Received 1 September 2024, accepted 12 October 2024, date of publication 16 October 2024, date of current version 30 October 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3481649

TOPICAL REVIEW

Recent Advancements in Deep Learning Techniques for Road Condition Monitoring: A Comprehensive Review

LORENZO MANONI, (Member, IEEE), SIMONE ORCIONI, (Senior Member, IEEE), AND MASSIMO CONTI, (Member, IEEE)
DII—Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, 60131 Ancona, Italy

Corresponding author: Massimo Conti (m.conti@univpm.it)

This work was supported in part by the Framework of the PNRR Italian National Center for Sustainable Mobility (MOST): SPOKE 7 "CCAM, Connected Networks and Smart Infrastructure," WP4.

ABSTRACT Road Condition Monitoring is a critical task for the management and maintenance of the pavement network infrastructure by the authorities. In recent years, the application of Artificial Intelligence (AI) techniques in this domain has experienced a significant growth, driven by the continuous advancements in AI algorithms. This paper presents a comprehensive review of the latest developments in Road Condition Monitoring approaches using AI methods, with a particular focus on Deep Learning techniques, covering works published from 2020 onwards. It highlights novel approaches that have not been thoroughly explored in previous literature reviews. The literature review categorizes studies based on the type of signal data, distinguishing between acoustic, vibrational, and vision-based approaches. For each data type, the paper examines and discuss the most recent advancements and improvements achieved through AI techniques. Additionally, it provides an overview of future directions and identifying key challenges that remain open in the field. In conclusion, relatively few studies have focused on the analysis of acoustic data, although some studies have reported promising results. Methods based on vibrational data typically integrate feature extraction in frequency and wavelet domain with Convolutional Neural Networks or Long Short-Term Memory Networks. Meanwhile, vision-based methods have experienced significant improvements, driven by the constant evolution of Deep Learning architectures. A total of 173 research articles are summarized across 10 tables.

INDEX TERMS Artificial intelligence, convolutional neural networks, deep learning, road condition monitoring, sensors.

I. INTRODUCTION

Road network agencies in Europe spend several billions euros annually for pavement maintenance. These interventions aim to improve pavement conditions and include repairing asphalt damage, reducing roughness, and partially or fully reconstructing pavement sections that have deteriorated due to weather, temperature changes, and heavy traffic.

Road condition monitoring (RCM) is a crucial part of a Pavement Management System. Real-time assessment of the overall infrastructure condition can save significant monetary resources on pavement maintenance by allowing for early corrective actions before severe deterioration occurs. Additionally, this leads to better infrastructure quality and enhanced safety and comfort for users. Automation in road transport plays a crucial role in advancing top policy priorities set by the European Commission. Cooperative, connected and automated mobility (CCAM) is the project that aims to make transport safer. Copernicus, the Earth observation component of the European Union's Space programme, supports the planning, design, construction, and monitoring of road infrastructure by drawing from satellite Earth

The associate editor coordinating the review of this manuscript and approving it for publication was Laxmisha Rai.

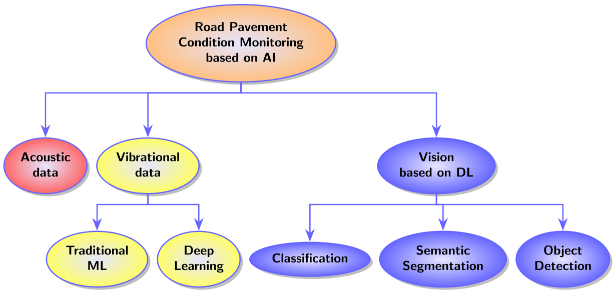

FIGURE 1. Road condition monitoring based on AI techniques.

observation and in-situ data, as reported in the Commission Staff Working Document [1].

In recent years, alongside advancement in scientific literature, numerous European-funded projects have focused on developing innovative solutions on pavement condition monitoring. For instance, the RPB HEALTECH [2] project aimed to deploy a system for detecting pavement defects, for determining their causes, and for assessing their degradation rate using Ground Penetrating Radar (GPR), infraRed thermography and air coupled ultrasonic technologies. An instrumented vehicle was used to perform measurements at a traffic speed without causing traffic disturbances. Conversely, The PAV-DT project [3], proposed an innovative low-cost technology that can be installed in any custom car to measure the International Roughness Index parameter. Furthermore, the PAVE-SCAN project [4] implemented a data driven solution employing participatory sensing to detect, classify and georeference road defects. To this aim, AI algorithms have been applied to data collected by integrated, low-cost sensors based on European Global Navigation Satellite System (EGNSS).

In the European framework, the Italian National Center for Sustainable Mobility (MOST, SPOKE 7 "CCAM, Connected Networks and Smart Infrastructure") [5] is developing solutions for the maintenance of infrastructures, as well as methods for the assessment of the resilience at a network level and for the management of infrastructural assets.

The adoption of AI techniques for Road Condition Monitoring has seen a significant growth in recent years. Sholevar at al., in their review [6] have proven Deep Learning (DL) techniques clearly outperform traditional AI algorithms such as Support Vector Machine (SVM), k-Nearest Neighbor (KNN), Decision Trees (DT), Random Forest (RF). These traditional techniques still offer a low-cost computational alternative for detecting and classifying road anomalies using vibrational data captured by accelerometers. Nevertheless, in their review [7], Ma et al. showed that for vision data, which include RGB images or 3D imaging performed by stereo vision or by GPR, DL methods are almost entirely replacing the older image processing-based algorithms.

Several review papers have covered this topic in recent years. In [8], Ranyal et al. reviewed studies employing AI algorithms and smart sensors for road condition monitoring, providing a comprehensive point of view focused on the advances in sensor technology used for RCM across various platforms such as instrumented ground vehicles, smartphones and Unmanned Aerial Vehicles (UAV). In contrast, the previously cited work [6] presented a review of Machine Learning (ML) techniques for RCM, focusing more on the algorithm perspective rather than on sensor technology.

However, to the best of our knowledge, literature reviews on this subject do not cover the most recent advances in RCM using AI and DL techniques over the last two years. During this period, these have seen significant improvements in algorithms performances for pavement damage classification and numerous novelties regarding system-level and employed technology. Approximately 85% of the research articles examined in this review article were published from 2020 to 2024. These articles have not been covered in previous review papers. The great amount of recent publications highlights the current significance and relevance of Artificial Intelligence methods in road condition monitoring. In this paper a review of the most recent advancements in AI approaches, in particular DL techniques, for RCM was performed by including papers published from 2020 onwards to address the aforementioned gap in literature.

The literature search was conducted by using the Web of Science (WoS) and Scopus databases with three different sets of keyword expressions, detailed in Tab. 1. These sets correspond to three type of sensor data: acoustic data captured by microphones, vibrational data obtained by low-cost and professional accelerometers, and vision data collected using RGB and stereo cameras, CCD sensors, laser and GPR.

The general framework of the review, categorized by AI algorithms and sensor types, is shown in Fig. 1.

Section II details the methods applied to acoustic data. Section III summarizes the methods used for accelerometer data. Section IV presents the DL techniques applied to road imagery. The results are summarized in tables reporting the key aspects of the different research works. Section V discusses the state of the art and future research trends in AI applications for road damage detection. Finally, conclusions are presented in the last section.

II. ACOUSTIC-BASED METHODS

Assessing road pavement quality through acoustic data involves measuring on road/tyre noise using one or more microphones positioned near the tyre or within its cavity.

A. TRADITIONAL MACHINE LEARNING

Existing approaches in the literature for pavement characterization using conventional Machine Learning techniques typically involve extracting features from Power Spectral Density (PSD) of the sound coming from the tyre/road interaction. Vázquez et al. [9] experimentally estimated the correlation between sound spectrum levels, recorded by a microphone positioned near the tyre, and the depth of road macrotexture measured by a laser profilometer. In another

TABLE 1. Keywords used for literature search in Scopus and WoS databases.

Data type	Keywords
Acoustic	("road" OR "pavement") AND (“damage” OR "pothole" OR "anomaly" OR "distress" OR "type") AND ("detection" OR "classification") AND ("deep learning") AND ("acoustic" OR "audio" OR "sound pressure")
Vibrational	("road*" OR "pavement") AND (“damage” OR "condition" OR "pothole" OR "anomal*" OR "distress*" OR "roughness") AND ("detect*" OR "classif*" OR "recognition" OR "estimat*" OR "eval*" OR "assess*") AND (acceler* OR "gyroscope" OR rotation* OR vibration*) AND ("AI" OR "machine learning" OR "deep learning" OR "neural network*" OR "*NN")
Vision	("road" OR "pavement") AND ("image" OR "vision" OR "laser") AND (damage* OR defect* OR condition OR anomal* OR pothole* OR crack* OR distress*) AND (detect* OR classif* OR identif* OR recogn* OR assess*) AND ("AI" OR "machine learning" OR "deep learning" OR "neural network" OR "NN")

work by Alonso et al. [10] a microphone was positioned on the car trunk and SVM technique was applied to signal spectrogram to classify road surface as dry/wet. However, the classification error was notably high due to engine noise interference.

