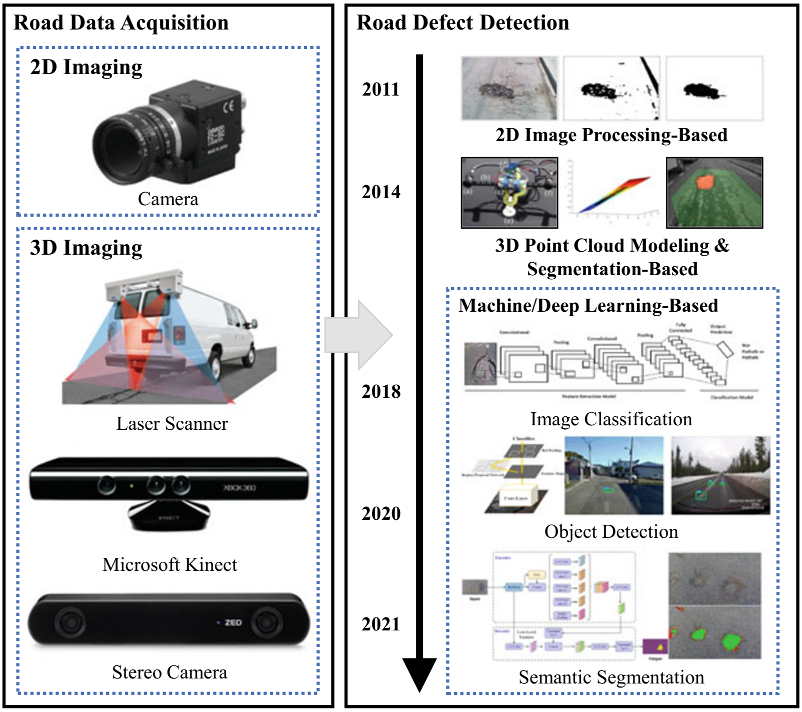


Fig. 3 Computer vision techniques for road data acqusition and defect detection [37]

always suffer from defect misdetection in spite of their advantages in terms of cost-effectiveness, small storage requirements, and real-time performance [32].

Before assessing the condition of the road, it is necessary to acquire road data using 2D imaging technology through cameras and distance sensors [38]. However, for 2D road images without overlapping regions, it is not easy to acquire the 3D geometry of the road surface [39]. Furthermore, these methods are severely affected by the environment, especially the lighting conditions [40]. In this regard, some researchers [32, 33] use laser scanners, Microsoft Kinect sensors, and stereo cameras to collect road data with the help of 3D imaging technology. In addition to the aforementioned three common types of 3D imaging technologies, photometric stereo [41, 42], interferometry [43, 44], structured light imaging [45, 46], and time-of-flight (ToF) [47, 48] are other alternatives for 3D geometry reconstruction. However, they were not widely used for 3D road data acquisition. The interested reader is referred to [49] for a detailed description of these technologies (both theoretical and practical aspects are covered).

Road defect detection is intrinsic to the intelligent road condition evaluation system. The road defect detection methods can be divided into: 2D image processing-based [50], 3D point cloud modeling and segmentation-based [51], and machine/deep learning-based [52]. Algorithms based on 2D image processing process road RGB or parallax/depth images through enhancement, compression, transformation, segmentation, *etc.*. 3D road point cloud modeling and segmentation-based algorithms build a geometric model from the observed road point cloud which is segmented by comparing the fitted and observed surfaces [53]. Machine/deep learning-based algorithms use image classification, object recognition, or semantic segmentation networks to solve the problem of road pothole detection.

2 Road Data Acquisition

The application of 2D imaging technologies for road data collection started as early as 1991 [39, 54]. However, the spatial structure cannot be explicitly illustrated in 2D road images [32]. Therefore, many researchers [32–34, 55] have resorted to 3D imaging technologies, which can better reflect the geometric characteristics of road defects and improve the robustness of intelligent road condition assessment systems.

With the development of digital transformation, various 3D imaging techniques have been used to calculate the depth of a road scene and reconstruct its 3D geometry model. The first reported effort [56] on leveraging 3D imaging technology for road data collection can be traced back to 1997. This section discusses four prevalent 3D road imaging technologies: laser scanning, infrared sensing, multi-view geometry, and shape (depth) from focus. Other 3D road imaging technologies are detailed in [49].

2.1 Laser Scanning

Laser scanning is a well-established imaging technology for accurate 3D road data acquisition. This technology is developed based on trigonometric triangulation [49]. As shown in Fig. 4, the laser receiver is located at a known distance from the laser's emitter. Accurate point measurements can be made by calculating the reflection angle of the laser light. Auto-synchronized triangulation [56] is a popular variation of classic trigonometric triangulation, and it has been widely utilized in laser scanners to capture the 3D geometry information of near-flat road surfaces [49]. However, laser scanners must be mounted on dedicated intelligent road condition assessment vehicles, such as a Georgia Institute of Technology Sensing Vehicle [57]. Nevertheless, such vehicles are not commonly used for intelligent road condition assessment because they are costly, and so is their routine maintenance [34].

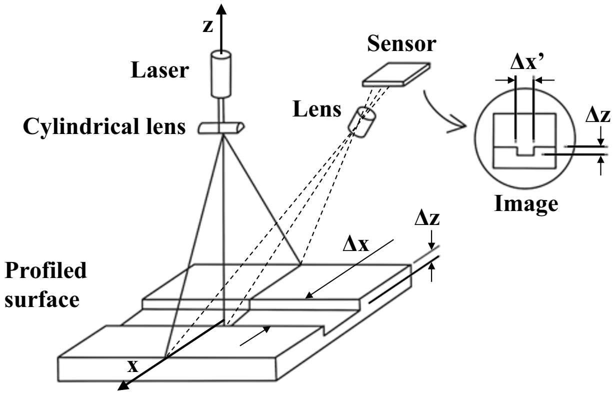
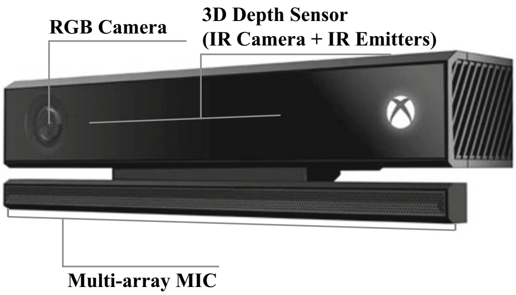


Fig. 4 Triangulation in laser scanners [58]

2.2 *Infrared Sensing*

Microsoft Kinect sensors [40], as illustrated in Fig. 5, are the most frequently used infrared sensors for road data acquisition. It was initially designed for the Xbox-360 motion-sensing games. The operating 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 [40, 59, 60] on 3D road data acquisition and defect detection using Microsoft Kinect sensors. However, the main downside of Microsoft Kinect sensors is that they suffer considerably from infrared saturation in direct sunlight [53].

