A review on computer vision and machine learning techniques for automated road surface defect and distress detection

Xuejing Chen [a], Sira Yongchareon [a,*] and Martin Knoche [b]

[a] *Computer Science and Software Engineering Department, Auckland University of Technology, Auckland, New Zealand*
E-mails: xuejing.chen@aut.ac.nz, sira.yongchareon@aut.ac.nz
[b] *N3T Limited, Whangarei, New Zealand*
E-mail: martin@n3t.kiwi

Abstract. As the pace grows in the development of image processing techniques and the current applications rise in machine learning and deep learning techniques for visual inspections and physical assessment, this article reviews the existing literature. It provides a detailed synthesis of the overview of surface pavement conditions, computer-vision-based technologies for road damage detection, various datasets and data collection methods. We analyse and compare different machine-learning methods and models proposed in the literature and identify challenges that need to be addressed in the future in road surface defect detection.

Keywords: Road condition assessment, distress detection, road defects analysis

1. Introduction

Road accidents have become one of the most significant reasons leading to severe injuries and deaths in recent years. Nearly 1.3 million people die on the world's roads every year, and 20 to 50 million suffer non-fatal injuries [60]. The degradation and defects of the pavement contribute to poor driving conditions due to insufficient time to call drivers' attention, sometimes leading to prolonged reactions, off-track driving and car accidents. It also impacts driving speed, skid resistance, rain drainage, wear abrasion, ride quality, engine operation, vehicle maintenance, and the gripping of tyres. Road environment contributes to 34% of traffic accidents via road layout and security facilities [44], whereas the vehicle factor accounts for 13% comparatively. Kwon [34] reported that the rating of road pavement is closely linked to traffic collisions and examined that car accident is up to 25 times different due to the grade of the pavement.

According to the report on New Zealand's rural roads [8], the average cost of maintaining and repairing a road surface was measured at $21,000 a year per year, assuming a 40-year life and $38,000 cost per year per km of

---

*Corresponding author. E-mail: xuejing.chen@aut.ac.nz.

road for a 10-year life cycle. It is predicted that maintaining the road pavement material was estimated to add approximately US$13.6, US$19.0 and US$21.8 billion, respectively, because of climate change by 2010, 2040 and 2070 [59]. However, if it was expected that the timber resource and the flow of logging trucks would continue on that path for ten years, the gross annual expense rose to $31,000 per km per year.

Transportation infrastructure systems significantly impact a country's development progress and well-functioning economic activities. For example, New Zealand's ten-year road safety strategy targets to reduce annual fatalities from 330 in 2019 by 40% by 2030, by which the government has already invested $1.4 billion over three years to update the most dangerous 3,300 km roads [62]. The maintenance and investment in road safety upgrades are essential for preventing fatalities and crashes. For example, the measures of GAMBIT, Poland National Road Safety Programme, were introduced to EU Directive and adopted on inspection of existing road networks in 2005 and were effective between 2007—2010 [31]. After changing the layout of the road structure, the police claimed that near those spots, the number of crashes, deaths and injuries dropped by 38%, 65% and 36%, respectively, in Korea [64]. As a result, detecting road conditions by monitoring techniques becomes imminent to avoid the high cost of repairing activities and multifunctional constructions.

2. Road surface pavement and other surface damages

Road surfaces consist of concrete, asphalt, chip seals, and unsealed ones. The adhesive material in asphalt is petroleum, in concrete is cement, and in chip-seal is a thin layer of stones in tars. Asphalt is found to be easy to patch, while, on the other hand, concrete cracks are more difficult to fix and unlikely to resurface. However, asphalt driveways require re-sealing, but the concrete surface does not need it. Asphalt has good grip capability but reduces skid resistance when wet, but the chip seal is generally smooth and lifts the risk of skidding. In rural areas, road surfaces usually are mud and gravel and cause great dangers because of loosened tops and raised dust. Here, we mainly focus on asphalt because New Zealand roads are primarily constructed with it. Our discussions revolve around the necessity to replace or refurnish the road pavement after physical assessment and visual inspections.
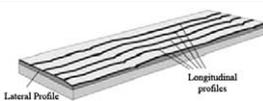
2.1. Road surface crack

Climatic reasons such as heavy rain and extreme heat cause most cracks, while some by load and traffic. Some happen because of pavement layers that are too thin or lack underlying support. Transverse cracks are also known as unrelated cracks, which run unconnected laterally through the pavement primarily due to the shrinkage of the surface layer or the underlying base layer. Longitudinal cracks are non-load related. They are formed due to the contraction or shrinkage of the surface layer or reflection from the underlying base layer joints. Fatigue cracking is inter-connecting cracks, block cracking is orthogonally intersected transverse and longitudinal, and slippage cracks are half-moon-shaped cracks caused by traffic. The occurrence of fatigue cracking suggests the end of the pavement's life cycle. Alligator cracking is characterised by interconnection or joining breaking within the black-top layer, taking after traffic loading. Edge crack is the formation of crescent-shaped splits close to the edge of a road [39]. It is caused by the need for support of the street edge, sometimes due to an ineffectively depleted drainage system. The cracking types' positions, dimensions, and orientations indicate distinct safety levels (Table 1).

2.2. Road surface damages

Road surface distresses and damages result in less attentive driving and may cause steering off the road and clashes. Rutting is permanent longitudinal deformation along the surface, followed by surface degradation like distress and fatigue cracks in the pavement, finally leading to cracking and disintegration in deeper road structures. It is caused by insufficient moisture control or gradual brokenness of the surrounding structure after construction. When the drainage system is not functioning well, the subgrade is poorly sealed, and ditches are filled with water on rainy days. The accumulation of water reduces skid resistance power and causes car accidence. If a road is not apparent at night or under a thin layer of water, drivers may not be aware of the possible dangers and may be unable to change the car's speed in rough areas [16]. In regions affected by moisture and frost, rutting causes expansion in

Table 1
Effects of road defects on road safety

Road deftec type	Condition			Effect
	Good	Fair	Poor	
Pavement roughness	IRI < 1.5 m/km [44]	1.5 m/km < IRI < 2.7 m/km [44]	IRI > 2.7 m/km [44]	Decreases single-vehicle accidents but increases multiple-vehicle accidents [4]
Rutting	RD < 5 mm [7]	5 mm < RD < 10 mm [7]	RD > 10 mm [7]	Has a significant impact during the night and rainy days but is visible during daytime and normal weather [16]
Cracking	X < 10 mm [47]	10 mm < X < 75 mm [47]	X > 75 mm [47]	Cracking width, length, and the surrounding area causing different levels of ravelling and risk of brokenness
Ravelling	$A < 1\ m^2$ [49]	$1\ m^2 < A < 10\ m^2$ [49]	$A > 10\ m^2$ [49]	Stripping area affects the safety level
Manhole	D < 20 mm [49]	20 mm < D < 50 mm [49]	D > 50 mm [49]	Manhole depth or height has intervention on the surrounding
Joint	X < 75 mm [49]	75 mm < X < 130 mm [49]	X > 130 mm [49]	The length of the joint seal defect needs monitoring and inspection
Edge break	X < 75 mm [49]	75 mm < X < 125 mm [49]	X > 125 mm [49]	Edge break exceeding a limit can be harmful
Patch	<10% of the surface [47]	10%–25% of the surface [47]	>25% of the surface [47]	Peculiar in urban areas and are highly affected by their shape

Pothole:

Depth (mm)	Diameter (mm)			
13–25	100–200	200–450	450–750	
25–50	Low	Low	Medium	
>50	Low	Medium	High	
	Medium	Medium	High	

Effect (Pothole): The shape, depth and average diameter have different impacts on road safety and comfort. It can lead to wheel damage, distraction, and swerving to put drivers in danger [47]

susceptible soil and road materials, and it further loses road carrying capacity in defrost season. The related accident rate is beginning to rise slightly as the routing depth approaches 7.6 mm [54]. Pavement roughness is the pavement surface's irregularness that adverts the ride of vehicles. The measurement to define the roughness of longitudinal profiles is known as the International Roughness Index (IRI), which is the ratio of a vehicle's suspension motion and the travelled distance. It has been estimated that a 1 m/km decrease in IRI would save $321 million per year [46]. On the other hand, an IRI rise of up to 4 m/km would increase the cost of repair and maintenance for passenger cars and heavy trucks by 10 per cent. This rise is up to 40 per cent for passenger cars and 50 per cent for heavy trucks at an IRI of 5 m/km.