Significantly improved results were obtained by placing microphone inside the tyre cavity, effectively utilizing it as a reverberation chamber. In a study conducted by Masino et al. [11] an acoustic sensor was positioned inside tyre cavity, and the SVM technique was applied to the PSD of acoustic signal to classify five different types of asphalt. The classification accuracy reached 69.9% in the testing set and could be further increased up to 91% by merging two classes and by smoothing classifier output.

B. DEEP LEARNING

Deep learning-based methods have significantly improved the performance of road type or status classification.

In the work of Pratico et al. [12], the seismic waves were captured by an isolated microphone placed on the asphalt near a passing car wheel to identify concealed cracks in the pavement. A dataset consisting by 1D time series data audio responses was used to train a simple one-layer convolutional neural network for crack classification. After hyperparameter optimization a 95.6% accuracy was achieved.

More detailed information about tyre/pavement interaction noise (TPIN) signal can be exploited by feeding a 2D convolutional neural network (2D-CNN) with images obtained through Wavelet transforms or time/frequency analysis instead of raw time-series. In [13] Lee et al. used a costant-Q filter bank was used to extract a 2D image from TPIN data (sampled at 51.2 kHz) captured by two microphones on the rear and front wheels of a car in order to classify 13 different types of asphalt obtained by combining pavement type with its condition snow, humid, dry or wet. A lightweight sequential 2D-CNN was trained using a mean-square-error loss function, optimizing signal window length, and achieving a 95.6% accuracy on the test set. The model was implemented in a real-time classification system in MATLAB Simulink environment.

TABLE 2. Pavement type/anomalies detection using DL with acoustic data.

Ref.	Model	Road type-distress	Sensor position	Classif. accuracy	Implement. platform
[12]	1-layer CNN	Concealed cracks	Seismic waves near wheel	95.6%	PC Desktop
[13]	2D-CNN	13 types of asphalt	front/rear trunk	95.67%	MATLAB Simulink
[14]	Tiny 2D-CNN	dirty and grass high roughness cracks and potholes	tyre cavity	92.7%	ESP32 Microcontroller 208.17 ms prediction latency

Staderini et al. [14] implement a real-time identification system that uses a ESP32-WROOM-32D module for data acquisition and to run inference. A very tiny 2D-CNN was trained with sound data capture by a microphone placed inside the tyre cavity to classify road types between: dirty and grass road, high roughness and road with cracks and potholes. The model, which takes as input 2D time-frequency analyses of audio signal, was converted in TensorFlow Lite (*TFLite*) for the microcontroller and a dedicated firmware was developed for data acquisition, preprocessing, inference and sending output result to BLE devices. A latency of 208.17 ms was obtained for the prediction.

Tab. 2, categorizes the papers that apply DL methods to acoustic data.

III. VIBRATIONAL-BASED METHODS

The recent literature trend concerning vibrational methods is focused to the application of ML and DL techniques.

Threshold techniques were initially favored thanks to their intuitive nature, simplicity and low computational requirements. Parameters, such as root mean square, standard deviation [15], local peaks [16] and derivative of vertical acceleration component [17] are generally used to identify abnormalities. While threshold techniques achieve high detection accuracy [8] and they are suitable for a real-time implementation, they struggle to differentiate multiple types of irregularities and often require a recalibration [18] to ensure reproducibility.

Machine and deep learning techniques demonstrated superior performance compared to threshold methods. Typically, this kind of studies requires initial collection and labeling of accelerometer and gyroscope data. Labeling is often synchronized with GPS or video data and acceleration to ensure accuracy. However, the high amount of data required for model training poses a significant challenge for ML approaches, requiring a standardized and automated data collection process. To address this problem, a customized app is often developed when smartphone sensors are used.

Some studies have proposed crowdsourcing-based solutions where individual users upload accelerometer, gyroscope and GPS data to cloud-based or server-based platforms to expedite the collection of training data for model development.

A. TRADITIONAL MACHINE LEARNING

Due to their low computational requirements, traditional ML methods offer a viable solution for implementing real-time identification systems.

In [19], Basavaraju et al. conducted a comparative analysis of SVM, DT and multi-layer perceptron (MLP) approaches to classify road condition (smooth, pothole and crack) using acceleration data from a smartphone fixed to the windshield. The proposed framework is suitable for a real-time implementation, primarily due to the computational burden associated with time, frequency and wavelet feature extraction, rather than the classification algorithms themselves. MLP classifier outperformed SVM and DT with a 92.12% accuracy on the test set.

A real-time speed bump and pothole identification system was proposed in [20] by Andrade et al. utilizing an Arduino 33 BLE device connected to a 3D digital accelerometer (LSM6DS3 by STMicroelectronics). An unsupervised algorithm with low computational cost was employed for classification. Similarly, Egaji et al. [21] developed a real-time pothole detection framework by extracting raw time features from smartphone 3-axis acceleration data and by feeding them to different ML classifiers as Naive Bayes (NB), Logistic Regression (LR), Support Vector Machine (KSVM), KNN and RF.

Ferjani and Alsaif [22], proposed a ML benchmark using a synthetic dataset generated with the Pothole Lab, presented in [23] by Carlos et al. and a real world dataset collected by González et al. [24]. The real world dataset, presented in [24], was obtained using multiple smartphones positioned in different positions within a car and connected to a tablet for acceleration data collection. The abnormalities like potholes, speed bumps, metal bumps, worn road and regular road were distinguished. The benchmark employed SVM, DT and MLP techniques on time, frequency and Daubeshies Wavelets features extracted from acceleration data. MLP trained with time and Wavelets features obtained the highest classification accuracy of 52% in the test set.

Recent advances in traditional ML methods have focused on candidate event window selection with threshold or Dynamic Time Warping (DTW) techniques. For instance, Sattar et al. [25] identified potential road anomalies with a threshold method based on the correlation between instantaneous and the averaged re-oriented acceleration vectors. 3-axis acceleration data was collected using smartphones affixed to the dashboards of multiple cars before a re-orientation was performed. Then unsupervised learning with a Gaussian Mixture Model approach was used to classify anomalies between high severity or lower severity. Du et al. [145] proposed a threshold method to identify potential event windows while compensating for vehicle speed variation effect, rarely incorporated into the features extracted for traditional ML. Zheng et al. [27] used a DTW method to select candidate events for potholes, speed bump and metal bumps from acceleration data provided by a smartphone fixed to the rear of an electric motor, followed by KNN classification. Additionally, Chibani et al. [28] employed a Dynamic Sliding Window (DSW) approach as an alternative to the traditional Static Sliding Window (SSW) method with a fixed window length, for event detection to face the problem of variable length events due to different speeds. Different classifiers, like KNN, SVM, Neural Network (NN), DT,RF were tested on synthetic datasets in [23] and [27] concluding that the DSW method outperforms the SSW technique.

An alternative to DTW or threshold method for vibrational signal analysis, was presented by Zhour et al. [110], employing DL image object detection to identify manholes from smartphone images. Acceleration and angular velocity coming from the same smartphone were combined with an SVM classifier to recognize manhole severity.

One of the critical aspects in vibrational-based techniques is the noise affecting the sensors. The issue of smartphone acceleration noise was addressed by Dong et al. [30], where sensors were mounted on car windshield to classify normal road conditions, potholes, patching, and distortion. Filtering process in the PSD domain followed by inverse Fast Fourier Transform (IFFT), was conducted to mitigate engine vibration noise before extracting time domain features for a KNN classifier.

Shtayat et al. used in [31] a high precision piezoelectric sensor (356B18 by PCB Piezotronics) to classify various types of road distress, including alligator cracks, edge cracks, longitudinal cracks, and patches, based on collected 3-axis acceleration data. SVM outperformed RF and DT classifiers after training with time domain features.

Furthermore, Martinelli et al. [32] proposed an original feature extraction methodology. This approach involved computing entropy and coefficient of variation of Short-Time Fourier Transform (STFT) sub-bands of a z-axis acceleration data. SVM, DT and KNN were used to classify no distress, long-term distresses, and short-term distresses in road condition, such as fatigue cracking or potholes, respectively.