Fig. 5 Microsoft Kinect sensor



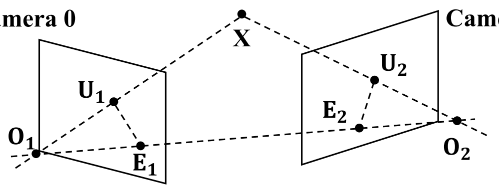


Fig. 6 Multi-view geometry [49]

2.3 Multi-view Geometry

The 3D geometry of a road surface can also be reconstructed using multiple images captured from different views [32]. The theory behind this 3D imaging technique is typically known as *multi-view geometry* [61], as shown in Fig. 6. These images can be captured using a single movable camera [62] or an array of synchronized cameras [33]. In both cases, dense correspondence matching (DCM) between two consecutive road video frames or between two synchronized stereo road images is the main problem to be solved.

Structure from motion (SfM) [63] and optical flow (OF) estimation [64] are the two most commonly used techniques for monocular correspondence matching, as illustrated in Fig. 7. SfM method estimates camera pose and 3D points of interest from a series of images taken from different viewpoints [65]. SfM typically starts with feature extraction and matching, followed by geometric validation. It then uses the resulting scene graph as the basis for the 3D reconstruction procedure that triangulates scene points, filters out outliers, and refines the reconstruction using bundle adjustment (BA) [66]. OF describes the motion of pixels between consecutive frames of a video sequence. OF method obtains the motion speed (or distance in recent deep learning-based works) of pixels between images based on the assumption that the correspondence intensity is constant. OF estimation methods replace the calculation of keypoints and descriptors in SfM, demonstrating superb (dense) results even in the case of limited texture information [67].

Stereo vision (also known as binocular vision, stereo matching, or disparity estimation) [32, 34] is typically employed for binocular DCM. Stereo vision acquires depth information by finding the horizontal positional differences (disparities) of the visual feature correspondence pairs between two synchronously captured road images [32]. Stereo vision has high research value and broad application fields because it can calculate the absolute scale to establish more accurate 3D models, without having to consider the complex environmental hypotheses [38].

The traditional stereo vision algorithms formulate disparity estimation as either a local block matching problem [32] or a global/semi-global energy minimization problem (solvable with various Markov random field-based optimization techniques) [68, 69], while data-driven algorithms typically solve stereo matching with convolutional neural networks (CNNs) [70]. Despite the low cost of digital cameras, stereo vision accuracy is always affected by various factors, most notably by poor illumi-

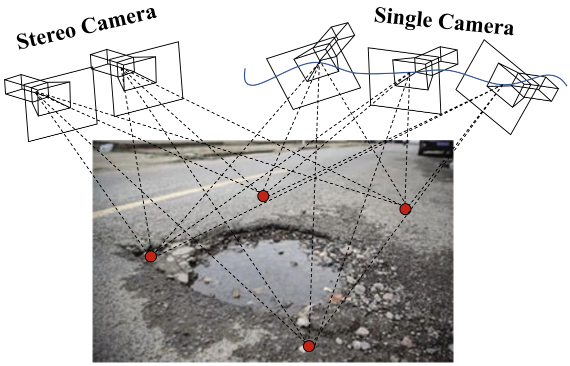


Fig. 7 3D road imaging with camera(s) [37]

nation conditions [53]. Furthermore, self/un-supervised learning [70, 71] that does not rely on human-labeled training data is the future of deep stereo matching. Moreover, the feasibility of stereo vision algorithms relies on the well-conducted stereo rig calibration [32]. Therefore, some systems incorporate stereo rig self-calibration functionalities to ensure their captured stereo road images are always well calibrated.

2.4 *Shape (depth) from Focus*

As shown in Fig. 8, the position of point O can determine the depth z_0 when fully focused. Point A and point B are located at positions Δz and $-\Delta z$ on either side of the focal plane. There are two situations when the point source is imaged through the lens: focusing on the image plane or out of focus. When a point light source cannot be focused, its image on the imaging plane is no longer a point, but a circle called the circle of confusion. In Fig. 8, points A and B show 'circle of confusion circles' with radii c_1 and c_2 in the imaging plane respectively. Although the circle of confusion makes the image appear blurry, the image at point O is obvious. We can use this feature to identify clear points in the image, called Shape from focus (SFF). SFF uses many 3D images taken at different focal lengths [72].

Among them, SFF is based on the same scene, and the depth information is estimated by taking images with different settings of cameras to ensure that the focal length of each image has a different value. At the same time, local focus changes in the image are used as depth cues. The focal length measurement (FM) operator is usually used to measure the focal length level of each image pixel [74]. The length between any two planes on either side of the focal plane is usually defined as the depth of field [75]. Any blur smaller than the pixel size will not be detected, and the confusion circle is within the acceptable range. Therefore, a sufficiently low depth of field is necessary to make the SFF system more accurate.

The SFF method is widely used in autofocus cameras today. This approach tends to rely on surface textures to perceive depth information. However, in many practical

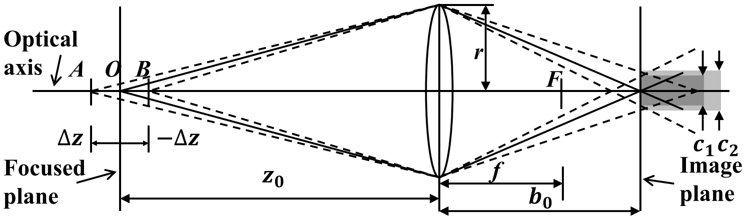


Fig. 8 Depth variation and resulting circles of confusion [73]

applications, there may be cases where the object's surface is smoothly shaded or lacks detectable texture. This situation can lead to poor results generated by SFF, resulting in inaccurate and sparse disparity maps. In this situation, the original SFF method can be extended to force a firm texture on the imaging surface by using dynamic lighting, making it suitable for weakly textured or untextured surfaces [76].