Pavement deformation involves the vertical and horizontal change of the layer, such as depression caused by the swelling of the surface structure accompanied by traffic and climatic influence. Depression is a bowl-shaped depression or broken part of a slab, which comes into effect because of the concrete's disconnection and reinforcement placement. Potholes start after rain leaks into breaks and down into the soil underneath the street surface. The soil turns into mud with no bolster, and a gap forms beneath the road. Repeated freezing and defrosting of overwhelming traffic causes the ground to extend. The ground returns to a lower level as the temperature rises, but the surface is raised. This makes a crevice between the asphalt and the ground underneath it. When a vehicle drives over it, the surface splits and falls into the hole, thus making a pothole. Potholes cause thousands of dollars of harm to vehicle wheels each year.

The surface imperfection includes bleeding and ravelling, shown as excessive bituminous cover on the top layer and insufficient asphalt, resulting in dislodging of the total. The causes of surface defects are related to bituminous and asphalt materials characteristics. Comfort is adversely influenced by the surface defects like ravelling and bleeding, whereas surface deformation like potholes, depression, rutting, and other road distresses like manholes, separation, pavement joints, and railroad crossings have an apparent impact on driving conditions [47].

2.3. Other surface damage scenarios

Other surfaces like bridges, tunnels, pipes, bricks, and building surfaces form similar degradation to road surfaces, such as cracking, spalling, corrosion, etc. The defect's magnitude helps predict future damage changes for visual inspection. Cha et al. [13] have collected 2366 images of civil infrastructure for four types of damage – concrete crack, steel corrosion, bolt corrosion, and steel delamination. Faghi-Roohi et al. [21] used the video data covering 700 km railway to label 22408 subjects into six surface classes normal, weld, light squat, moderate squat, severe squat, and joint. Kim et al. [26] classified the images collected from the Internet into cracks, single line joints/edges, multiple lines join/edges, intact surfaces, and plants. In another work, the crack and non-crack images from the concrete and brickwork buildings are generalised for inspection during building monitoring [43]. Different construction surfaces are assessed in the previous studies by comparing distinct defect classes and severities, which has provided useful data resources for DL model development. Nevertheless, the data types can still be improved to include more comprehensive categories and degradation levels. In addition, it is worth discovering technical methods to detect similar kinds of defects in various construction surfaces.

3. Computer vision-based image processing techniques

Civil infrastructure like tunnels, pipes, roads, and bridges are under continuous assessment to ensure the safety and serviceability of engineering facilities and building structures. Usually, pavement assessment procedures are carried out manually by inspectors and road assessors. However, Computer Vision (CV) techniques have developed swiftly to take the intensive job of visual inspection of surface defects to save time and increase accuracy.

Feature engineering based on mathematical algorithms has been proposed to detect road damage in a street context in the past decades. The traditional image processing techniques [25] mainly include such methods as intensity-thresholding [14,24,40], edge detection [1], filtering [51,67], and wavelet transforms [15,70]. These techniques can be attributed to feature-based methods, model-based methods, and pattern-based methods [. Abdel-Qader et al. [1] showed fast Haar transform is more reliable than other edge detection techniques like a fast Fourier transform, Sobel and Canny on detecting concrete bridge crack images. Yamaguchi et al. proposed a percolated technique to extract a continuous texture to reflect the connectivity of brightness and shape [63]. This method was used for noise reduction and has achieved high precision-recall and receiver operating characteristics. [24] presents subtraction pre-processing with smoothed image and a line filter based on the Hessian matrix to highlight the cracks-like structures. Fujita et al. finally utilised thresholding processing to separate cracks from the background. Salman et al. [51] showed that the Gabor filter is highly effective in detecting multidirectional cracks by analysing images with a large degree of surface texture that makes crack detection hard. The Gabor filtering technique is further developed by Medina and Llamas [67] with three different methodologies for setting the threshold of the classifiers. The AdaBoost algorithm is found to show the best results. Zhong et al. [70] showcased the application of continuous transforms

Table 2
Some examples of image processing and ML techniques for crack detection

Work	Type of damages	Techniques	Performance	Methods	Drawbacks/Limitations
Abdel-Qader et al. [1]	Bridge crack	Haar Wavelet	Acc = 86%	feature-based	Noisy image data questionable
Yamaguchi et al. [63]	Large surface crack	Scalable local percolation-based	Pr = 70% R = 90%	Model-based	rely on user input to initialise the seed pixels
Abdel-Qader et al. [2]	Cracks on a bridge surface	Principle component analysis	Acc = 73%	ML	Camera pose and distance make results different
Salman et al. [51]	Pavement cracks	Gabor filter	Pr = 98%	feature-based	Not optimal if one seeks broad spectral information
Lattanzi and Miller [35]	Cracks on different surfaces	Clustering method based on Canny and K-Means	Acc (Canny) = 87.5% Acc (K-Means) = 86.5%	feature-based & ML	smaller objects not recognised as one continuous crack
Zhong et al. [70]	Cracks on pavement surfaces	Wavelet transforms	NA	feature-based	Results not validified on datasets
Prassana et al. [45]	Cracks on bridge decks	Multi-feature Adaboost	Acc = 95%	ML	Can only analyse high-resolution images
Shi et al. [52]	Road pavement cracks	CrackForest	Pr > 90%	ML	Computational complex

representing a cracked simply-support beam using finite elements. Another work adapting the wavelet method by a 2D matched filter [15] uses Markov Random Field (MRF) to detect multi-scale and segment fine structures of the image. The overview of image and ML techniques on crack detection is summarised in Table 2.

4. Machine learning in road defect detection

The development of machine learning (ML) has grown in the interest of defect/damage detection studies in the past decade. For example, it was well applied to studying automated pavement crack detection problems. In addition, by eliminating the need for manually tuning threshold parameters, ML helps develop parameter choices.

4.1. Traditional machine learning methods

The development of vision-based ML, surface crack and distress investigation, has prompted plenty of techniques, such as PCA [2], AdaBoost [45], K-Means [35], SVM [3] and Random Forests [52]. Abdel-Qader et al. [2] applied PCA on raw data, implemented linear structure modelling, and segmented local information to enhance local detection with linear modelling over global. Lattanzi and Miller [35] used robust feature extraction and ML methods based on Canny and

K-Means for crack segmentation. They showed that the clustering techniques explore the inherent characteristics of fracture images to achieve consistent performance and improve classifier outcomes. Prassana et al. [45] investigated support vector machines (SVM), AdaBoost, and Random Forest as classifiers with different Laplacian pyramid feature vector combinations. They used this algorithm to make robust curve fitting so that the potential crack regions are spatially localised even in the presence of noise. CrackForest was developed by Shi et al. [52] by introducing random structured forests with integral channel features and proposing a new crack descriptor to characterise cracks and differentiate them from noises.

Many works integrated traditional ML and image processing techniques to enhance performance. Wang et al. [61] detect surface cracks of wind turbine blades by images captured by UAVs. They used a Parallel data-driven crack location method: parallelised sliding windows connected into the cascading classifier and parallel Jaya K-means crack contour detection methods. Specifically, a parallel sliding window method is utilised to scan images

for locations, cascading classifier to classify sliding windows into two classes: cracks and non-cracks, and crack windows on Haar-like features to locate cracks by the extended cascading classifier (ECC). Finally, the Jaya K-means algorithm is developed to cluster each pixel in crack windows into crack and non-crack segments to obtain crack contour. Jaya algorithm tunes the parameters and simplifies the approach. Gaussian filter is then used for edge preservation.

In practice, WT blade images are captured by UAVs from commercial wind farms. The raw images have different backgrounds under different lighting conditions. Each pixel of crack windows is labelled by RGB and HSV features. The performance evaluation shows that 95.83% of cracks are detected, and no false alarm is observed. For contour detection: Jaya K-means algorithm yields the smallest values of minimal, maximal and average values of SSEs (sum of square errors) over Sequential sliding window, Generic K-means and PSO K-means. For crack locations: The parallel sliding window can reduce the detection time by 7.9 times. For contour detection: the Jaya algorithm has a faster convergence speed. The integration of cascading classifier and parallel sliding window accelerated image-based crack detection and developed parallel Jaya K-means-based crack contour detection. The benefit of the approach is the flexibility to access remote areas, collect multimedia information on blades, and prevent inspection dangers. But then, robustness needs to be proposed for variations of the illumination and background.