B. DEEP LEARNING

Deep Learning methods achieve significant performance improvement over traditional ML classifiers not only because they learn to extract the optimal features for the problem, but also because they can incorporate additional input signals. For example, 3-axis rotation has been included by Basavaraju et al. [19] or vehicle speed by Sabapathy and Biswas [33]. Speed is a key variable that influences the acceleration response, as reported in the work of Wang et al. [34]. However, real-time classification implementation poses a challenge since prediction must be made within the duration of the signal window. Consequently, most studies perform offline classification.

Luo et al. achieved in in [35] a very high accuracy for road anomaly classification using Deep Feedforward Networks (DFN), 2D-CNN and Recurrent Neural Networks (RNN) with raw time data, including 3-axis acceleration, rotation velocity, wheel speeds, spindle, and shock responses. However, multiple anomaly categories (pothole, bump, gravel, and other different pavement types) were merged in a single class, thus obtaining a simple binary normal/abnormal classification problem.

Numerous studies have focused on the feature extraction procedure, a critical factor for the overall performance of the classification framework.

Baldini et al. [36] made a comparison between 1D-CNN with raw time data as input and a 2D-CNN with STFT magnitude, STFT phase and Morlet Continuous Wavelet Transform (CWT) to classify road anomalies and obstacles using acceleration and rotation data from a sensor (Xsens Mti by Movella) fixed on car dashboard. The STFT magnitude applied to re-orientated z-axis acceleration provided the best accuracy, achieving 97.21% in test set. In [37], Cheng et al. compared the performance of STFT and Wavelet Transform (WT) applied to 3-axis smartphone acceleration data, by feeding them into a 2D-CNN to distinguish normal road from transverse cracks or manholes. WT transform performed best, reaching a 97.53% accuracy. Martinez-Ríos et al. [38] proposed the Generalized Morse Wavelet to detect pavement transverse cracking from the vertical acceleration captured by a high cost piezoelectric sensor (CT1100L) mounted on a car tire suspension knuckle. An FFT-based denoising technique was applied before computing the 2D image, which was then input to well known pre-trained Deep Neural Network (DNN) architectures used for images: GoogleNet [39], SqueezeNet [40], ResNet18 [41]. ResNet18 achieved an accuracy of 91.18%.

Other studies have focused on more specific pavement characterizations. For example in [42], Varona et al. used a 3-axis smartphone acceleration data together with a 1D-CNN, Long Short-Term Memory (LSTM) network and the Reservoir Computing approach, presented by Bianchi et al. [43], to distinguish between road anomalies and acceleration variations due to some driver actions. In [33], Sabapathy and Biswas classified road pavement conditions as good, medium or bad using the Pavement Surface Evaluation and Rating (PASER) system labels available for the cities considered. An On-Board Diagnostic II data logger was used to collect 3-axis acceleration, rotation velocity and speed data which were input to a 1D-CNN. A bump characterization system was presented by Salman and Mian in [44], where acceleration, rotation and speed data of a smartphone fixed on car dashboard were used to train a 1D-CNN to classify different types of bumps such as: flat-top, sinusoidal, thump, round-top.

Advances in DNN model in this context include hybrid-architectures and multiple features or data fusion approaches.

In the work by Setiawan et al. [45], an unrolled Generative Adversarial Networks (GAN) was employed for data augmentation, aiming to extend a small dataset comprising smartphone acceleration data. The objective was to discern various road conditions, including flat road, potholes, speed bumps and rough road ultimately enhancing classification performance through Deep Convolutional Neural Network (DCNN).

Setiawan et al. [46] proposed a 1-dimensional semantic segmentation methodology, where in each signal window was assigned a time-varying label. Data acquisition involved mounting smartphones on a motorcycle to capture different environmental factors such as road conditions (flat, potholed, speed-bumped, rough), human activity, and machine-induced vibrations. Furthermore, an hybrid Bidirectional Long Short-Term Memory (BiLSTM) Skip-U-Net architecture was developed, and trained with raw time z-axis acceleration/rotation data. This model demonstrated superior performance over conventional 1D and 2D U-Net, as evidenced by higher accuracy and mean Intersection over Union (mIoU) scores.

Menegazzo et al. [47], achieved a classification of distinct road surfaces (paved, unpaved and cobblestone) utilizing data from multiple gyroscope, accelerometer and magnetometer sensors (MPU-9250 by InvenSense) affixed to different vehicles. Information regarding 3-axis acceleration, rotation, and speed was integrated to discern the nuances of road types, facilitating comprehensive analysis and classification. A 1D-CNN, LSTM and a hybrid 1D-CNN-LSTM model were trained on raw time data, CNN model gave the highest test accuracy 93.17%. The benchmark is available in [48]. It can be also configured for speed bump detection as in [49] where a CNN-LSTM outperforms the other state-of art models. In a related work by Narit et al. [50], a squeeze-excitation module 1D-ResNet was introduced for the same PVS benchmark as discussed in [47], resulting in a notable enhancement in accuracy. Singh et al. [51] proposed a binary feature extraction custom layer tailored for an LSTM architecture, aiming for pothole detection using smartphone accelerometer and gyroscope data. Among the classical LSTM variants examined, the Improved-LSTM model developed in this research exhibited the highest accuracy.

Gated Recurrent Unit (GRU) and LSTM performance were compared by Ibrahim et al. [52] for classifying bump, pothole, rough and smooth road using raw smartphone acceleration data, with GRU achieving better accuracy.

Siddiqui et al. [53] addressed the issue of road anomaly event length variability by implementing an initial binary BiLSTM detector to classify normal versus abnormal event followed by another BiLSTM classifier to differentiate among specific the anomaly type, including cat's eye, manhole, pothole and speed bump. Accelerometric sensors mounted on car wheel and dashboard were used to collect 3-axis acceleration data, from which instantaneous frequency and spectral entropy features were extracted and used as model inputs.

A lightweight hybrid model was implemented by Raslan et al. [54] to classify road segments into categories such as normal road, speed bump, potholes, bad road. The dataset comprised 3-axis acceleration, rotation and speed from a MPU-9250 sensor. A SepConv1D-BiLSTM architecture was proposed, employing raw time data for the BiLSTM block and FFT data for a Separable-1D-Conv block. The features from these block were fused to generate the final prediction. This model not only achieved the highest accuracy compared to the other state-of-art networks, but also maintained a low computational cost.

Pandey et al.in [55] presented an peculiar framework that fused image and acceleration data using a hybrid 2D-CNN + 1D-CNN network for pothole detection. Data were collected from smartphone mounted on windshield, capturing both 3-axis acceleration sampled at 100 Hz and 128 × 128 images. Data augmentations techniques such as scaling and permutation were applied to acceleration data, while geometric transformation and brightness variation were applied to images. The hybrid model outperformed 1D-CNN and 2D-CNN models, which utilized only acceleration or image inputs, achieving a test accuracy of 95.71%.

The existing literature on International Roughness Index (IRI) estimation using ML methods is relatively sparse. The primary challenge in this domain is the need of ground truth IRI measurement for supervised learning, which typically requires expensive inertial or laser profilometers.

Aboah Adu-Gyamfi [56] employed a DL method based on entity embedding to estimate road IRI from smartphone acceleration data and historical IRI values. Reference roughness data were sourced from an on line viewer portal (*MoDOT's ARAN*). Raw time-series data were preprocessed using Empirical Mode Decomposition (EMD) techinque to extract different resolution modes, and PSD parameters of each mode were computed as input features for a CNN. Instead of direct data collection, vehicle response to road profiles can be obtained using quarter-car models. In Jeong and Jo [57] a road profile with a IRI values was generated following the method presented by Tyan et al. [58], and the acceleration response simulated using an half-car model was employed to train a 2D-CNN model.

Liu et al. [59] proposed a semi-supervised method for IRI estimation by fitting a nonlinear PSD model using already available IRI values within a collaborative framework. Previously calculated IRI values from an inertial profilometer data were used by Jeong et al. [60] to develop a 2D-CNN model named IRI-Net. This model was trained with smartphone acceleration, angular velocity and speed data collected with several cars and phone positions. Peigen et al. [61] used a randomly generated profile with variable variance to simulate vehicle acceleration response, training an LSTM with vertical acceleration and speed data. Transfer learning from Vehicle-1 and Vehicle-2 dataset was employed to improve performance. The method was validated using real data collected with a 9-axis accelerometer and inclinometer with BLE connection (WT901BLECL5.0 by WIT motion) and real IRI values measured with a three-meter straight edge.