3 Road Defect Detection

In our recently published survey paper article [37], the state-of-the-art (SOTA) road defect detection approaches are categorized as (1) traditional 2D image processing-based, (2) 3D road surface modeling and segmentation-based, (3) machine/deep learning-based, and (4) hybrid.

3.1 Traditional 2D Image Processing-Based Approaches

The traditional image processing-based road defect detection approaches typically consist of four main procedures: (1) image pre-processing, (2) image segmentation, (3) shape extraction, and (4) object recognition [77]. [78] is a representative example of this algorithm type. It first uses a histogram-based thresholding algorithm [50] to segment (binarize) the input gray-scale road images. The segmented road images are then processed with image filters, such as a median filter and morphology operators, to reduce the redundant noise. Finally, the defective areas are extracted by analyzing a pixel intensity histogram. Additionally, [50] solved road defect detection with similar techniques. It first applies the triangle algorithm [79] to segment gray-scale road images. An ellipse then models the boundary of an extracted potential road defect.

Researchers also performed image segmentation algorithms on depth/disparity images for road defect detection. This allows the intelligent road condition assessment system to identify road defects with a higher accuracy [40, 57, 80]. For instance, [40] acquires the depth images of pavements using a Microsoft Kinect sensor. The depth images are then segmented using the wavelet transform algorithm [81]. In their practical experiments, the Microsoft Kinect sensor must be mounted as perpendicularly as possible to the pavements. In 2018, [57] proposed to detect road defects from depth images acquired using a highly accurate laser scanner mounted on a road inspection 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 the watershed method [82] for road defect detection. In 2019, [80] leveraged a so-called disparity transformation algorithm to process dense road disparity images, making the road defects highly distinguishable. The road defects are then detected by applying a histogram-based thresholding method on the transformed disparity images.

3.2 Machine/Deep Learning-Based Approaches

The CNN-based approaches typically address the road defect detection problem in an end-to-end manner, based on image classification, object inspection, or semantic segmentation networks [37].

3.2.1 Image Classification Networks

Since 2012, deep CNNs have been the first choice for many image classification problems. The image classification networks have been widely used to detect, recognize and localize road cracks, as the features learned by CNNs can replace the traditional hand-crafted features [83]. For example, in 2016, [84] proposed a robust road crack detection network. An RGB road image is fed into a CNN consisting of a collection of convolutional layers. The learned feature serves as input to a fully connected (FC) layer to produce a scalar indicating the probability that the image contains road cracks.

In 2019, [85] proposed a self-supervised monocular road defect detection algorithm, which can not only reconstruct the 3D 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 unimpaired or defective with a classification CNN (the images used for network training were automatically annotated via a 3D road point cloud thresholding algorithm).

Recently, [87] conducted a comprehensive comparison among 30 SOTA image classification CNNs for road crack detection. The extensive experiments demonstrate that (a) the performances achieved by these deep CNNs are very similar [87]; (b) Learning road crack detection does not require a large amount of training data (10,000 images are sufficient to train a well-performing CNN [87]). However, the pre-trained CNNs typically perform unsatisfactorily on additional test sets. Therefore, unsupervised domain adaptation (UDA) [88], capable of aligning two different domains, has become a hot research topic that requires more attention. Furthermore,

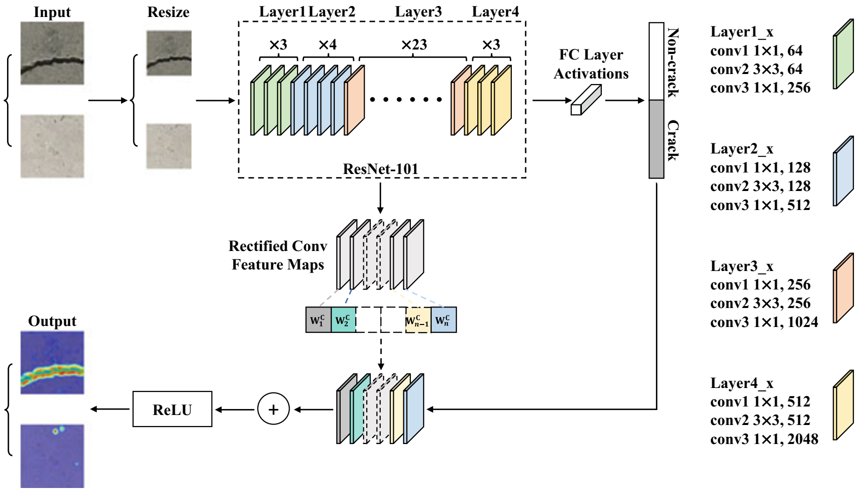


Fig. 9 An example of CNN's decision-making explanation with Grad-CAM++ [86]

explainable AI algorithms, such as Grad-CAM++ [89] (an example is shown in Fig. 9), have become essential for image classification applications to understand CNN's decision-making.

3.2.2 Object Detection Networks

Region-based CNN (R-CNN) [90–92] series and you only look once (YOLO) [93–96] series are two representative groups of modern deep learning-based object detection algorithms for road defect detection. As discussed above, the SOTA image classification CNNs can categorize an input image into a specific class. However, we sometimes attempt to obtain the exact location of a specific object in the image, which requires the CNNs to have the ability 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 (usually taken from the original image) as negative (the object is absent from the patch) or positive (the object is present in the patch). However, such an approach is usually costly as it selects many regions of interest (RoIs).