Ai et al. [3] proposed a probabilistic generative model (PGM) and a support vector machine (SVM) based fusion algorithm. A PGM-based method is used to generate a probability map based on pixel intensity information. The SVM-based method generates probability maps based on multi-scale neighbourhood information. A novel fusion algorithm can merge the multiple probability maps into a fused map to detect cracks with high accuracy. Weighted dilation operation enhances the recognition of borderline pixels and improves crack continuity. The method outperforms the two baselines regarding recall and F1-score while achieving precision (90.7%) close to the baseline approaches at 2-pixel margin vicinity. They fuse the probability vector to make Max, Min, Multiply and Mean operations, respectively. They found that the Max operation can help detect most cracks as non-crack pixels are classified as cracks. Multiply operation uses a small number of crack pixels, which misses several cracks. But Min and Mean operations are in between the Max and Multiply operations. The results show that the accuracy is higher than any original probability maps. It detailed the borders and the widths of the cracks applicable to noisy environments. The method deals with cracks with heterogeneous intensity, complex topology in morphology, and bad illumination conditions. However, it is necessary to continue to investigate how the neighbourhood affects the detection results and need to determine features extracted from the neighbourhood. They also need to accelerate the detection using CUDA and further optimise the detection algorithms.

4.2. Deep learning methods

The visual inspection of asphalt pavements and other surfaces has been developed over the last three decades. However, detecting damages is a challenging problem due to the variations of image sources, non-uniformity of cracks, inadequate brightening, and other similar features. Mathematical method-based algorithms show limitations in detecting cracks in different morphology and situations, demanding tedious efforts to calculate the crack pixel locations. The recent achievements brought by Deep Learning (DL) methods have shed light on automatic pavement image investigation and propelled the applications to move from handcrafted features to data-driven approaches. The results of large neural networks outperform small and medium neural networks in precedence over traditional learning algorithms by the tide of AI development. More dynamic analysis and algorithmic strategies are developed to address the challenges. Due to the flourishing deployment of portable hardware, powerful-capacity computation, parallel device usage and database systems, Deep Learning (DL) approaches have reached outstanding achievements and are brought into the applications in street view detections and surface texture investigation.

In the 1980 s, the idea of non-linear dynamics was introduced into a neural network. In the same period, the backpropagation (BP) algorithm was applied as an efficient gradient descent algorithm, a reliable learning method for multilayer neural networks. Neural networks, characterised by hidden layers of numerous neurons, extract features by training forth and back through the architecture levels constituting non-linear information processing units. Deep networks, usually Convolutional Neural Networks (CNN), implement foundational layers like pooling and convnet to target multi-class classification problems. Deeper neural networks, however, often encounter over-fitting issues when coming in the form of complicated structures to deal with small crack image datasets. Hence, fine-tuning

process, the use of a dropout layer, and regularisation are necessary to adjust the parameters of the classification layer or the entire network.

Many up-to-date techniques of deep learning models use CV visions to produce good recognition results. Hinton et al. [30] integrate backpropagation learning with a multilayer neural network to recognise an object with high computational efficiency. The authors train a deep learning model to generate sensory data and learn non-linear distributed data for each layer. Fu et al. [23] proposed a recurrent CNN (RCNN) for object recognition. They discovered that this model could fix the number of parameters, facilitate the learning process, and outperform the state-of-art models. The number of hidden nodes is shown to be more significant than the algorithm or depth of the model to achieve high performance in Coates et al. [19]. Only a single layer of features can be used to obtain good results when pushed to limits. He et al. [29] demonstrated a residual learning framework to ease the training of deeper neural networks. The result shows that the residual networks are easier to optimise and can achieve higher accuracy when tested to conduct visual recognition tasks. A novel recurrent attention convolutional neural network (RA-CNN) [23] is proposed by Fu et al. to reflect the mutually correlated relation between region detection and feature representation. Again, a deep CNN network by Krizhevsky et al. [33] is shown to classify 1.3 million high-resolution images with a low top-5 error. The current CV machine learning studies focus on a result-oriented engineering process. However, it lacks target-oriented motives to modify models to tailor them to user-defined objects.

1) *Existing DL work on pavement image analysis and automated distress detection* CNNs, which comprise convolutional layers, activation layers and pooling, are commonly used for DL applications on a wide range of pavement image analysis problems. CNNs and DCNNs have deeper architectures than shallow neural network structures, resulting in more reliable performance and application efficacy. In the past five years, there has been a vast growth of studies on CNN applications to address various challenges in vision-based automated pavement distress detection. The large-scale public annotated image datasets serve for pavement image classification problems by applying various the-state-of-the-art CNN techniques.

   For example, Gopalakrishnan et al. [26] built a simplified vision-based pavement crack detection system using a pre-trained deep learning model through transfer learning and domain adaptation approach. The pre-trained DCNN uses Keras deep learning framework and is implemented on the VGG-16 model. The pre-trained deep Convolutional Neural Networks (DCNN) were studied by transfer learning for automated pavement distress detection. The truncated VGG-16 DCNN is a deep feature generator that vectorises the labelled pavement images and then trains a machine learning classifier to predict the labels. Various machine learning classifiers are trained using the semantic image vectors, and a neural network classifier gives the best result. A single-layer neural network classifier (with 'adam' optimiser) is trained on ImageNet pre-trained VGG-16 DCNN with 144 million parameters. It contains 16 convolutional layers with very small receptive fields of size 3×3, and five max-pooling layers of size 2×2 for carrying out spatial pooling, followed by three fully-connected layers, with the final layer as the soft-max layer. Rectification nonlinearity (ReLu) activation and dropout regularisation are applied to all hidden and fully-connected layers. Pavement images were sampled from the FHWA/LTPP ImageNet database with 1056 pavement images. The pre-processed image size is set to 1000×500 pixels, and VGG-16 DCNN default images were cropped to 224×224 pixels.

   There are 25,088 deep transfer learning features as inputs. The experiments were conducted on an Intel CoreTM i7-5600U CPU on 64-bit Windows 10 OS. The NN classifiers were trained on ImageNet pre-trained VGG-16 DCNN. They yielded the best performance in accuracy, precision, recall, F1-score, Cohen's Kappa score, ROC curves, and area-under-the-ROC-curve (AUC) values. VGG, AlexNet, and GoogLeNet are learned in the deep transfer learning approach, which was insensitive to surface colour and texture variation. Using a pre-trained deep learning model and fine-tuning with smaller datasets achieves a higher order of complexity. A challenging machine learning for a few instances to the VGG-16 model has learned to extract features and distinguish one class from another. A significantly higher order of complexity is introduced to prove model robustness. However, it fails to learn to distinguish cracks from joints in PCC-surfaced pavements and requires high computational complexity. Table 3 lists examples of recent deep learning models for surface defect detection.