Tab. 3 reports the available datasets and benchmarks for the vibrational signals. A great amount of accelerometric data (trips, cars, distress type) are freely available to be used as benchmark. Video and GPS coordinates are associated to the 3axis accelerometer data. Different types of accelerometer were used, each with its own signal-to-noise ratios and accuracy levels.

Tab. 4 summarizes the state of the art for DL techniques for anomaly detection using vibrational signals. The techniques are primarily employed to detect the presence of potholes on the road. The feature extraction process is a critical factor in the overall performance of the classification. The most commonly used architecture is the CNN. The dataset vary in terms of acquisition systems and sensor accuracy, making direct comparisons difficult. Nevertheless, the accuracy of the results is over 90%.

Tab. 5 summarizes data acquisition setups of vibrational signals utilized by DL methods, including both publicly available and proprietary systems. Various types of 3-axis accelerometers, ranging from piezoelectric sensors to MEMS embedded in common smartphones, each offering different signal-to-noise ratios and accuracy levels. The sensor position in the vehicle is usually selected for ease of application, no discussion is presented on the relevance of its position. The availability of low-cost hardware implementations for the solutions presented in the table facilitates data acquisition.

C. CROWDSOURCING STUDIES

Some recent crowdsourcing solutions for road condition monitoring with vibrational data have performed aggregation and/or clustering of single users predictions on a cloud or server in order to produce a reliable global estimation of road segments condition.

A pothole system called *DeepBus* was presented by Bansal et al. [65]. In this system 3-axis acceleration and angular velocity data were collected with a accelerometric sensor (MPU-6050 by InvenSense) on a motorbike and transmitted to a Raspberry Pi3, running a ML classifier implemented in Python, sent predictions to a server. Additionally, a companion app was developed to display road potholes on a map in real-time.

An Android app called *RoadCare* was developed by Tiwari et al. [66]. This app collects 3-axis acceleration data and includes an embedded NN model that classifies road condition as good, medium or bad. Users' prediction confidences are sent to a server, which computes a global 1-10 score rating by applying an unsupervised clustering algorithm to the prediction confidences.

Other works preferred to leave feature extraction and/or road status estimation to servers or super nodes within

TABLE 3. List of available benchmarks for the vibrational signals.

Reference	Distress type	Dataset details	Sensor data	Sensor details	Sampling frequency
[23]	Potholes Metal bumps Asphalt bumps Worn out road Regular road	Different car models. More than 500 events were recorded	3-axis acc.	smartphone accelerometers	50 Hz
[48], [50]	Unpaved Cubblestone paved(normal)	9 datasets: combination of different cars and drivers	3-axis acc. 3-axis rotation speed	MPU-9250	100 Hz
[54]	Normal Cateye Manhole Pothole Speed bump	10 cars 2454 anomalies	3-axis acc.	ADXL362 VK2828U7G5LF sensors	93 Hz
[39]	Transverse cracks	327 samples	3-axis acc.	Piezoelectric CT1100L	1380 Hz

TABLE 4. Resume of state-of-art deep learning anomaly detection with vibrational signals.

Ref	Distress type	Model arch.	Input data	Dataset details	Performance	Main novelty	Main disadvantage
[43]	Potholes speed bumps driver actions	CNN LSTM RC	1D Time	608 normal events 302 potholes 218 speed bumps 98 street gutters Augmentation by stretching	CNN pavement type: 85% CNN driver actions: 93%	Identify different kinds of surface, distinguish potholes speed bumps and driver actions	Low number of different driver action samples
[37]	Normal obstacles road anomalies	CNN	1D Time STFT Spectrogram Morse CWT	12 cars and 20 loops 4 obstacles 15 anomalies 9600 segments 11/12 train 1/12 test	Spectrogram data outpeforms all other domains. Top accuracy: 97.21%	Proposed spectrogram instead of 1D time, STFT or CWT	Only 19 different physical anomalies & obstacles
[38]	Transverse cracks manhole normal	CNN	STFT CWT	1792 training samples 768 testing	WT has best accuracy: 97,53%	Comparison between WT and STFT	Data collected with a unique constant speed
[48], [50], [51]	Unpaved Cubblestone paved(normal)	CNN LSTM LSTM+CNN	1D Time	9 datasets different cars and drivers	CNN best perf. accuracy: 93.17%	Available PVS benchmark	Sensor position below & above suspension not practical
[55]	normal road speed bump potholes bad road	SepConv1D + BiLSTM	1D Time to SepConv1D FFT to BiLSTM	301 sequences	Proposed model acc: 93.1%	Proposed a lightweight architecture. Comparison with state-of-art models. Estimation of Road Quality Index	Proposed a model for IoT but did not evaluate execution performance on an IoT device
[63]	potholes	2D-CNN	raw time	2 datasets 432 and 1132 samples	5 architectures best model: 93.24% acc	Proposed 5 deep network architectures	No comparison with baseline methods. Same car, driver and sensor position for all datasets
[54]	Normal Cateye Manhole Pothole Speed bump	BiLSTM	Istantaneous frequency Spectral Entropy	10 cars 2454 anomalies	VISC trespasses FLSC and VLSC. Accuracy> 90% for all classes	Variable length event model Events detection and classification networks Vehicle indepedent model	Very simple classifier for variable length event detection
[47]	float road pothole speed bump rough surface human movement machine vibration	BiLSTM Skip-U-Net	1D time	Motorcycle used 6 routes 26 km length	Hybrid network has 90 % accuracy	Performs segmentation on raw time data	No comparison with high performing state-of-art semantic segmentation architectures
[52]	Pothole smooth road	Improved LSTM	1D time	5 trips with one vehicle	ILSTM has 99.45% accuracy	LSTM with custom feature extraction layer Comparison with other models	Obtains high performances but focuses only on pothole detection
[56], [64]	Potholes	1D-CNN + 2D-CNN	1D time + 2D image	33360 images and data samples Train/val/test: 70/15/15	Fused model accuracy: 95.71%	Images and acceleration data fusion	Data collected only with one vehicle
[65]	Potholes	LSTM	1D time acc+video	Different cars 10 m segments	Join model best accuracy: 92.17%	Crowdsourcing system with acceleration and video data fusion	Feature extraction technique from video and acceleration could be improved

the network, which have virtually unlimited computational capabilities.

Mihoub et al. [67] developed the *Road Scanner* app to classify road segments as normal, pothole, bump or other. However, the app requires users to actively select the anomaly type on the app and collect data. Wu et al. [68] conducted a comparative study of ML classifiers to identify road condition as good, poor or extremely bad, for a crowdsourcing monitoring system.

A LoRaWAN based network was designed in [69], where smartphones are connected to gateways that send smartphone acceleration data to a server. An app was developed to monitor road condition in real-time.

A 2D-CNN + GRU model was proposed in the crowdsourcing approach in [70] to mark road as flat, satisfactory or unsatisfactory by using CFS features extracted from smartphone acceleration data.

Instead, Bustamante et al. presented in [71] a V2I-Fog computing architecture based on on-board unit, which collects and sends the data to a roadside Unit, which is responsible for preprocessing and classification using ML methods to categorize road condition as plain, pothole, speed

TABLE 5. Resume of data acquistion setups for vibrational signal DL methods.

Ref	Sensor data	Sensor details	Sampling frequency	Sensor position	Speed dependency	Hardware
[37]	Z-acc. Y-ang. velocity	Xsens Mti	50,100, 200, 250 Hz	fixed on car dashboard	Different speeds each loop	GNSS receiver
[38]	3-axis acc.	smartphone sensor	100 Hz	back seats	constant 30 km/h	camera
[48], [50], [51]	3-axis acc 3-axis angular velocity speed	MPU-9250	100 Hz	Front axle left/right Control arm	from 0 to 91 km/h	HP Webcam HD-4110 GPS Raspberry Pi 3
[55]	3-axis acc. 3-axis ang. velocity speed	MPU-9250	71 Hz	car dashboard	Included in model input	Raspberry Pi 4 GPS
[54]	3-axis acc.	ADXL362 VK2828U7G5LF sensors	93 Hz	dashboard car wheel	n/a	GPS PIC18F26K22
[34]	3-axis acc. speed	Data Logger OBD II	24 Hz	inside vehicle	included in model input	OBD Data Logger
[63]	3-axis acc. ang velocity	smartphone accelerometer	100 Hz	fixed in vehicle	maximum 50 km/h	Two smart mobile devices
[52]	3-axis acc. ang. velocity	smartphone sensor	450 Hz	windscreen	different speed each trip	two smartphones
[39]	vertical acc.	Piezoelectric CT1100L	1280 Hz	tire suspension knuckle	30,40,50 km/h	MCC USB-231 notebook PC

bump, or curve/turn. The on-board devices used MPU6050 sensors, while the roadside devices were Raspberry Pi 3 units equipped with Spark, Sklearn, NumPy, Pandas and TensorFlow Python libraries. The results are then sent to a central database for storage.