To solve this dilemma, R-CNN [90] employs a selective search algorithm [97] to extract only 2,000 RoIs (generally known 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 classes, as shown in Fig. 10. In 2015, Fast R-CNN [91] was introduced as a faster object detection algorithm based on R-CNN. Instead of feeding region proposals to the classification CNN, Fast R-CNN [91] directly feeds the original image to the CNN and produces a convolutional feature map (CFM). The region proposals are identified from the CFM using selective search, which are

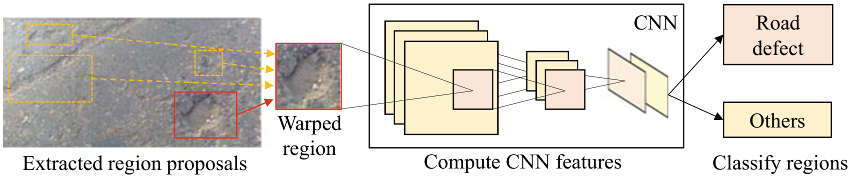


Fig. 10 R-CNN [90] for road defect detection

then reshaped into a fixed size with an RoI pooling layer before being fed into an FC layer. A softmax layer is subsequently utilized to predict the region proposals' classes. Nevertheless, both R-CNN [90] and Fast R-CNN [91] employ a selective search algorithm to determine region proposals, which is time-consuming and slow. To overcome this disadvantage, [92] proposes Faster R-CNN, which leverages a network to learn the region proposals.

The YOLO series, on the other hand, is very different from the R-CNN series. YOLOv1 [93] formulates object detection as a regression problem. It first splits an image into an $M \times M$ grid, from which B bounding boxes are selected. The network then produces class probabilities and offset values for these bounding boxes. Since YOLOv1 [93] makes a significant number of localization errors and its achieved recall is unsatisfactory, YOLOv2 [94] was proposed to improve YOLOv1 [93] in the following respects: (1) it fine-tunes the classification network at the full 448×448 resolution; (2) it adds batch normalization on all the convolutional layers; (3) it removes the fully connected layers from YOLOv1 [93] and uses anchor boxes to predict bounding boxes. In 2018, [95] made several tweaks to further improve YOLOv2.

The aforementioned object detectors have been extensively applied for road defect detection. [98] used a Faster R-CNN [92] to recognize road potholes, while [99] trained a YOLOv2 [94] object detector to achieve the same objective. [100] utilized YOLOv3 [95] to detect road potholes from RGB images captured by a smartphone mounted on a car. [101] employed three different YOLOv3 [95] architectures to detect road potholes. Recently, [102] compared the performances of YOLOv2 [94] and Mask R-CNN [103] for road defect detection. Nonetheless, these approaches typically leverage an unmodified object detector which has been specifically designed for the task of road defect detection. Furthermore, these object detectors can only provide instance-level predictions instead of pixel-level predictions. Therefore, in recent years, semantic segmentation has become a more desirable technique for pixel-level road defect detection.

3.2.3 Semantic Segmentation Networks

Semantic segmentation aims at assigning each image pixel with a specific class [104]. It has been widely used for road defect detection in recent years. According

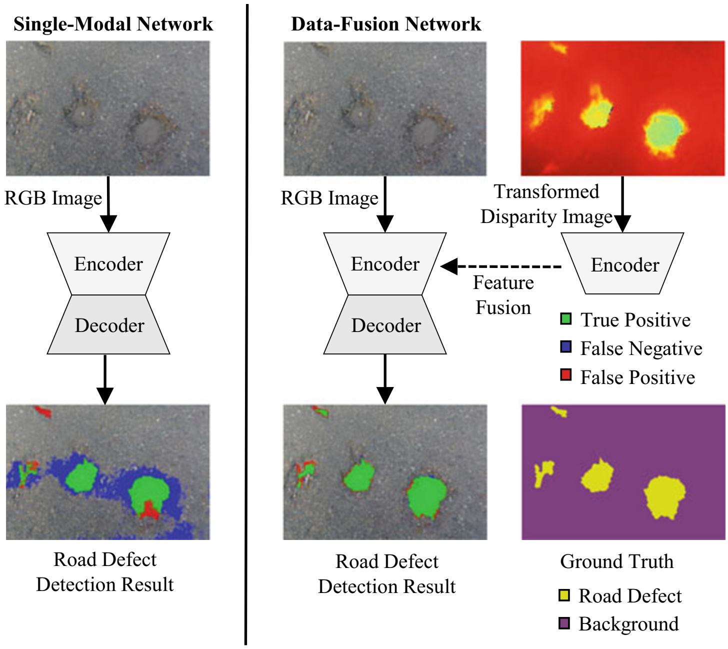


Fig. 11 Single-modal semantic segmentation versus data-fusion semantic segmentation for road defect detection

to the number of encoders, the SOTA semantic image segmentation networks can be grouped into two categories: (1) single-modal [105] and (2) data-fusion [106, 107], as shown in Fig. 11. The former inputs only one type of vision sensor data, such as RGB images or depth images, while the latter typically learns visual features from different types of vision sensor data, such as RGB images and transformed disparity images [108] or RGB images and surface normal information [106]. Researchers have employed both types of semantic segmentation networks to detect road defects/anomalies (the approaches designed to detect road defects can also be leveraged to detect road anomalies, as these two types of objects are below and above the road, respectively).

Recently, [108] developed a data-fusion semantic segmentation CNN for road anomaly detection. The authors also conducted a comprehensive comparison of data fusion performance for different modalities of visual features, including (1) RGB images, (2) disparity images, (3) surface standard images [109], (4) elevation images, (5) HHA images, and (6) transformed disparity images [80]. The extensive experimental results demonstrated that the transformed disparity image is the most informative visual feature in terms of road defect/anomaly detection.