Table 3
Some examples of deep learning models on surface defect detection

Work	Type of damages	Techniques	Performance Acc/F1/...	Detection problem	Drawbacks/Limitations
Gopalakrishnan et al. [27]	Pavement cracks	Transfer learning & VGG	Acc = 90%	Classification	Fails to learn to distinguish cracks from joints in PCC-surfaced pavements
Rajadurai et al. [48]	Surface cracks	AlexNet	Acc = 81–89%	Classification	Accuracy degraded by shadows, surface roughness, scaling, edges, holes, and background debris
Cha et al. [11]	Concrete and steel surfaces	CNN & sliding window	Acc = 98%	Classification	The incapability of sensing internal features due to the nature of the photographic image
Faghih-Roohi et al. [22]	Railway defects	DCNN	F1-score – 92%	Classification	Doubles computational time
Maeda et al. [14]	Road defects	SSD	Pr > 75% R > 75%	Detection	Hard to detect rare types of damage
Stricker et al. [56]	Road defect	GAN & Auto-decoder	Acc = 60%	Detection	The performance can be improved

1.1) *Deep learning models on road defect classification*. Kim et al. [32] detect real structures (various apparent conditions) over cracks and intact surfaces into five classes (intact surfaces, cracks, multiple joints and edges, and a single joint or edge). They used transfer learning to form a probability map (softmax layer) to increase the robustness of a sliding window and parametric study of thresholds. The database was scraped from the Internet, the classifier was developed by transfer learning of Alexnet (CNN), and the probability map detected cracks. They used AlexNet of five convolutional layers, max-polling layers, three fully-connected layers with a 1000-way softmax, and a non-linear unit (ReLU) (activation function) at the end of neurons to reduce the vanishing gradient effect for real-time video frames of an unmanned aerial vehicle under field conditions. UAV-based inspection https://youtu.be/5sNbfEaRwkU has 40 images. The experiments were conducted on CPU: Intel(R) Core(TM) i3-6100 and RAM:8192 MB GPU: NVIDIA Geforce 1060 3 GB. Pixel-level test results turn at 90% of accuracy at 86.73% average precision and 88.68% average recall. Accuracy vs. threshold shows that when at average accuracy, the result changes slightly. At average precision, it slowly increases from 0% to 50% and remains constant afterwards. The average recall decreases slightly from 0% to 50% and drops after 50%. A threshold is determined as 35%, and precision and recall are higher than 90%. The automated DL method is to detect cracks in natural structures. Detailed categorisation can achieve high accuracy of trained CNN misclassification due to shape similarity.

Cha et al. [12] extract defect features like cracks in concrete and steel surfaces with a vision-based method using a deep architecture of convolutional neural networks (CNNs) for detecting concrete cracks without calculating the defect features and works without the conjugation of IPTs for extracting features. The trained CNN is combined with a sliding window technique to scan any image size larger than 256 × 256 pixels resolutions. The designed CNN is trained on 40 K images of 256 × 256 pixels resolutions and tested on 55 images of 5,888 × 3,584 pixels. It achieves 98% accuracy and overcomes extensively varying real-world situations (e.g., lighting and shadow changes).

Faghih-Roohi et al. use a deep convolutional neural network with three DCNN structures (small, medium, and large) considered [21]. The implementation is based on the framework in Torch 7. The images are obtained from many hours of automated video recordings for 350 kilometres of track manually labelled 22408 objects, 10 per cent of the samples for testing and use the remaining 90 per cent for training. The rail defect classes can be classified with almost 92% by F1-score measurement. The large DCNN primarily performs better than the rest. However, this improvement comes at the price of doubling the computation time as opposed to the small DCNN for extracting suitable features for detecting rail surface defects. The approach took advantage of the DCNN to skip elaborate procedures of feature extractions

that are required in classical learning approaches. It uses raw images as the only input to the classification model and optimises the network using a mini-batch gradient descent method. The DCNN classification accuracy is very high, and only the normal type can reach 95.74%. However, there is no cross-validation with other models.

1.2) *Deep learning models on object detection of road defects*. The SSD Inception V2 and SSD MobileNet by Maeda et al. [38] were trained on the dataset and compared the accuracy and runtime speed on both, using a GPU server and a smartphone. For the results, they achieved recalls and precisions greater than 75% in the best-detectable category and achieved an inference time of 1.5 s on a smartphone. Stricker et al. [55] used a generative adversarial network (GAN) along with Poisson blending and a variational autoencoder to generate road damage images as new training data to improve the accuracy of road damage detection. The addition of a synthesised road damage image to the training data improves the F-measure by 5% and 2% when the number of original images is small and relatively large. The enlisted groups participating in the RDDC2020 challenge achieved an average f1-score of more than 60%. Their detection techniques include Yolov5, Ensemble model, Ensemble Prediction, and data augmentation such as TTA are also adopted as efficient methods to lift performance. Duplicate or overlapped predictions generated in the process are filtered using the non-maximum suppression (NMS) algorithm [5].

1.3) *Deep learning models on crack segmentation*

1.3.1) *Pixel-wise segmentation*. Zou et al. [72] found that more detailed representations are made in larger-scale and smaller-scale feature maps. They proposed DeepCrack: an end-to-end trainable deep CNN. It is built on the encoder-decoder architecture of Segnet. They made a skip-layer fusion to connect the encoder and decoder networks and pixel-wise semantic segmentation to learn multi-scale features. There is one convolutional layer before the pooling layer at each scale of the encoder network concatenated crack dataset of 260 pavement images (https://sites.google.com/site/qinzoucn). There are three crack datasets: pavement image datasets of 100 and 315 images each and stone surface of 331 images. Human experts manually label the ground truth cracks. It is implemented on GeForce GTX TITAN X GPU by three metrics of F-measurement: ODS, OIS and AP, all have the highest value of 0.872. The precision-recall achieves the highest performance. It is compared with HED, RCF, SegNet, SRN, U-net, SE, CrackForest, DeepCrack, and Crack-Tree (Traditional low-level feature-based method). The performance efficiency is 6 FPS, slower than HED, RFC, and SRN. They removed the skip-layer connection to decrease performance. The model is trained from scratch to obtain better performance than trained from a pre-trained model. They also reduce the ground-truth crack label and shift crack labels to overcome declined performance, whilst noisy crack labels have little influence. Smaller crack weight decreased performance. They design a new neural network architecture for crack detection, using a fusion of multi-scale features at each encoder-and-decoder level to infer cracks from the background. They construct DeepCrack with different scales and weights assigned to each scale and train DeepCrack with or without pre-trained models of SegNet and Pascal Voc, sampling operation of max-pooling indices to bilinear interpolation and detection of bright cracks. However, it is not practical only for crack information extraction. Ground truth marks are laboured by handwritten compared with decision tree and three other none-DL models. More network layers lead to DeepCrack handling images more slowly than HED, RCF and SRN.

1.3.2) *Patch-wise segmentation*. Eisenbach et al. [50] introduced the German asphalt pavement distress (GAPs) dataset as the first free and publicly available massive pavement distress images dataset suitable for training high-performance DCNNs. However, almost all previous studies in this domain used their datasets collected and annotated differently. Thus the performance of the developed distress detection algorithms and techniques could not be compared to a standard benchmarking dataset. Furthermore, even when the pavement distress datasets were made publicly available, they were not big enough to directly implement the DCNNs [50]. Thus, the GAPs dataset (http://www.tu-ilmenau.de/neurob/data-sets-code/gaps/) seems to be the first attempt at

creating a standard benchmarking pavement distress images dataset for deep learning applications. It includes 1969 grayscale pavement images (1418 for training, 51 for validation, and 500 for testing) with various distresses, including crack, potholes, inlaid patches, applied patches, open joints, and bleeding.

2) *Temporal-spatial analysis of pavement conditions*. Chen et al. [18] proposed a DL convolutional neural network (CNN) and a Naïve Bayes data fusion CNN that detects crack patches in each video frame. The data fusion scheme maintains the spatiotemporal coherence of cracks in the video. Naïve Bayes decision-making discards false positives effectively. Configurations of convolution, pooling and fully connected layers follow the model used in TensorFlow. The videos are with 30 ft/s and 720×540 pixels resolution 2. 5326 crack image patches manually annotated. It is augmented to have 147 344 crack and 149 460 non-crack image patches. Two Intel Xeon E5-2620 v4 CPUs, 256-GB DDR4 memories, and four NVIDIA Titan X Pascal GPUs. It turns out that there is a 98.3% hit rate against 0.1 false positives per frame with NB-CNN with data fusion LBP-SVM with data fusion LBP-SVM. The CNN architecture detects crack patches, and a registration procedure maintains the spatiotemporal coherence of cracks in videos. A Naïve Bayes data fusion scheme discards false positives effectively by aggregating information from multiple frames.