An LSTM model was proposed by Bhosale et al. [72] for a cloud-based fusion system for road hazard detection using smartphone motion data. Road segments were classified into three categories: no hazard, road defect, and obstacles hazards.

Sabor et al. [73] implemented a 2D-CNN with Low-High Frequency Features Mixer block for a crowdsourcing system designed to detect asphalt bump, pothole or metal bump. The method used an an initial 2-way classifier to detect the presence of anomaly, followed by a 3-way classifier to identify the anomaly type.

Pandey et al. [55] suggested the potential of a 6G connected autonomous vehicle framework that fuses vibrational and images data using DL technique. The framework is based on federated learning, enabling distributed training of a DL model without accessing private user data.

Ramesh et al. [74] proposed a cloud-based collaborative fusion approach using Amazon Web Services (AWS). This method employed smartphone motion data for damaged/undamaged classification and a vision DL method to recognize the damage type. An LSTM model was trained with raw smartphone acceleration data. Cloud system gathered user prediction and performed a data clustering to aggregate results.

Xin et al. [64] introduced an interesting approach that combines acceleration and video data in a crowdsourcing solution for pothole detection. Features from a video slide of n frames were extracted with a 2D-CNN block and concatenated with raw acceleration data to produce a global feature used for prediction. A spatial density-based clustering algorithm was used for single user prediction fusion. This joint model significantly outperformed ML and LSTM classifiers in terms of accuracy.

Tab. 6 summarizes the crowdsourced studies based on vibrational signals. Similar observations can be made regarding data acquisition setups used by DL methods.

IV. VISION-BASED METHODS

In the following subsections, we considered three main aspects of the AI applied to vision data: object classification, image segmentation and object detection. Classification approaches aim to recognizing distress categories in vision data. Image segmentation techniques emphasize the small objects or details. Object detection focuses on identifying instances of semantic objects within a certain class.

A. CLASSIFICATION

Classification approaches using Deep Learning on vision data typically subdivide the collected images into smaller patches, which are then used to train the neural network and perform classification.

This is done for two main reasons: first, to avoid adding excessive computational burden to the method, which would result in much longer training time and impracticality for real-time implementation; and second, because classification is mainly used to detect localized distresses, as shown by Qureshi et al. [75].

One of the first available benchmarks for image classification was proposed by Eisenbach et al. [76], where the well-known *GAPs* dataset was collected with an S.T.I.E.R. mobile mapping system for large-scale pavement surveys. The distress categories considered–such as cracks, potholes, inlaid patches, applied patches, open joints and bleedings were labeled with bounding boxes around each distress in high resolution images captured by a CCD sensor (KAI-2093

TABLE 6. Resume of crowdsourced studies based on vibrational signals.

Ref	Distress type	Algorithm	Sensor data	Sensor details	Sampling frequency	System details
[66]	potholes	Logistic Reg, SVM, KNN, NB, DT,RF, EV	3-axis acc. ang velocity	MPU6050	10 Hz	Raspberry Pi 3 detection Predictions sent to server
[67]	Good Medium Bad	CNN	3-axis acc. speed	smartphone accelerometer	50 Hz	RoadCare app Real-time detection Clustering and ranking predictions on server
[70]	Pothole Bump patched road damaged road normal road	SVM CNN	3-axis acc.	smartphone accelerometer	16 Hz	Devices connected to LoRaWAN network
[72]	plain pothole speed bump curve/turn	ANN KNN	3-axis acc.	MPU6050	1000 Hz	V2I-Fog computing arch.
[56]	potholes	1D-CNN + 2D-CNN	3-axis acc. image	smartphone acc, camera	100 Hz	6G connected vehicles Federal Learning scheme
[65]	potholes	LSTM	3-axis acc, video	smartphone accelerometer dashcam	50 Hz	Joint acc+video model Classification and clustering on multiple local servers

by Kodak). The distresses boxes were extracted and resized to 64 × 64.

An extension of the *GAPs* dataset was proposed by Stricker et al. [77] increasing the number of images from 1969 to 2468 with a patch size of 100 × 100. A custom network called ASINVOSNet, with 3 convolution blocks, achieved a F_1-score of 91.74% on the test set.

Zhang et al. [78], proposed a patch-wise classification method to distinguish clean road, patch, potholes, linear and reticular cracks. Granular segmentation masks were of a full-scale resolution images demonstrated the the method's effectiveness. The proposed custom CNN outperformed ASINVOSNet in term of accuracy.

Chun et al. [79] used medium-sized, partially overlapped patches of 256 × 256 pixels from 1024 × 1024 images collected with a 3D Mobile Mapping System to distinguish several types of cracks and road markings. They proposed an iterative training technique for ResNet50, which adds incorrectly classified images to the training set at each iteration and evaluates model performance on a fixed, separated set, achieving a final accuracy of 94%.

The variability of road distresses dimensions can significantly limit classification performance. In this context, Eslami and Yun [80] proposed a multi-scale CNN which processes 50 × 50, 250 × 250, and 500 × 500 mm^2 images' Region of Interests, all resized to 50 × 50 pixels. The network comprises three convolution blocks and a final attention block that combines the features. The multi-scale model achieved the highest F_1-score of 92.0% on the *UCF-Pave 2017* dataset, which contains images with four types of distress anomalies and five types of non-distress anomalies.

1) THERMAL, INFRARED AND 3D LASER IMAGES

Since the performance of vision techniques based on RGB images can be significantly affected by weather conditions, camera noise, and illuminance variations, recent research in road damage detection has explored the application of thermal, infrared and laser 3D images, which are more robust to these issues.

The feasibility of using thermal images taken by a camera (FLIR ONE by Teledyne Flir) for pothole detection was first studied by Aparna et al. [81]. ResNet-18,34,50,102,150 models were trained and tested on a dataset made by 4500 images resized to 240 × 295 and augmented with data augmentation techniques. All the models reached an accuracy above 89%.

Chen et al. [82] presented an RGB, thermal, and Multi-band Dynamic Imaging (MSX) fusion approach with 500 RGB and 500 thermal images for each of the nine considered categories, including various types of transverse, longitudinal, and alligator cracks, potholes, joint or patches, manholes, shadows, road markings, and oil stains. EfficientNet B0-7 [83] performances were compared after training on single RGB, thermal and fused images datasets. The fused images approach outperformed RGB, thermal and MSX for all EfficientNet versions and categories, EfficientNet B4 achieving the highest average accuracy of 98.34%.

The same data collection setup and categories as [82] were used in Chen et al. [84], where an RGB-thermal concatenation was employed to train a CNN including custom hierarchical residual blocks,channel attention modules and spatial attention modules.

Liu et al. [85] proposed a visible-infrared images fusion framework combined with EfficientNet B0-B7 to classify three levels of severity for cracks and absence of crack. Fusion technique improved accuracy compared to the visible dataset for all state-of-art networks, with EfficientNet B3 achieving the highest accuracy of 94.14%.

Another approach involves using range images, which measure the spatial distance of the surface from a focal point. Zhou et al. [86] employed a 3D laser imaging system by AMES to collect RGB and range images of pavement to detect cracks. Three approaches were compared: Net-A utilized a heterogeneous dataset made by both intensity and range images, Net-B received fused intensity-range images as input, and Net-C had two distinct blocks for intensity and range images, later fusing their features. Net-B, with image fusion, achieved the best detection accuracy of 99.16%.

2) GENERATIVE ADVERSARIAL NETWORKS

Generative Adversarial Networks (GANs) are architectures used to generate synthetic data that closely resemble real data. They represent a rapidly expanding and innovating field within DL, applicable to both supervised and unsupervised learning.

GANs architectures consist of two main components: a generative block, which typically takes random noise as input and generates fake samples, and a discriminative block, which aims to distinguish between real samples and those produced by the generator.

Training a GAN involves solving a min-max optimization problem. The discriminator's weights are adjusted to maximize the probability of correctly identifying generated data as fake, while the generator's weights are adapted to maximize the probability of generated data being classified as real.

In this context GANs are utilized to address the issue of small datasets, offering an alternative to traditional data augmentation methods.

A Variational Autoencoder was used to extract input features for the generator in [87], extending a dataset of both phone and camera images for crack detection. A combined dataset of 3000 real crack images and 3000 DCGAN-generated images was produced, resulting in an improvement in crack detection accuracy on the original dataset and with Faster-RCNN from 68.25% to 90.32%.