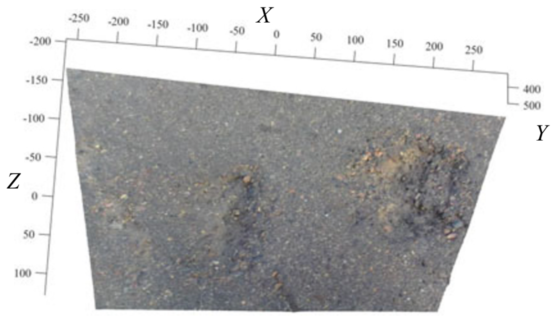


Fig. 12 An example of 3D road pothole cloud acquired using the stereo matching algorithm introduced in [34]. The accuracy of the 3D road point cloud is around 2.23 mm

3.3 3D Road Surface Modeling and Segmentation-Based Approaches

The 3D road surface modeling and segmentation-based approaches [53, 110–112] typically formulate the 3D road point clouds (an example is shown in Fig. 12) as an explicit model $f(\mathbf{p}, \mathbf{a})$, where $\mathbf{p} = (X; Y; Z)$ is a 3D point on the road surface in the camera coordinate system and $\mathbf{a}$ stores the model coefficients. The optimal $\mathbf{a}$ can be obtained by minimizing [110]:

$$E = \sum_{i=1}^{n}(f(\mathbf{p}_i, \mathbf{a}) - Y_i). \tag{1}$$

By comparing the difference (namely, $f(\mathbf{p}, \mathbf{a}) - Y$) between the actual and modeled road surfaces, the road defects can be extracted. Nevertheless, the 3D road surface modeling process is very sensitive to outliers. Hence, [112] employs random sample consensus (RANSAC) [113] to further improve its robustness. Furthermore, the actual road surface is sometimes uneven, which makes quadratic surface modeling somewhat problematic [34].

3.4 Hybrid Approaches

In 2017, [114] proposed an intelligent road defect detection system based on the analysis of 2D LiDAR data (providing road profile information) and RGB road images (providing road texture information). To obtain a large field of view, this system uses two LiDARs. This hybrid system is pretty robust to electromagnetic waves and poor road conditions. Such a hybrid system can combine multi-modal vision sensor data's advantages to improve the accuracy of road defect detection.

In 2019, [53] also introduced a hybrid framework for road defect detection. It first applies disparity transformation, discussed in Sect. 3.1, and histogram-based thresholding method to extract potential road defects. In the next step, a quadratic surface was fitted to the original disparity image, where the RANSAC algorithm was employed to enhance the surface fitting robustness. Road defects can be successfully detected by comparing the actual and fitted disparity images. This hybrid framework uses a 2D image processing algorithm along with a 3D road surface modeling algorithm to significantly improve the road defect detection performance.

4 Future Insights

DX technology has a bright future in many fields of study. With the use of DX techniques, smart cities can optimize their service efficiency. Ericsson, IBM, Cisco, Microsoft, and many other companies are also primarily involved in the digital revolution. These companies are conducting innovative research on how to apply digital technologies (such as wireless network sensors, digital twins, semantic web, and cloud computing) to address the challenges facing future cities [115].

[116] defined ten pillars of smart city that drive benefits to the digital age, as shown in Fig. 13. There are five fundamental pillars– governance, economy, talent, funding, and infrastructure [117]:

- Building an intelligent city first requires wise governance. Creating a technology-enabled vision and developing a coherent implementation plan is the beginning of a smart city.
- Only when smart cities have a practical economic development plan can they find their place in global trade and promote a good economic environment.
- Smart cities are centers of talent. How to attract talents and establish a creative cultural center is the lifeblood of innovative city development.
- To keep pace with the digital age, cities must be more innovative in their financing techniques, capital sources, budgeting methods, and business models.
- The construction of interconnected and well-maintained infrastructure is the foundation for smart cities to expand their digital areas.

Fig. 13 The ten pillars of smart city transformation [117]

In addition to the foundational pillars, the five science and technology pillars are also critical– mobility, environment, public safety, public health, and payment systems [117]:

- Smart city layouts in the future should be multi-modal, involving various options that are fully integrated and connected.
- As smart cities mature, environmental sustainability issues, energy use, resource allocation, and climate hazards should also be addressed.
- The creation of a harmonious, safe, and healthy environment also underpins a flourishing smart city.
- Moreover, Digital payment systems will not only help businesses to reduce costs but also increase revenue.

Transportation is the artery of modern society and the economy [118]. After obtaining the road data using the sensing technologies mentioned above, the department of highways and surrounding traffic management can use the data for digital transformation [119]. Extensive interdisciplinary research using data science technologies is receiving more and more attention [120]. However, the following issues need to be taken into consideration in the future DX of smart city transportation [121]:

- Improve the efficiency of the entire process. In order to make the traffic assessment prediction more timely, the time spent on data collection and analysis should be as short as possible [122]. Dedicated intelligent road condition assessment vehicles were devised to make data collection more accessible and faster. After obtaining road data, relevant departments should also be able to carry out digital management, where road restoration and traffic control work should be performed collectively to benefit end-users.
- Improve the accuracy of intelligent road condition assessment. The road defect detection algorithms with deep learning models have achieved appealing results. However, such models are generally developed based on supervised learning. Labeling road data is a highly labor-intensive and time-consuming process [34]. Therefore, developing un/self/semi/low-shot learning strategies for road defect detection has become a popular area of research that requires more attention.
- Expand the service scope for the digital outreach. The service group is not only limited to roadway engineers but also to all groups involved in the whole road condition assessment process. Ultimately, strategic decisions can be transparent and open, and the strategic decision implementation results can be traced.
- Enhance information sharing and cooperation while establishing data fusion models [123]. By adopting a unified modular data interface, participants can implement the interoperability and consistent data standards to share data resources from the city context [124]. In the future, an interdisciplinary and trans-department digital urban transport system will be established.

- Improve network security. There are many privacy and security risks in smart cities, which need to be rethought and improved [125]. The rise of smart cities entails considerable risks. In order to prevent malicious use of road data, it is imperative to enhance the anti-attack capability of the gateway. Therefore, in the early stage of innovative city development, smart cities must be fully prepared for cyberattacks.

5 Summary

Digital transformation has been extensively used to improve the efficiency and accuracy of road condition assessment. This chapter gave readers an overall picture of the state-of-the-art road condition assessment systems. The state-of-the-art road data collection technologies, including remote sensing, vibration sensing, laser scanning, infrared sensing, multi-view geometry, and shape (depth) from focus, were first discussed. The existing machine vision and intelligence approaches, including classical 2D image processing algorithms, 3D road surface modeling and segmentation algorithms, machine/deep learning algorithms, and hybrid algorithms, designed to detect road defects were then presented. Finally, future insights on the digital transformation of smart cities were discussed. Establishing the digital transformation of road inspection with 3D road imaging technology and combining different types of machine vision and intelligence algorithms for road defect detection are the future trends in this research area.