3) *3D detection of road distress*. The methods for recognition, location, measurement, and 3D reconstruction of concealed cracks are developed based on convolutional neural networks. Tong et al. [58] proposed three different CNNs (recognition, location, and feature extraction). They used a feed-forward algorithm to train data and decrease errors. The recognition CNN is to distinguish concealed cracks from other types of damage in a GPR image. The location CNN determines the location and length. The 3D reconstruction makes feature extraction by CNN models. In recognition CNN: CNN is composed of two convolutional layers (C1 and C2 based on the activation function sigmoid) and two subsampling layers (S1 and S2 based on different feature matrixes to avoid excessive useless information), followed by two full connection layers (F1 and F2) and the output layer. Location CNN is similar to the above. Recognition CNN uses 500 grey-scale maps divided into 256×256 pixel sized 6832 images, 2200 as the training samples for the CNNs. The types of damages are quadrature encoded. Location CNN uses 5233 data 4000 labelled for training. Feature extraction CNN made 4000 labelled images out of 5233. Recognition CNN split 4632 divided images for testing. Location CNN has 5233 data and 1233 for testing. Feature extraction CNN made 1233 images out of 5233.

The experiments are conducted on Inter(R Core(TM) i7-6700 CPU, 8.00 GB RAM, and NVIDIA GeForce GTX 1060 6GB GPU. The recognition CNN has zero errors, and CNN locates concealed cracks correctly. The MSE of the edge box and length is 0.327 cm and 0.732 cm. The length recognition possesses a 0.2543 cm mean-squared error, a 0.978 cm maximum length error, and a 0.504 cm average error in the recognition. CNN was designed to distinguish concealed cracks from other pavement damage in GPR radar images with no error. Measuring errors satisfies the demand for highway engineering detection. The recognition CNN produces no erroneous results. However, there is more to explore on how 3D reconstruction models can predict the growth tendency of cracks. Detection efficiency could be improved for classification. Width coordinates and feature point extraction are too rough and not verified to be the best-performed method due to a lack of comparison with other methods.

Zhang et al. [68] explicitly the objective of pixel-perfect accuracy for the set of predicted class scores for all pixels. CrackNet, an efficient architecture, is based on the Convolutional Neural Network (CNN). CrackNet does not have any pooling layers that downsize previous layers' outputs. It ensures pixel-perfect accuracy using the newly developed invariant image width and height technique through all layers. CrackNet comprises five layers and includes more than one million parameters trained in the learning process. The input data of the CrackNet are feature maps generated by the feature extractor using the proposed line filters with various orientations, widths, and lengths and trained with 1,800 3D pavement images and tested with another set of 200 3D pavement images. It achieved high precision (90.13%), recall (87.63%) and F-measure (88.86%). The spatial size of the input data is invariant through all layers. The ground truth of training data is prepared for pixel-to-pixel supervised learning 3. Last, an individual pixel is compared with its neighbours through local connections.

4) *Road defect measurement*. Tong et al. [57] handled image sizes and shapes from asphalt pavement images and transferred learning of DCNN to use weights and bias for length classification. RGBs are transformed into grey-scale images to calculate the threshold and properties by k-means clustering. The DCNN is initialised with parameters in the fine-tuning process to recognise crack length. Quadrature encoding improves response rate and training convergence. The stochastic gradient descent (SGD) speeds up the training rate and decreases weight errors in the feed-forward algorithm and is composed of two convolutional layers, two subsampling layers, and two fully connected layers. The dataset is divided into 200*200 images as grey-scale pictures. Training samples contain 7500 images with cracks and 500 images without cracks, and 1200 images are randomly selected for the pre-train process. Five hundred images are selected from the training sample as the testing sample. For image Extraction: Inter ® Core ™ i7-6700 CPU and 8.00GB Random Access Memory (RAM) and NVIDIA GeForce 1060 6GB GPU 2. For length classification: Intel ® Core ™ i5-2520 CPU, 4.00 GB RAM and NVIDIA GeForce GT 630. Recognition accuracy achieves 94.36%. Maximum length error equals 1 cm, mean squared error equals 0.2377, length recognition = 3 ms. It is found that the image resolution has little influence on accuracy in the range of 54–300 PPI. The light condition has little impact on the accuracy of DCNN because of the extracting features by k-means. The training strategy overcomes the lack of crack labels and improves accuracy. They can extract discriminative crack features. Crack labels fine-tune the trained network to recognise length. The developed DCNN has lower requirements for hardware capacity and enhances efficiency. The developed DCNN is transferrable to other asphalt pavement at 6–8 cm cracks with high error rates. The model's efficiency and accuracy are not verified by comparing it with other benchmark models.

5. Datasets

A conventional approach includes high labouring and survey cost. It also renders workers in unsafe conditions with unreliable inspection results. For instance, an inspector can measure the 2D dimensions of surface damage but cannot calculate its underground depth and variation extent. Due to the availability of low-cost and highly accessible sensors, the emergence of technology of visual evaluation is replacing manual inspection for pavement distress acquisition. Some deploy GPS in the data acquisition vehicle, coordinating with the information from a camera, laser and other sensors for automatic processing in spatial resolution.

5.1. Data collection and image capture

Digital cameras are a widely applied technology in distress detection. Most 2D image studies have led to a high accuracy rate in recognition and classification. The camera devices can be carried in vehicles whilst capturing other information such as vehicle speed, acceleration, longitudinal and transverse position, height, and environmental factors. After collecting the relevant information, the raw image data needs to be cleaned and pre-processed to avoid deformations and exclude noises.

Line-Scan cameras are normally used for crack detection from 2D images because of their high acquisition rate and high resolution. One of the difficulties for computer vision algorithms is the variability of the lighting conditions. Crack detection using 3D data is a new line of research and application [10], which can be acquired by 3D scanners and LiDAR devices and transferred to point cloud data. An example in Fujita et al. [24] builds a Laser Road System (LRIS) on a vehicle employing a Differential Global Position System (DGPS), a high-definition camera, and a high-speed area scan camera. The geometrical information is obtained with 3D detection captured by the laser line projected with the scanning camera.

Recently, new pavement defect detection applications were developed that utilised Unmanned Aerial Vehicles (UAVs) to avoid blocking the traffic flow and to access remote sensing. Branco et al. [9] use a CMOS sensor attached to a UAV platform called MaNIAC-UAV (Methodology for Asphalt Automatic Characterization – using Unmanned Aerial Vehicles). The sensor's resolution was 16 megapixels, while the flight altitude was 130 m with a spatial resolution of 4.0 cm per pixel. Zhang et al. [69] designed programs for UAV intelligent control modulus and defect detection modulus. After the automatic flight task, they combined image processing techniques with typical

edge detection algorithms. The method improved the accuracy and efficiency of remote monitoring and established the basis for large-scale automatic detection.

5.2. Other sensor-based data

Chellaswamy et al. [17] proposed an internet of things-based road monitoring system (IoT-RMS) to identify potholes and humps in the road. An accelerometer has been included with the ultrasonic sensor to measure the degree of variety present in the signal, applying the honey-bee optimisation (HBO) technique. Moreover, Due to the common limitation of the nature of photographic images, the internal features of objects cannot be fully discovered. In order to overcome the drawbacks of vision-based methods, Yang et al. [66] proposed an approach for detecting cracks in infrared thermal imaging steel sheets using Faster R-CNN. Their study also collected 3,000 infrared thermograms labeled for penetrating cracks, non-penetrating cracks, and surface scratches.

Furthermore, Ground Penetrating Radar (GPR) can recognise subsurface damage using the emission of short pulses of electromagnetic energy without deepening into the buried area. Lit et al. [37] proposed an effective method to automatically perform the recognition and location of concealed cracks based on 3-D ground penetrating radar (GPR) and YOLO deep learning models. A dataset containing 303 GPR images and 1306 cracks was constructed as part of their study.

5.3. Dataset of road defects and other surface damages

2D images are the primary data source discussed in this section because it is well studied and widely applied, which has obtained a solid base to compare and analyse. Some video frames, 3D images and stereo vision data are enlisted but very limited and less pertinent to our research focus. After obtaining the 2D images from camera devices, they are cleaned and pre-processed before going through image-processing or modelling procedures. Table 4 gives an overview of a public dataset on road defects with annotations to provide benchmarks for solid evaluation.