Xu et al. [149] compared VGG-19 classification accuracy on a dataset before and after GAN augmentation. Images captured with high resolution camera were labeled as no crack, linear crack and reticulated crack after being subdivided in smaller patches. Accuracy improved from 80.15% to 91.61%.

A comparison between traditional data augmentation and expansion using GAN-generated data was performed by Yun et al. [89] to classify horizontal, vertical, and alligator cracks, pothole, and non-crack labeled images captured with a camera (AC7 by ORDRO) from a top down perspective. Traditional data augmentation techniques included flipping, rotation, cropping, adding Gaussian noise, and color jittering. A modified VGG16 network was used for classification, achieving a 93.27% accuracy on the augmented set with image processing, and a 97.01% accuracy on the augmented set with GAN.

3) PARTICIPATORY SENSING SYSTEMS

Several recent studies have focused on classification within participatory sensing solutions in crowdsourced systems.

Bibi et al. [90] proposed a framework with edge implementation of anomaly recognition among crack, pothole, bump, and no anomaly within a vehicular ad hoc network (VANET). In this system vehicles communicate with each other and with the infrastructure, transmitting information about road anomalies for safety purposes. ResNet18 and VGG-11 were trained on a dataset obtained by merging three existing databases of smartphone images, achieving an accuracy of 99.92%.

For real-time edge implementation of the DL model, Yue et al. [91] developed a custom network named *Mobile-crack* to classify road anomalies such as: cracks, sealed cracks, pavement markings, or matrices. Smaller patches were extracted from high resolution images, and an optimal patch size of 200 × 200 pixel was chosen to balance the trade-off between accuracy and inference time, measured on an Intel Core i7-6700 CPU. The model achieved an accuracy of 86.5% on test set was with an inference time of 47 ms.

Jana et al. [92] presented a benchmark for model performance comparison using a participatory sensing dataset of 1590 images captured by users' smartphones for pothole detection.

Patra et al. [93] introduced a mobile participatory sensing framework named *PotSpot* for pothole detection. Users collect images using a distributed application, that can either send the images along with GPS location to a Firebase server or run pothole detection on their devices and send the detection information to the server. The model was initially developed in Tensorflow, then converted to Tensorflow-Lite, and integrated into the Android application.

Table 7 summarizes the state-of-art of DL classification techniques on vision data.

B. SEMANTIC SEGMENTATION

Most of semantic segmentation literature for road condition assessment focuses on cracks rather than other types of damages, as cracks are well-suited for characterization with a binary mask where each pixel is classified as belonging to a crack or not. However, a few studies concentrated on other types of distresses, such as potholes.

In contrast to vibrational-based methods, a significant number of available datasets and benchmarks exists for segmentation methods. Table 8 summarizes the main publicly available datasets for road crack semantic segmentation. These datasets provide a large number of images, including high-resolution images, with the primary distress type being cracks. Preferred data collection setups generally involve high-resolution cameras with fixed angle capturing only pavement view, CCD sensors mounted on inspection vehicles for top-view images, and smartphone-based data collection setups.

Existing datasets have been extensively used to evaluate and compare the performances of DL segmentation models. This has led to the development of high-quality performing crack segmentation DL architectures from 2020 onwards.

TABLE 7. Resume of state-of-art of deep learning classification techniques on vision data.

Ref.	Distress type	Architecture	Dataset details	Input resolution	Data acquisition	Performance	Main novelty	Main disadvantage
[81]	crack crack seal patch pothole	Custom multi-scale CNN with Attention block	UCF-PAVE 2017 dataset 1215 pavement images 719 asphalt, 496 concrete	3 different scales resized to 50x50	RGB images Four different high-speed line-scanning cameras top down view	Average F1-score = 92.0% Best among other state-of-art models	Proposed a Multiscale model with Attention block	Higher computational cost & inference time than baseline multiscale models
[83]	Transverse cracks Longitudinal cracks Alligator cracks Joint or patches Potholes Manholes Shadows Road Markings Oil Stains	EfficientNet B0-B7	500 RGB, 500 thermal for each category Data augmentation with noise, thermal colorbar scaling 18945 augmented train/val/test: 60%/20%/20%	n/a	RGB images & Thermal images FLIR ONE camera	Accuracy from 96.91% and 99.52%	RGB + thermal fusion Comparison of EfficientNet versions	Does not use a fusion model for RGB and thermal images
[86]	no crack low severity crack medium severity crack high severity crack	MobileNet-V2 ResNet 34-152 DenseNet 121-161 EfficientNet B0-B7	2316 total images train/test: 80%/20%	224 × 224	RGB & infrared images Infrared camera FLUKE TiX58	EfficientNet B3 best: acc=93.28% RGB acc=86.55% infrared acc=94.14% fused	RGB & infrared fusion for crack severity classif.	Transfer learning improves performances only for visible images
[87]	cracks	Custom Net-A: same CNN with intensity or range Net-B: fused intensity & range Net-C: dual architecture	train/val/test: 18000/6000/6000 50%/50% crack/non-crack	256 × 256 patches	intensity & range images 3D laser imaging system by AMES	Net-B best: acc=99.6%	crack detecton with intensity & range images fusion	Focuses only on crack detection
[90]	horizontal crack vertical crack alligator crack pothole non-crack	Custom GAN gener. & discrim. Improved VGG16 for classif.	T-set: 5200 images G-set: 12000 images	64 × 64 patches	ORDRO AC7 camera top down	Improved VGG16 acc.: T-set=93.27% G-set=97.01%	Compares GAN augm. with image processing augm.	Proposes only a slight modification of VGG16 for classification
[91]	pothole bump crack no anomaly	ResNet18 VGG-11	Pothole Image Dataset Speed Hump/Bump Dataset Pothole and plain road images Merge + data augmentation	224 × 224	Web Scrapped	Accuracy: 99.92%	Edge-AI based framework for VANET	Poorly consistent dataset
[92]	crack sealed crack pavement marking pavement matrix	MobileCrack	train/val/test: 10,000/1000/400	100 × 100 200 × 200 400 × 400 patches	road inspection vehicle with camera	Optimal performance with 200 × 200 acc=86.5% Inference time=47 ms	Proposed a lightweight network MobileCrack	No comparison with other lightweight models
[94]	potholes	Custom CNN	train/test: 3264/160	64 × 64	users' mobile camera	Custom CNN acc=97.5%	Participatory sensing system Android App with integrated NN model Data collected by a Firebase cloud service	Focuses only on pothole detection

The following subsection provides a detailed review of the most recent high-performance DL architectures for crack segmentation, tested on available benchmarks.

Additionally, we explore recent innovations focused on specific types of problems or approaches. Examples include hybrid methods that combine segmentation with classification or detection to enhance performances, the application of pavement 3D images collected by laser or stereo vision systems, and alternative methods to address the imbalance between crack and non-crack pixels in images.

1) NOVEL DEEP LEARNING ARCHITECTURES FOR CRACK SEGMENTATION

One of the most significant breakthroughs in crack segmentation was the introduction of the Hierarchical Feature (HF) approach by Zou et al. [94]. HF outperformed classical segmentation architectures like UNet of Jenkins et al. [95] and Fully Connected Networks (FCN) of Long et al. [96].

HF approach extracts prediction masks from the multiple intermediate-scale decoder features, and creates a fused prediction through concatenation and convolution operations. Training uses a combined loss by summing contributions from the multiple prediction masks.

The application of Feature Pyramid with Hierarchical Boosting (FPHB) to a custom UNet on the *Crack500* dataset by Yang et al. [97], which was collected in the same study, resulted to a model significantly outperforming existing methods as FCN, CrackForest, holistically-nested edge detection (HED), richer convolutional features (RCF). In their work, the authors measured the segmentation quality with an indicator known as ODS, which is defined as the maximum F_1-score with respect to value of the threshold used to extract the binary mask from the predicted probabilities.

Han et al. [98] proposed a modification of UNet, which consisted of an upsample block nested into the main encoder-decoder structure to perform crack segmentation on *Crack500* dataset enlarged by self collected images. A 77.32% IoU was obtained in the test set by using 720 × 1280 image resolution.

Wang and Su [99] achieved a 4.53% IoU improvement with respect to the UNet version proposed in [100] for *Crack500* segmentation by using a DenseNet121 backbone as encoder combined with a decoder based on global attention upsample blocks.

In [101] Liu et al. proposed a network called DeepCrack, along with a dataset of the same name comprising 537 crack images of 544 × 384 resolution. They used the first 13 convolutional layers of VGG16 as a backbone, combined with HF approach and aggressive data augmentation, achieved an 86.5% F1-score on the DeepCrack test set.