References

1. Bohnsack R et al (2022) Sustainability in the digital age: intended and unintended consequences of digital technologies for sustainable development, pp 599–602
2. Elliott A (2021) Contemporary Social Theory: An Introduction. Routledge, London
3. Piepponen A et al (2022) Digital transformation of the value proposition: a single case study in the media industry. J Bus Res 150:311–325
4. Lasi H et al (2014) Industry 4.0. Bus Inf Syst Eng 6(4):239–242
5. Nandico OF (2016) A framework to support digital transformation. In: El-Sheikh E, Zimmermann A, Jain LC (eds) Emerging Trends in the Evolution of Service-Oriented and Enterprise Architectures, vol 111. ISRL. Springer, Cham, pp 113–138. https://doi.org/10.1007/978-3-319-40564-3_7
6. Henriette E et al (2015) The shape of digital transformation: a systematic literature review. In: MCIS 2015 Proceedings, vol 10, pp 431–443
7. Goran J et al (2017) Culture for a digital age. McKinsey Q 3(1):56–67
8. Davenport TH et al (2019) Artificial Intelligence: The Insights You Need From Harvard Business Review. Harvard Business Press, Boston
9. Kodama M (2020) Digitally transforming work styles in an era of infectious disease. Int J Inf Manag 55:102172
10. Haggerty E (2017) Healthcare and digital transformation. Netw Secur 2017(8):7–11

11. Gardy A et al (2016) Digital trends & opportunities for airports. In: ACI-NA World Annual Conference
12. Scriney M, Roantree M (2016) Efficient cube construction for smart city data. In: EDBT/ICDT Workshops, vol 2016
13. Fan R, Wang H, Bocus MJ, Liu M (2020) We learn better road pothole detection: from attention aggregation to adversarial domain adaptation. In: Bartoli A, Fusiello A (eds) ECCV 2020, vol 12538. LNCS. Springer, Cham, pp 285–300. https://doi.org/10.1007/978-3-030-66823-5_17
14. Yesner R (2017) Accelerating the digital transformation of smart cities and smart communities, Microsoft, October 2017
15. Bordeleau F-È, Felden C (2019) Digitally transforming organisations: a review of change models of industry 4.0. In: European Conference on Information Systems
16. Schmarzo B (2017) What is digital transformation. Accessed 8 Mar 2018
17. Ebert C, Duarte CHC (2018) Digital transformation. IEEE Softw 35(4):16–21
18. Erol T (2020) Digital transformation revolution with digital twin technology. In: 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), pp 1–7. IEEE
19. Tao F et al (2019) Digital Twin Driven Smart Manufacturing. Academic Press, Cambridge
20. Catarci T (2019) A conceptual architecture and model for smart manufacturing relying on service-based digital twins. In: 2019 IEEE International Conference on Web Services (ICWS), pp 229–236. IEEE
21. The digital twin IoT platform. https://www.digitaltwincorporation.com/. Accessed 12 May 2022
22. Batty M et al (2012) Smart cities of the future. Eur Phys J Spec Top 214(1):481–518
23. Zhao F et al (2021) Smart city research: a holistic and state-of-the-art literature review. Cities 119:103406
24. Forward N (2016) Building a smart city, equitable city
25. Sarker IH (2022) Smart city data science: towards data-driven smart cities with open research issues. Internet Things 19:100528
26. Zhang Y (2010) Interpretation of smart planet and smart city [j]. China Inf Times 10:38–41
27. Nam T, Pardo TA (2011) Conceptualizing smart city with dimensions of technology, people, and institutions. In: Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, pp 282–291
28. Yovanof GS, Hazapis GN (2009) An architectural framework and enabling wireless technologies for digital cities & intelligent urban environments. Wirel Pers Commun 49(3):445–463
29. Chu Z et al (2021) A smart city is a less polluted city. Technol Forecast Soc Chang 172:121037
30. Xue Q (2010) Smart healthcare: applications of the internet of things in medical treatment and health. Inf Constr 2010(5):56–58
31. Su K (2011) Smart city and the applications. In: 2011 International Conference on Electronics, Communications and Control (ICECC), pp 1028–1031. IEEE
32. Fan R et al (2018) Road surface 3D reconstruction based on dense subpixel disparity map estimation. IEEE Trans Image Process 27(6):3025–3035
33. Fan R et al (2019) Real-time dense stereo embedded in a UAV for road inspection. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE Computer Society, pp 535–543
34. Fan R et al (2021) Rethinking road surface 3-D reconstruction and pothole detection: from perspective transformation to disparity map segmentation. In: IEEE Transactions on Cybernetics, 2021. https://doi.org/10.1109/TCYB.2021.3060461
35. De Zoysa K, Keppitiyagama C, Seneviratne GP, Shihan W (2007) A public transport system based sensor network for road surface condition monitoring. In: Proceedings of the 2007 Workshop on Networked Systems for Developing Regions, pp 1–6
36. Eriksson J, Girod L, Hull B, Newton R, Madden S, Balakrishnan H (2008) The pothole patrol: using a mobile sensor network for road surface monitoring. In: Proceedings of the 6th International Conference on Mobile Systems, Applications, and Services, pp 29–39