5.3.1. Image-level dataset

Özgenel et al. [42] collected various concrete images from walls and concrete floors in campus buildings of Middle East Technical University (METU). The dataset is obtained by extracting 40k 224 × 224-pixel image patches from 500 (4032 pixels to 3024 pixels). Here, we abbreviate the dataset to CCIC (crack image classification). The two classes of concrete images – con-crack and crack, are balanced in number. The total number of CCIC dataset are

Table 4

Some examples of public datasets for road defect

Reference	Dataset	Goal	Label level	Data size/unit	Total number
Zou et al. [71]	CrackTree206	Detect road cracks with similar background texture	Pixel-level	800 × 600 pixels	206
Eisenbach et al. [20]	GAPS v1	Pavement distress detection	Batch-level	64 × 64 pixels	32k
Stricker et al. [56]	GAPS v2	Pavement distress detection on different classes and patch levels	Batch-level	64 × 64 pixels and various scales	50k
Li et al. [36]	AEL	Road crack detection	Pixel-level	800 × 800 pixels	35
Shi et al. [53]	CFD	Road crack detection	Pixel-level	480 × 320 pixels	118
Yang et al. [65]	Crack500	Pavement crack detection	Pixel-level	2000 × 1500 pixels	3368
Guzmán et al. [28]	CCSAD	Hazardous road conditions	Stereo-level	1096 × 822 pixels	40 sequences
Nienaber et al. [41]	Sunny	South African roads with potholes	Image-level	3880 × 2760 pixels	48913 frames
Özgenel et al. [42]	CCIC	Classify cracks and non-cracks on concrete surfaces	Image-level	224 × 224 pixels	40k
Maeda et al. [14]	RDDC2018	Classification of 8 categories of road damages	Image-level	600 × 600 pixels	9053
Arya et al. [6]	RDD2022	Classification of 9 categories of road damages	Image-level	600 × 600 pixels	47,420

randomly split into 70%, 15% and 15% of training, validating and testing data. The CCIC dataset provides enough data for training and testing and is used to test prototype models for road crack image classification. It highlights the crack features with a bright look and high-quality effect. As a result, concrete damages show clear boundaries and contrasts with the background in colour and intensity. In their study, a multidimensional performance analysis of highly acknowledged pre-trained networks (e.g., AlexNet, GoogleNet, ResNet50, ResNet101, ReNet152, Vgg16, and Vgg19) concerns the size of the training dataset, depth of networks, the number of epochs for training, and the expandability to other material types.

Nienaber et al. [41] annotated 53 images containing 97 potholes from the newly created pothole image library. The video frames were obtained while driving at 40 km/h, from which the images of resolution size 3880 × 2760 pixels are selected to involve one pothole in a single frame. They included pictures from different situations, for example, driving while pointing toward the sun, having the sun on the right of the vehicle, and potholes close to trees with huge shadows. The entire dataset was partitioned into simple and complex groups. Each dataset contains folders containing the training (positive and negative) images and a set of positive test images.

Maeda et al. [14] have collected the road damage dataset as a benchmark for evaluating state-of-art computer vision techniques for detection and classification. The Road Damage Detection Classification (RDDC2018) Dataset is based on eight classes: wheel mark part, longitudinal construction joint part, equal interval, transverse construction joint part, partial & overall pavement, rutting, bump, pothole, and separation, white line blur, and crosswalk blur. The Japanese government uses the eight road categories as a classification standard to inspect the road. It is the most completely categorised and is extensively used in road damage detection applications for object detection purposes. RDDC includes 2k to 4k instances and less than 1k instances of other classes.

Nevertheless, the data structure distribution is imbalanced as there are more road cases of cracks than potholes or significant damage. In 2022, Arya et al. [6] released dataset of 47,420 road images collected from India, Japan, the Czech Republic, Norway, the United States, and China and it was used in the Crowdsensing-based Road Damage Detection Challenge (CRDDC), a Big Data Cup organized as a part of the IEEE International Conference on Big Data'2022.

5.3.2. Patch-level dataset

The German Asphalt Pavement distress (GAPs) dataset provides high-quality images collected by the mobile mapping system STIER deployed with laser sensors and camera systems for surface measurement. The distress annotations are provided according to Road Monitoring and Assessment (RMA), which fulfils German federal regulations. GAPs v1 [20] contains 32k small batches of 64 × 64 pixels extracted from annotated regions and intact surfaces in HD images. The training, validation and testing set come in 32k samples in which we use 32k instances for training and 13k for testing. The GAPs v2 dataset [56] has 2 468 grey-valued images in total, separated into 1417 training images, 51 validation images, 500 validation-test images, and 500 test images. They created a training set of 50 000 samples with 10,000 samples of the validation set, validation-test set, and test set. Surface defect classes in the GAPs dataset comprise single/multiple cracking, longitudinal/transversal cracking, alligator cracking, and sealed/filled cracks.

5.3.3. Pixel-level dataset

The CrackTree206 dataset [71] consists of 206 road surface images taken from spacious road textures. It has 800 × 600 pixels RGB crack images manually labelled as ground truth with a delineated centre line. The pavement images are used to detect road cracks automatically. AEL [36] comprises three Aigle-RN, ESAR and LCMS data, including 58 crack images of around 800 × 800 pixels. The dataset is gained at traffic speed for occasionally checking the French asphalt surface condition utilising the Aigle-RN framework. ESAR has 15 images captured by a static acquisition system with no controlled lighting. LCMS contains five pixel-wise annotated crack images. Yang et al. [65] collected a pavement crack dataset Crack500 which composes of 500 images captured by a mobile phone on the main road of Temple University, which are of a size of around 2,000 × 1,500 pixels. The dataset is divided into 250 images of training data, 50 images of validation data, and 200 images of test data. They cropped the data to obtain 1896 images for training data, 348 images for validation data, and 1124 images for test data. Shi et al. [53] proposed an annotated road crack dataset CFD (https://github.com/cuilimeng/CrackForest-dataset) with manually labelled crack contours. The dataset consists of 118 images with 480 × 320 pixels resolution. The device used to acquire the images is an iPhone5 with a focus of 4 mm, aperture of f/2.4 and exposure time of 1/135 s.

5.3.4. Stereo-level dataset

Guzmán et al. [28] collected around 500 GB of data corresponding to more than 96,000 stereo pairs distributed in 42 sequences in 1 h and 20 minutes of stereo sequences. The stereo pairs were chosen at 20 fps using a rectified radiometric 8-bit pixel depth and made into a CCSAD dataset (http://camaron.cimat.mx/Personal/jbhayet/ccsad-dataset) of 1096 × 822 pixels in hazardous road condition of developing countries including abundant potholes, varying speed bumpers and peculiar flows of pedestrians. In addition, they recorded high-resolution stereo datasets complemented with direction and acceleration information captured from IMU and GPS in Mexico from a moving vehicle. The dataset, however, is limited to highway traffic with annotations relevant to work zones.

6. Conclusion and future work

The development of computer vision and the deployment of visual technologies have exponentially increased in the past few decades. We reviewed the computer-vision techniques, from image processing to machine-learning methods, across different challenge-focused works of automatic road damage detection. The different classifications of road defects are identified, and the most up-to-date data collection technologies are discussed.

The advantages and limits of the existing methods for automatic road defect detection are outlined in the discussion. There are still plenty of challenges to face in computer vision to support civil engineering research in the future. The common problem is the over-fitting resulting from the DL models that conduct the high precision performance on training data but fails to generalise and identify other datasets. The current studies have not detected crack severity levels nor analysed the impact on road safety issues, despite being one of the most critical factors in road inspection. Due to the variety of road distress types in sparse areas, the cost increases to carry heavy visual devices to different locations for acquiring specific damage classes.

In addition, the abnormalities of asphalt surfaces and the variety of different types of distinct surface materials lead to few successes in automatic surface distress detection with high precision and comprehensiveness.

Conflict of interest

None to report.

References

[1] I. Abdel-Qader, O. Abudayyeh and M.E. Kelly, Analysis of edge-detection techniques for crack identification in bridges, *Journal of Computing in Civil Engineering* **17**(4) (2003), 255–263. doi:10.1061/(ASCE)0887-3801(2003)17:4(255).

[2] I. Abdel-Qader, S. Pashaie-Rad, O. Abudayyeh and S. Yehia, PCA-based algorithm for unsupervised bridge crack detection, *Advances in Engineering Software* **37**(12) (2006), 771–778. doi:10.1016/j.advengsoft.2006.06.002.