Deeplabv3+ [102], by Chen et al., is one of the highest performing segmentation architectures due to the Atrous Spatial Pyramidal Pooling Module, which provides robustness against objects scale variations.

Qu et al. explored in [103] the application of a custom DeepLab with Feature Piramid (FP) approach and multi-scale feature fusion in the decoder, reaching a 67.5% F1-score on *Crack500*, nearly matching the performance obtained by Song et al. [104] with CrackSeg.

In [105] Fan et al. assessed the application of a multi-dilation convolution module in a customized UNet combined with FP approach, proposing the U-HDN network, which was trained and tested on *CFD* and *AigleRN* datasets.

A significant performances boost on *Crack500* was achieved by Qu et al. [106]. Gao et al. presented in [107] an encoder-decoder network with a Res2Net101 backbone

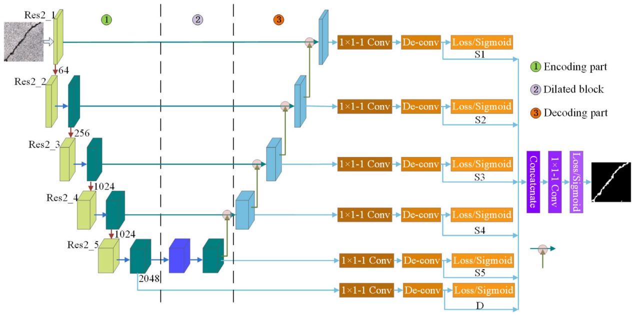

FIGURE 2. Architecture proposed in [106]. Copyright 1558-0016 © 2021 IEEE.

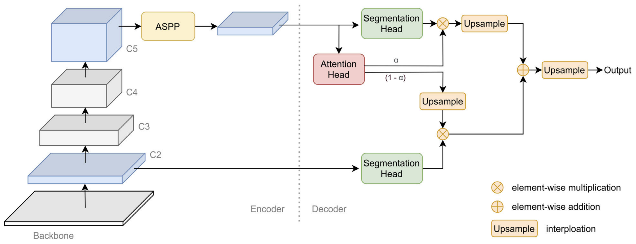

FIGURE 3. DMA-Net architecture proposed in [108]. Copyright 1558-0016 © 2022 IEEE.

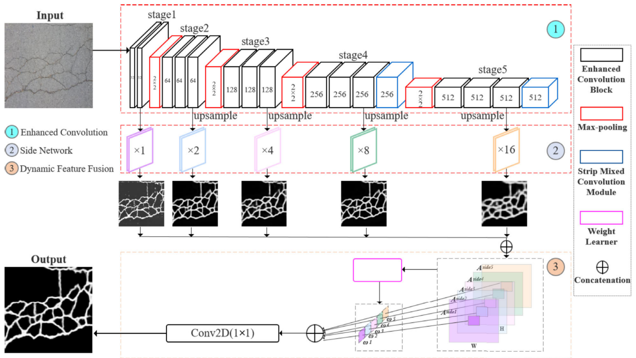

FIGURE 4. ECDFFNet architecture proposed in [110]. Copyright 1558-0016 © 2022 IEEE.

encoder and HF approach achieving 75.3% ODS on *Crack500* test set, which is 2.4% higher than the ODS obtained by the DeepCrack in the same work. Fig. 2 reports a scheme of this architecture.

The first architecture to slightly outperform DeepCrack performances is DMA-Net, proposed by Sun et al. [108] and reported in Fig. 3. This network utilized a Deeplabv3+ encoder with a pre-trained ResNet101 backbone on Imagenet [109], a double-scale attention fusion decoder and hierarchical feature approach.

An improvement of 3.1% ODS over DeepCrack was achieved by Zhou et al. [110] with a ECDFFNet model. This model featured an encoder composed of enhanced blocks with horizontal, vertical, and square parallel convolutions, strip blocks made up of 1D and 2D large convolutions, and HF approach. Additionally, a Dynamic Fusion Strategy was introduced to replace the traditional fixed weight fusion method, enabling the generation of a global prediction mask from the multiscale features. Fig. 4 reports a scheme of this architecture.

Al-Huda et al. [111] proposed a transfer learning approach based on the fusion of the features produced by a segmentation network and a class activation map (CAM) architecture. The segmentation model was based on a Xception encoder while the CAM network was custom. The proposed hybrid network called KTCAM-Net outperformed DMA-Net [108] on *Crack500* and *DeepCrack* datasets by more than 1% F1-score.

To the best of our knowledge, the model that achieved the highest performance boost over DeepCrack was MST-Net, presented by Yang et al. [112]. MST-Net employed a multi-scale encoder with ResNet block, a triple attention module for each scale containing channel, spatial, and pixel-wise attentions, and the classical HF. Performances comparison on *Crack500* and *DeepCrack* demonstrated significant improvement not only over *DeepCrack* but also over the most recent architectures, including ECDFFNet.

Table. 9 summarizes the most recent high-performing DL architectures for crack segmentation. A comparison between the performances of the different architectures can be partially performed when the same dataset is used.

2) OTHER RECENT NOVELTIES

Other recent researches focused on addressing various challenges such as developing real-time segmentation architectures, a challenging task causing a high computational burden, leveraging 3D images captured by laser or stereo systems, utilizing UAVs for monitoring systems, and combining segmentation with object detection to select distress regions for improving accuracy.

As stated by Paszke et al. [113], achieving at least 30 FPS is necessary for a DL segmentation model to be suitable for a real-time implementation. Wang et al. [114] explored the possibility of using the BiseNet architecture proposed by Yu et al. [115] for real-time processing on desktop GPU. An inference speed of 31.3 FPS was reached on 1024 × 512 images using NVIDIA GTX1080Ti GPU, and demonstrated a comparable segmentation performance on *Crack500* dataset as Deeplabv3+ and FC-DenseNet103 [116].

A custom CNN was designed by Pengfei et al. [117], featuring an encoder with asymmetric convolution enhanced blocks that include vertical, horizontal, and square parallel convolutions, and a 5 × 5 kernel convolution. In the decoding stage, a multi-scale feature fusion approach was used. This network proved feasible for real-time implementation on a NVIDIA Geforce RTX2080 TI GPU, achieving 33.61 FPS with 1280 × 1024 images. Although BiseNet showed better computational performances, the proposed model had superior segmentation accuracy.

Contrary to the aforementioned studies that utilized a desktop GPU, a low-computational cost architecture, feasible for mobile implementation was presented by Doğan at al. in [118], with an encoder based on MobilenetV2 bottlenecks, presented by Sandler et al. [119]. This architecture demonstrated superior segmentation performances compared to [114], additionally a lower computational cost than MobilenetV2 was proven.

Liu et al. [120] employed hybrid detection-segmentation approaches to overcome the problem of imbalance between crack and non-crack pixels or samples. In [121] Nguyen et al. trained a custom classification network to classify 96 × 96 patches from higher resolution images as containing cracks or not, followed by segmentation of the classified patches with another custom CNN. This combined method outperformed traditional one-step methods. Similarly, Qiaoning et al. [122] proposed an hybrid method to handle the high number of non-crack images collected during inspections by selecting images classified by a pre-trained VGG16 for segmentation with a UNet++.

3D laser imaging is another expanding research direction for crack segmentation. Zhang et al. [123] addressed this issue. They proposed the CrackNet network, trained on a dataset collected with the PaveVision3D system using 4096 × 2048 3D images, achieving an 88.86% F1-score in the test set. Fei et al. [124] developed an evolution of this model named CrackNet-V, with a significantly fewer parameters than [123] and comparable segmentation performance on a private 3D images dataset. The problem of imbalance between crack and non-crack pixels in both 3D and color images was intelligently addressed by Tang et al. [125], where the label binary mask was projected into a lower dimensional feature-space using an autoencoder initially trained to reconstruct the input label. A segmentation network was trained to estimate the low dimensional features produced by the autoencoder from the binary labels, achieving a 97.82% F1-score in the test set, thus significantly outperforming CrackNet-V and UNet with ResNet34 encoder.

UAV images are becoming an attractive alternative for road condition monitoring to the inspection with equipped vehicles, offering feasibility with solutions for real-time monitoring. A highway road crack segmentation dataset of UAV images was published by Hong et al. [126], which contains 1157 images captured with a 5 cm resolution at the height of 200 m. This dataset was used to evaluate the performance of a custom light UNet architecture with channel and spatial attention modules after the encoder, which was trained on an existing dataset [127] originally used for object detection. UNet-CBAM achieved a 98.87% F1-score on the evaluation set.