37. Ma N et al (2022) Computer vision for road imaging and pothole detection: a state-of-the-art review of systems and algorithms. In: Transportation Safety and Environment. https://doi.org/10.1093/tse/tdac026
38. Fan R (2021) Long-awaited next-generation road damage detection and localization system is finally here. In: 2021 29th European Signal Processing Conference (EUSIPCO), pp 641–645. IEEE
39. Mahler DS et al (1991) Pavement distress analysis using image processing techniques. Comput Aided Civ Infrastruct Eng 6(1):1–14
40. Jahanshahi MR et al (2013) Unsupervised approach for autonomous pavement-defect detection and quantification using an inexpensive depth sensor. J Comput Civ Eng 27(6):743–754
41. Woodham RJ (1980) Photometric method for determining surface orientation from multiple images. Opt Eng 19(1):191139
42. Barsky S, Petrou M (2003) 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
43. Fujimoto JG et al (2000) Optical coherence tomography: an emerging technology for biomedical imaging and optical biopsy. Neoplasia 2(1–2):9–25
44. Walecki WJ, Van P (2006) Determining thickness of slabs of materials by inventors, 3 Oct 2006, uS Patent 7,116,429
45. Scharstein D, Szeliski R (2003) High-accuracy stereo depth maps using structured light. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings., vol 1, p I. IEEE
46. Muzet V et al (2009) Surface deflection measurement using structured light. Testing in Civil Engineering, Nantes, France
47. Oggier T et al (2004) 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, pp 534–545. International Society for Optics and Photonics
48. Anderson D et al (2005) Experimental characterization of commercial flash ladar devices. In: 2005 International Conference of Sensing and Technology, vol 2, pp 17–23. Citeseer
49. Mathavan S et al (2015) A review of three-dimensional imaging technologies for pavement distress detection and measurements. IEEE Trans Intell Transp Syst 16(5):2353–2362
50. Koch C, Brilakis I (2011) Pothole detection in asphalt pavement images. Adv Eng Inform 25(3):507–515
51. Chang K et al (2005) Detection of pavement distresses using 3D laser scanning technology. Comput Civ Eng 2005:1–11
52. Lin J, Liu Y (2010) Potholes detection based on SVM in the pavement distress image. In: 2010 Ninth International Symposium on Distributed Computing and Applications to Business, Engineering and Science, pp 544–547. IEEE
53. Fan R et al (2019) Pothole detection based on disparity transformation and road surface modeling. IEEE Trans Image Process 29:897–908
54. Koutsopoulos HN et al (1993) Primitive-based classification of pavement cracking images. J Transp Eng 119(3):402–418
55. Fan R et al (2018) Real-time stereo vision for road surface 3-D reconstruction. In: 2018 IEEE International Conference on Imaging Systems and Techniques (IST), pp 1–6. IEEE
56. Laurent J et al (1997) Road surface inspection using laser scanners adapted for the high precision 3D measurements of large flat surfaces. In: Proceedings. International Conference on Recent Advances in 3-D Digital Imaging and Modeling (Cat. No. 97TB100134), pp 303–310. IEEE
57. Tsai Y-C et al (2018) Pothole detection and classification using 3D technology and watershed method. J Comput Civ Eng 32(2):04017078
58. IIT: Lasers in quality controlline. http://www.iitk.ac.in/celt/lecture_laser/. Accessed 2 Apr 2022
59. Joubert D et al (2011) Pothole tagging system. In: 2011 th Robotics and Mechatronics Conference of South Africa, CSIR International Conference Centre, Pretoria, pp 23–25

60. Moazzam I et al (2013) Metrology and visualization of potholes using the microsoft kinect sensor. In: 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013), pp 1284–1291. IEEE
61. Andrew, AM (2001) Multiple view geometry in computer vision, Kybernetes
62. Jog G et al (2012) Pothole properties measurement through visual 2D recognition and 3D reconstruction. Comput Civ Eng 2012:553–560
63. Ullman S (1979) The interpretation of structure from motion. Proc R Soc Lond Ser B. Biol Sci 203(1153):405–426
64. Wang H et al (2021) CoT-AMFlow: adaptive modulation network with co-teaching strategy for unsupervised optical flow estimation. In: Conference on Robot Learning (CoRL), pp 143–155. PMLR
65. Schonberger JL, Frahm J-M (2016) Structure-from-motion revisited. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 4104–4113
66. B. Triggs *et al.*, "International workshop on vision algorithms," *Bundle adjustment–a modern synthesis*, pp. 298–372, 1999
67. Fleet D, Weiss Y (2006) Optical flow estimation. In: Paragios N, Chen Y, Faugeras O (eds) Handbook of Mathematical Models in Computer Vision, pp 237–257. Springer, Boston, MA. https://doi.org/10.1007/0-387-28831-7_15
68. Hirschmuller H (2007) Stereo processing by semiglobal matching and mutual information. IEEE Trans Pattern Anal Mach Intell 30(2):328–341
69. Sun J et al (2003) Stereo matching using belief propagation. IEEE Trans Pattern Anal Mach Intell 25(7):787–800
70. Wang H et al (2021) PVStereo: pyramid voting module for end-to-end self-supervised stereo matching. IEEE Robot Autom Lett 6(3):4353–4360
71. Wang H et al (2021) Co-Teaching: an ark to unsupervised stereo matching. In: 2021 IEEE International Conference on Image Processing (ICIP), pp 3328–3332. IEEE
72. Danzl R et al (2009) 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, pp 484–491. Citeseer
73. Vilaça JL et al (2010) 3D surface profile equipment for the characterization of the pavement texture-TexScan. Mechatronics 20(6):674–685
74. Pertuz S et al (2013) Analysis of focus measure operators for shape-from-focus. Pattern Recogn 46(5):1415–1432
75. Conrad J (2006) Depth of field in depth. Large Format Photography, pp 1–45
76. Sundaram H, Nayar S (1997) Are textureless scenes recoverable? In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 814–820. IEEE
77. Buza E et al (2013) Pothole detection with image processing and spectral clustering. In: Proceedings of the 2nd International Conference on Information Technology and Computer Networks, vol 810, p 4853
78. Ryu S-K et al (2015) Image-based pothole detection system for its service and road management system. Math Probl Eng 2015
79. Zack GW et al (1977) Automatic measurement of sister chromatid exchange frequency. J Histochem Cytochem 25(7):741–753
80. Fan R, Liu M (2019) Road damage detection based on unsupervised disparity map segmentation. IEEE Trans Intell Transp Syst 21(11):4906–4911
81. Beylkin G et al (2009) Fast wavelet transforms and numerical algorithms. In: Fundamental Papers in Wavelet Theory, pp 741–783. Princeton University Press
82. Najman L, Schmitt M (1994) Watershed of a continuous function. Signal Process 38(1):99–112
83. LeCun Y et al (2015) Deep learning. Nature 521(7553):436–444
84. Zhang L (2016) Road crack detection using deep convolutional neural network. In: 2016 IEEE International Conference on Image Processing (ICIP), pp 3708–3712. IEEE