[3] D. Ai, G. Jiang, L.S. Kei and C. Li, Automatic pixel-level pavement crack detection using information of multi-scale neighborhoods, *IEEE Access* **6** (2018), 24452–24463. doi:10.1109/ACCESS.2018.2829347.

[4] H.R. Al-Masaeid, Impact of pavement condition on rural road accidents, *Canadian Journal of Civil Engineering* **24**(4) (1997), 523–531. doi:10.1139/l97-009.

[5] D. Arya et al., Global road damage detection: State-of-the-art solutions, 2020, arXiv preprint arXiv:2011.08740.

[6] D. Arya, H. Maeda, Y. Sekimoto, H. Omata, S.K. Ghosh, D. Toshniwal, M. Sharma, V.V. Pham, J. Zhong, M. AlHammadi, M.B. Shami, D. Nguyen, H. Cheng, J. Zhang, A. Klein-Paste, H. Mork, F. Lindseth, T. Seto, A. Mraz and T. Kashiyama, RDD2022 – The multi-national Road Damage Dataset released through CRDDC'2022, 10 2022. [Online]. Available: https://figshare.com/articles/dataset/RDD2022Road Damage Dataset released through The CRDDC, multi-national 2022/21431547.

[7] S. Baskara et al., *Influence of Pavement Condition Towards Accident Number on Malaysian Highway*, IOP Conference Series: Earth and Environmental Science, Vol. 220, IOP Publishing, 2019, p. 012008.

[8] J. Bluett, M. de Aguiar and N. Gimson, Impacts of exposure to dust from unsealed roads April 2017, 2017.

[9] L.H.C. Branco and P.C.L. Segantine, August. MaNIAC-UAV-a methodology for automatic pavement defects detection using images obtained by Unmanned Aerial Vehicles, in: *Journal of Physics: Conference Series*, Vol. 633, IOP Publishing, 2015, p. 012122.

[10] W. Cao, Q. Liu and Z. He, Review of pavement defect detection methods, *IEEE Access* **8** (2020), 14531–14544. doi:10.1109/ACCESS.2020.2966881.

[11] Y.-J. Cha, W. Choi and O. Büyüköztürk, Deep learning-based crack damage detection using convolutional neural networks, *Computer-Aided Civil and Infrastructure Engineering* **32** (2017), 361–378. doi:10.1111/mice.12263.

[12] Y.J. Cha, W. Choi and O. Büyüköztürk, Deep learning-based crack damage detection using convolutional neural networks, *Computer-Aided Civil and Infrastructure Engineering* **32**(5) (2017), 361–378. doi:10.1111/mice.12263.

[13] Y.J. Cha, W. Choi, G. Suh, S. Mahmoudkhani and O. Büyüköztürk, Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types, *Computer-Aided Civil and Infrastructure Engineering* **33**(9) (2018), 731–747. doi:10.1111/mice.12334.

[14] K. Chaiyasarn, Damage detection and monitoring for tunnel inspection based on computer vision, 2014.

[15] S. Chambon, P. Subirats and J. Dumoulin, Introduction of a wavelet transform based on 2D matched filter in a Markov random field for fine structure extraction: Application on road crack detection, in: *Image Processing: Machine Vision Applications II*, Optics and Photonics, Vol. 7251, International Society for, 2009, p. 72510A.

[16] C.Y. Chan, B. Huang, X. Yan and S. Richards, Investigating effects of asphalt pavement conditions on traffic accidents in Tennessee based on the pavement management system (PMS), *Journal of advanced transportation* **44**(3) (2010), 150–161. doi:10.1002/atr.129.

[17] C. Chellaswamy, H. Famitha, T. Anusuya and S.B. Amirthavarshini, IoT based humps and pothole detection on roads and information sharing, in: *2018 International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC)*, 2018, March, IEEE, 2018, pp. 084–090.

[18] F. Chen and M.R. Jahanshahi, NB-CNN: Deep learning-based crack detection using convolutional neural network and naïve Bayes data fusion, *IEEE Transactions on Industrial Electronics* **65** (2018), 4392–4400. doi:10.1109/TIE.2017.2764844.

[19] A. Coates, A. Ng and H. Lee, An analysis of single-layer networks in unsupervised feature learning, in: *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics, JMLR Workshop and Conference Proceedings*, 2011, pp. 215–223.

[20] M. Eisenbach, R. Stricker, D. Seichter, K. Amende, K. Debes, M. Sesselmann, D. Ebersbach, U. Stöckert and H.-M. Gross, How to get pavement distress detection ready for deep learning? A systematic approach, in: *Int. Joint Conf. on Neural Networks (IJCNN)*, IEEE, Anchorage, USA, 2017, pp. 2039–2047.

[21] S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska and B. De Schutter, Deep convolutional neural networks for detection of rail surface defects, in: *2016 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2016, pp. 2584–2589.

[22] S. Faghih-Roohi, S. Hajizadeh, A. Núñez, R. Babuska and B. De Schutter, Deep convolutional neural networks for detection of rail surface defects, in: *2016 International Joint Conference on Neural Networks (IJCNN)*, Vancouver, BC, Canada, 2016, pp. 2584–2589. doi:10.1109/IJCNN.2016.7727522.

[23] J. Fu, H. Zheng and T. Mei, Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4438–4446.

[24] Y. Fujita, Y. Mitani and Y. Hamamoto, A method for crack detection on a concrete structure, in: *18th International Conference on Pattern Recognition (ICPR'06)*, Vol. 3, IEEE, 2006, pp. 901–904. doi:10.1109/ICPR.2006.98.

[25] K. Gopalakrishnan, Deep learning in data-driven pavement image analysis and automated distress detection: A review, *Data* **3**(3) (2018), 28. doi:10.3390/data3030028.

[26] K. Gopalakrishnan, S.K. Khaitan, A. Choudhary and A. Agrawal, Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection, *Construction and Building Materials* **157** (2017), 322–330. doi:10.1016/j.conbuildmat.2017.09.110.

[27] K. Gopalakrishnan, S.K. Khaitan, A. Choudhary and A. Agrawal, Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection, *Construct. Building Mater.* **157** (2017), 322–330. doi:10.1016/j.conbuildmat.2017.09.110.

[28] R. Guzmán, J.B. Hayet and R. Klette, Towards ubiquitous autonomous driving: The CCSAD dataset, in: *Computer Analysis of Images and Patterns. CAIP 2015*, G. Azzopardi and N. Petkov, eds, Lecture Notes in Computer Science, Vol. 9256, Springer, Cham, 2015. doi:10.1007/978-3-319-23192-1_49.

[29] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[30] G.E. Hinton, Learning multiple layers of representation, *Trends in cognitive sciences* **11**(10) (2007), 428–434. doi:10.1016/j.tics.2007.09.004.

[31] K. Jamroz, M. Budzyński, A. Romanowska, J. Żukowska, J. Oskarbski and W. Kustra, Experiences and challenges in fatality reduction on Polish roads, *Sustainability* **11**(4) (2019), 959. doi:10.3390/su11040959.

[32] B. Kim and S. Cho, Automated vision-based detection of cracks on concrete surfaces using a deep learning technique, *Sensors* **18**(10) (2018), 3452. doi:10.3390/s18103452.

[33] A. Krizhevsky, I. Sutskever and G.E. Hinton, Imagenet classification with deep convolutional neural networks, *Advances in neural information processing systems* **25** (2012), 1097–1105.

[34] J. Kwon, An analysis of factors affecting water-resevoir phenomenon in rainy situation on highway, Hanyang University, Master Thesis, 2009.

[35] D. Lattanzi and G.R. Miller, Robust automated concrete damage detection algorithms for field applications, *Journal of Computing in Civil Engineering* **28**(2) (2014), 253–262. doi:10.1061/(ASCE)CP.1943-5487.0000257.

[36] H. Li, J. Zong, J. Nie, Z. Wu and H. Han, Pavement crack detection algorithm based on densely connected and deeply supervised network, *IEEE Access* **9** (2021), 11835–11842. doi:10.1109/ACCESS.2021.3050401.

[37] S. Li, X. Gu, X. Xu, D. Xu, T. Zhang, Z. Liu and Q. Dong, Detection of concealed cracks from ground penetrating radar images based on deep learning algorithm, *Construction and Building Materials* **273** (2021), 121949. doi:10.1016/j.conbuildmat.2020.121949.