85. Hu Y, Furukawa T (2019) A self-supervised learning technique for road defects detection based on monocular three-dimensional reconstruction. In: 2019 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, vol 59216, p V003T01A021. American Society of Mechanical Engineers
86. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-cam: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision, pp 618–626
87. Fan J et al (2021) Deep convolutional neural networks for road crack detection: qualitative and quantitative comparisons. In: 2021 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE
88. Hoffman J et al (2018) Cycada: cycle-consistent adversarial domain adaptation. In: International Conference on Machine Learning, pp 1989–1998. PMLR
89. Chattopadhay A (2018) Grad-cam++: generalized gradient-based visual explanations for deep convolutional networks. In: 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), pp 839–847. IEEE
90. Girshick R et al (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 580–587
91. Girshick R (2015) Fast R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp 1440–1448
92. Ren S et al (2015) Faster R-CNN: towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst 28:91–99
93. Redmon J et al (2016) You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 779–788
94. Redmon J, Farhadi A (2017) Yolo9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp 7263–7271
95. Redmon J, Farhadi A (2018) Yolov3: an incremental improvement, CoRR
96. Bochkovskiy A et al (2020) Yolov4: optimal speed and accuracy of object detection, *CoRR*
97. Uijlings JR et al (2013) Selective search for object recognition. Int J Comput Vis 104(2):154–171
98. Wang W (2018) Road damage detection and classification with faster R-CNN. In: 2018 IEEE International Conference on Big Data (Big Data), pp 5220–5223. IEEE
99. Suong LK, Kwon J (2018) Detection of potholes using a deep convolutional neural network. J Univ Comput Sci 24(9):1244–1257
100. Camilleri N, Gatt T (2020) Detecting road potholes using computer vision techniques. In: 2020 IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP), pp 343–350. IEEE
101. Ukhwah EN (2019) Asphalt pavement pothole detection using deep learning method based on yolo neural network. In: 2019 International Seminar on Intelligent Technology and Its Applications (ISITIA), pp 35–40. IEEE
102. Dhiman A et al (2019) Pothole detection using computer vision and learning. IEEE Trans Intell Transp Syst 21(8):3536–3550
103. He K et al (2017) Mask R-CNN. In: Proceedings of the IEEE International Conference on Computer Vision, pp 2961–2969
104. Fan R et al (2020) Computer stereo vision for autonomous driving, *CoRR*
105. Fan R et al (2021) Learning collision-free space detection from stereo images: homography matrix brings better data augmentation. IEEE/ASME Trans Mechatron 27(1):225–233
106. Fan R, Wang H, Cai P, Liu M (2020) SNE-RoadSeg: incorporating surface normal information into semantic segmentation for accurate freespace detection. In: Vedaldi A, Bischof H, Brox T, Frahm J-M (eds) ECCV 2020, vol 12375. LNCS. Springer, Cham, pp 340–356. https://doi.org/10.1007/978-3-030-58577-8_21
107. Wang H et al (2021) SNE-RoadSeg+: rethinking depth-normal translation and deep supervision for freespace detection. In: 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp 1140–1145. IEEE

108. Wang H et al (2021) Dynamic fusion module evolves drivable area and road anomaly detection: a benchmark and algorithms. IEEE Trans Cybern. https://doi.org/10.1109/TCYB.2021.3064089
109. Fan R et al (2021) Three-filters-to-normal: an accurate and ultrafast surface normal estimator. IEEE Robot Autom Lett 6(3):5405–5412
110. Ozgunalp U (2016) Vision based lane detection for intelligent vehicles, Ph.D. dissertation, University of Bristol
111. Zhang Z (2013) Advanced stereo vision disparity calculation and obstacle analysis for intelligent vehicles, Ph.D. dissertation, University of Bristol
112. Wu R et al (2021) Scale-adaptive road pothole detection and tracking from 3d point clouds. In: 2021 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE
113. Hast A, Nysjö J (2013) Optimal RANSAC-towards a repeatable algorithm for finding the optimal set. J WSCG 21(1):21–30
114. Kang B-H, Choi S-I (2017) Pothole detection system using 2D lidar and camera. In: 2017 Ninth International Conference on Ubiquitous and Future Networks (ICUFN), pp 744–746. IEEE
115. Clohessy T et al (2014) Smart city as a service (SCaaS): a future roadmap for e-government smart city cloud computing initiatives. In: 2014 IEEE/ACM 7th International Conference on Utility and Cloud Computing, pp 836–841. IEEE
116. ThoughtLab E (2018) Smarter cities 2025 building a sustainable business and financing plan. https://econsultsolutions.com/wp-content/uploads/2018/11, ESI-ThoughtLab_Smarter-Cities-2025_ebook_FINAL. pdf (letöltve: 2019.07. 26.)
117. Yoon SY et al (2020) Smart city pathways for developing Asia: an analytical framework and guidance, ADB Sustainable Development Working Paper Series
118. Zhe W et al (2015) Traffic patterns in the silk road economic belt and construction modes for a traffic economic belt across continental plates. J Resour Ecol 6(2):79–86
119. Tsakalidis A et al (2020) Digital transformation supporting transport decarbonisation: technological developments in EU-funded research and innovation. Sustainability 12(9):3762
120. Parti K, Szigeti A (2021) The future of interdisciplinary research in the digital era: obstacles and perspectives of collaboration in social and data sciences-an empirical study. Cogent Soc Sci 7(1):1970880
121. Singh S et al (2020) Convergence of blockchain and artificial intelligence in IoT network for the sustainable smart city. Sustain Urban Areas 63:102364
122. Leduc G et al (2008) Road traffic data: collection methods and applications. Work Pap Energy Transp Climate Chang 1(55):1–55
123. Sidek O, Quadri S (2012) A review of data fusion models and systems. Int J Image Data Fusion 3(1):3–21
124. Jeong S et al (2020) City data hub: implementation of standard-based smart city data platform for interoperability. Sensors 20(23):7000
125. Al-Turjman F et al (2022) An overview of security and privacy in smart cities' IoT communications. Trans Emerg Telecommun Technol 33(3):e3677