[38] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama and H. Omata, Road damage detection using deep neural networks with images captured through a smartphone, 2018, arXiv preprint arXiv:1801.09454.

[39] J.S. Miller and W.Y. Bellinger, Distress identification manual for the long-term pavement performance program, *United States. Federal Highway Administration. Office of Infrastructure ...* (2003).

[40] A. Miyamoto, M.-A. Konno and E. Bruhwiler, Automatic crack recognition system for concrete structures using image processing approach, *Asian Journal of Information Technology* **6**(5) (2007), 553–561.

[41] S. Nienaber, M.J. Booysen and R.S. Kroon, Detecting potholes using simple image processing techniques and real-world footage, in: *SATC*, Pretoria, South Africa, 2015.

[42] Ç.F. Özgenel, Concrete crack images for classification, *Mendeley Data V* **2** (2019). doi:10.17632/5y9wdsg2zt.2.

[43] Ç.F. Özgenel and A.G. Sorguç, Performance comparison of pretrained convolutional neural networks on crack detection in buildings, in: *ISARC, Proceedings of the International Symposium on Automation and Robotics in Construction*, Vol. 35, IAARC Publications, 2018, pp. 1–8.

[44] J. Park and D. Yun, Analysis of road accident according to road surface condition, 2017.

[45] P. Prasanna et al., Automated crack detection on concrete bridges, *IEEE Transactions on automation science and engineering* **13**(2) (2014), 591–599. doi:10.1109/TASE.2014.2354314.

[46] S.S. Pulugurtha, V. Ogunro, M.A. Pando, K.J. Patel and A. Bonsu, Preliminary results towards developing thresholds for pavement condition maintenance: Safety perspective, *Procedia-Social and Behavioral Sciences* **104** (2013), 302–311. doi:10.1016/j.sbspro.2013.11.123.

[47] A. Ragnoli, M.R. De Blasiis and A. Di Benedetto, Pavement distress detection methods: A review, *Infrastructures* **3**(4) (2018), 58. doi:10.3390/infrastructures3040058.

[48] R.-S. Rajadurai and S.-T. Kang, Automated vision-based crack detection on concrete surfaces using deep learning, *Appl. Sci.* **11** (2021), 5229. doi:10.3390/app11115229.

[49] Routine Maintenance Guidelines, Transport and Main Roads, Queensland, Australia, 2017.

[50] R. Salakhutdinov, A. Mnih and G. Hinton, Restricted Boltzmann machines for collaborative filtering, in: *Proceedings of the 24th International Conference on Machine Learning*, 2007, pp. 791–798. doi:10.1145/1273496.1273596.

[51] M. Salman, S. Mathavan, K. Kamal and M. Rahman, Pavement crack detection using the Gabor filter, in: *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, IEEE, 2013, pp. 2039–2044. doi:10.1109/ITSC.2013.6728529.

[52] Y. Shi, L. Cui, Z. Qi, F. Meng and Z. Chen, Automatic road crack detection using random structured forests, *IEEE Transactions on Intelligent Transportation Systems* **17**(12) (2016), 3434–3445. doi:10.1109/TITS.2016.2552248.

[53] Y. Shi, L. Cui, Z. Qi, F. Meng and Z. Chen, Automatic road crack detection using random structured forests, *IEEE Transactions on Intelligent Transportation Systems* **17**(12) (2016), 3434–3445. doi:10.1109/TITS.2016.2552248.

[54] M.R. Start, J. Kim and W.D. Berg, Potential safety cost-effectiveness of treating rutted pavements, *Transportation Research Record* **1629**(1) (1998), 208–213. doi:10.3141/1629-23.

[55] R. Stricker, M. Eisenbach, M. Sesselmann, K. Debes and H.-M. Gross, Improving visual road condition assessment by extensive experiments on the extended gaps dataset, in: *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019, IEEE, pp. 1–8.

[56] R. Stricker, M. Eisenbach, M. Sesselmann, K. Debes and H.-M. Gross, Improving visual road condition assessment by extensive experiments on the extended GAPs dataset, in: *Int. Joint Conf. on Neural Networks (IJCNN)*, IEEE, Budapest, Hungary, 2019, pp. 1–8.

[57] Z. Tong, J. Gao, A. Sha, L. Hu and S. Li, Convolutional neural network for asphalt pavement surface texture analysis, *Computer-Aided Civil and Infrastructure Engineering* **33**(12) (2018), 1056–1072. doi:10.1111/mice.12406.

[58] Z. Tong, J. Gao and D. Yuan, Advances of deep learning applications in ground-penetrating radar: A survey, *Construction and Building Materials* **258** (2020), 120371. doi:10.1016/j.conbuildmat.2020.120371.

[59] B.S. Underwood, Z. Guido, P. Gudipudi and Y. Feinberg, Increased costs to US pavement infrastructure from future temperature rise, *Nature Climate Change* **7**(10) (2017), 704–707. doi:10.1038/nclimate3390.

[60] W.H.O.D.o. Violence, I. Prevention, W.H.O. Violence, I. Prevention and W.H. Organization, *Global Status Report on Road Safety: Time for Action*, World Health Organization, 2009.

[61] L. Wang, Z. Zhang and X. Luo, A two-stage data-driven approach for image-based wind turbine blade crack inspections, *IEEE/ASME Transactions on Mechatronics* **24**(3) (2019), 1271–1281. doi:10.1109/TMECH.2019.2908233.

[62] L. Xiong et al., A good practice towards top performance of face recognition: Transferred deep feature fusion, 2017, arXiv preprint arXiv:1704.00438.

[63] T. Yamaguchi, S. Nakamura, R. Saegusa and S. Hashimoto, Image-based crack detection for real concrete surfaces, *IEEJ Transactions on Electrical and Electronic Engineering* **3**(1) (2008), 128–135. doi:10.1002/tee.20244.

[64] B.-M. Yang and J. Kim, Road traffic accidents and policy interventions in Korea, *Injury control and safety promotion* **10**(1–2) (2003), 89–94. doi:10.1076/icsp.10.1.89.14120.

[65] F. Yang, L. Zhang, S. Yu, D. Prokhorov, X. Mei and H. Ling, Feature pyramid and hierarchical boosting network for pavement crack detection, *IEEE Transactions on Intelligent Transportation Systems* (2019).

[66] J. Yang, W. Wang, G. Lin, Q. Li, Y. Sun and Y. Sun, Infrared thermal imaging-based crack detection using deep learning, *IEEE Access* **7** (2019), 182060–182077. doi:10.1109/ACCESS.2019.2958264.

[67] E. Zalama, J. Gómez-García-Bermejo, R. Medina and J. Llamas, Road crack detection using visual features extracted by Gabor filters, *Computer-Aided Civil and Infrastructure Engineering* **29**(5) (2014), 342–358. doi:10.1111/mice.12042.

[68] A. Zhang, K.C.P. Wang, B. Li, E. Yang, X. Dai, Y. Peng, Y. Fei, Y. Liu, J.Q. Li and C. Chen, Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network, *Computer-Aided Civil and Infrastructure Engineering* **32** (2017), 805–819. doi:10.1111/mice.12297.

[69] L. Zhang, W. Xu, L. Zhu, X. Yuan and C. Zhang, February. Study on pavement defect detection based on image processing utilizing UAV, in: *Journal of Physics: Conference Series*, Vol. 1168, IOP Publishing, 2019, p. 042011.

[70] S. Zhong and S.O. Oyadiji, Detection of cracks in simply-supported beams by continuous wavelet transform of reconstructed modal data, *Computers & structures* **89**(1–2) (2011), 127–148. doi:10.1016/j.compstruc.2010.08.008.

[71] Q. Zou, Y. Cao, Q. Li, Q. Mao and S. Wang, CrackTree: Automatic crack detection from pavement images, *Pattern Recognit. Lett.* **33** (2012), 227–238. doi:10.1016/j.patrec.2011.11.004.

[72] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang and S. Wang, Deepcrack: Learning hierarchical convolutional features for crack detection, *IEEE Transactions on Image Processing* **28**(3) (2018), 1498–1512. doi:10.1109/TIP.2018.2878966.