Chao et al. [128] implemented a system for road 3D reconstruction using stereo vision with a high resolution camera mounted on a vehicle. The aim of the work was to perform a high-precision volume estimation of potholes and cracks. The 3D reconstruction involved estimating camera coordinates with feature tracking between successive frames, followed by a triangulation procedure to generate the 3D point cloud. Semantic segmentation was performed with a UNet based on depthwise convolutions to select the pothole or damages regions from the orthoimages and depth images generated with the aid of 3D construction.

Instead of performing a pixel-level segmentation, Tang et al. [129] developed an iterative patch-level segmentation technique to classify 300 × 300 patches from full resolution images as containing several types of cracks, raveling, repair or being normal road. A classification network was iteratively trained with the output produced in the previous iteration as target labels by starting from a constant patch label along the full-scale image. The proposed method outperformed EfficientNet in terms of Precision and Recall.

In order to perform crack segmentation, Mei et al. [130] proposed a GAN which employed a conditional generator along with the Wasserstein distance instead of the classical Kullback-Leiber divergence. The network was trained on a dataset made up of 600 images collected with a GoPro Hero mounted at the back of a vehicle. High resolution collected images were split into 256×256 patches. The method reached superior segmentation accuracy than UNet, ResNet152-FCN and VGG19-FCN.

Segmentation of road images is generally not performed on very high resolution images due to the computational and memory constraints. Reduced image resolution typically limits the segmentation accuracy of DL models. To address this, Shim et al. [131] proposed an image resolution enhancement technique based on GAN, to increase road damage images resolution from 512 × 512 to 1024 × 1024. A segmentation network, also based on a GAN, was then used to produce the map from the generated high resolution images. This approach was tested with several generators and discriminators for the two GANs, resulting in a significant improvement over traditional low resolution supervised segmentation methods.

C. OBJECT DETECTION

As opposed to semantic segmentation methods which mainly concentrate on cracks, several categories of road damages are typically considered. Unlike semantic segmentation methods that primarily focus on cracks, object detection DL architectures handle multiple categories of road damages by performing both localization and classification tasks simultaneously.

Additionally, while segmentation methods usually work on datasets collected from inspection vehicles or professional cameras with top-down view images, object detection is suitable for distress detection in wide view images captured by user cameras and real-time monitoring systems based on UAVs.

One of the most well known state-of-art object detection benchmarks for road pavement anomalies was published by Maeda et al. [137]. A total of 9053 600 × 600 images were collected in Japan with a LG Nexus smartphone mounted on car dashboard, including 15435 abnormalities instances belonging to eight different categories.

TABLE 8. Resume of the main available datasets for semantic segmentation and classification techniques.

Ref.	Dataset name	Distress	Method	Number of images	Acquisition device	Domain	Original resolution
[133]	CrackForest dataset (CFD)	cracks	semantic segmentation	118	smartphone camera	RGB	480 × 320
[134]	AELLT	cracks	semantic segmentation	269	Aigle RN (62), ESAR (30), LCMS (65), LRIS (89) and Tempest (23)	RGB laser	varying
[135]	YangCD	cracks	semantic segmentation	800	n/a	RGB	from 72 to 300 dpi
[102]	DeepCrack	cracks	semantic segmentation	537	n/a	RGB	544 × 384
[102]	CrackTree260	cracks	semantic segmentation	260	area-array camera	gray-scale	512 × 512
[77]	GAPs	crack pothole inlaid patches applied patches applied patches open joint bleeding	detection	1969 images each image has a 64 × 64 bounding box with a distress	Kodak KAI-2093 CCD scanner	gray-level	1920 × 1080
[78]	GAPsv2	Normal Applied patch crack inlaid patch open joint pothole	detection	Extended GAPs 2 468 images	Kodak KAI-2093 CCD scanner	gray-level	1920 × 1080
[98]	GAPs384	cracks	semantic segmentation	384 crack images from GAPs	Kodak KAI-2093 CCD scanner	RGB	1920 × 1080
[98]	Crack500	cracks	semantic segmentation	500 collected images each cropped in 16 regions 3368 total images	smartphone camera	RGB	2000 × 1500
[131]	EdmCrack600	cracks	semantic segmentation	600	GoPro Hero 7 Black	RGB	1920 × 1080
[130]	CQU-BPDD	transverse crack, massive crack alligator crack, crack pouring longitudinal crack, raveling repair, normal	patch-wise classification	60059	in-vehicle cameras of a professional pavement inspection vehicle	gray-level	1200 × 900
[136]	-	cracks	semantic segmentation	11300 images Own dataset merged with Crack500, Gaps, CFD, AELLT, CrackTree260	n/a	RGB	256 × 256
[137]	CrackSC	cracks	semantic segmentation	197 images	iPhone 8	RGB	320 × 480
[127]	-	cracks in highway pavements	semantic segmentation	1157	UAV	RGB	5 cm resolution

Maeda et al. [138] proposed RDD2019 which is an augmentation of RDD2018 made with a GAN approach. An additional category was added such as ‘‘Utility Hole’’ and the number of images was increased up to 13159 containing 30989 instances.

Arya et al. [139] recategorized RDD2019 by merging classes and obtaining only four different categories such as longitudinal cracks, transverse cracks, alligator cracks and potholes. The dataset was also enlarged with images collected with a dedicated smartphone app in India and Czech Republic. A total of 26336 images dataset was obtained, including 31000 instances.

Furthermore Arya et al. [140] proposed the most recent version of this dataset called RDD2022 by expanding RDD2020 up to 47420 images and 55000 instances by adding images collected in Norway with high resolution camera, from Google street view in USA and from China with drones and smartphones mounted on motorbikes.

Traditional deep learning architectures for object detection, such as Single Shot Multibox Detector(SSD) [141], Faster R-CNN [142], and YOLOv3 [143], have been extensively tested on available benchmarks.

Cao et al. [144] established a benchmark employing the RDD2018 [137] dataset to compare the performances of SSD Mobilenet V1-2, Faster R-CNN Inception V2, Inception ResNet V2, and ResNet50-101. Despite being computationally efficient, SSD Mobilenet V1-2 achieved significantly lower detection accuracy than Faster R-CNN, which reached a 54.75% mean Average Precision (mAP) with Inception ResNet V2 backbone.

A better trade-off between computational complexity and detection accuracy can be achieved by YOLO-series, as shown by Du et al. [145], where cracks, patched cracks, potholes, patched potholes, net, patched net, and manholes were considered as distress classes. YOLOv3 obtained slightly lower average precision (AP) than Faster R-CNN but had a lower inference time than SSD when tested on NVIDIA Geforce Titan X.

Zhu et al. [146] presented the largest available UAV collected dataset for detection, comprising 3151 images of size 512 × 512, considering several types of cracks, sealed cracks, and repaired patches, potholes as categories. YOLOv3 achieved a 56.6% mAP, outperforming Faster R-CNN, which did not reach 50% mAP, and even its evolution YOLOv4.

TABLE 9. Resume of state-of-art novel DL architectures for crack segmentation on available benchmarks.

Ref.	Architecture details	Dataset	Input resolution	Training method	Performance	State-of-art comparison
[99]	CrackWNet: UNet with round trip skip-level upsample block	Crack500 + self collected images	720 × 1280	scratch	IoU = 77.32% Rec = 57.37%	IoU > 7% higher than FCN, UNet
[98]	Custom UNet with Hierarchical Feature approach	Crack500 GAPs384 CrackTree200 CFD AigleRN	Depends on dataset	scratch	Crack500: ODS = 60.4% GAPs384: ODS = 22.0% CrackTree200: ODS = 51.7% CFD: ODS = 68.3% AigleRN: ODS = 49.2%	Clearly outperforms FCN, HED, RCF and CrackForest
[106]	U-HDN: UNet + multi-dilation convolution module + Hierarchical Feature approach	CFD AigleRN	320 × 480	scratch	CFD: ODS = 93.5% AigleRN: ODS = 92.7%	Higher ODS than UNet, FPHBN[98]
[104]	DeepLab + multiscale feature fusion + Hierarchical Feature approach	Crack500 DeepCrack CrackDataset (eval. only)	n/a	scratch	Crack500: F1 = 67.5% DeepCrack: F1 = 85.8% CrackDataset: F1 = 61.5%	Same perf. as CrackSeg on Crack500 Same perf. as DC[102] on DeepCrack
[107]	Encoder-decoder: Res2Net encoder blocks + Hierarchical Feature approach	Crack500 DeepCrack (eval. only) CFD (eval. only) CrackTree260 (eval. only)	n/a	scratch	Crack500: ODS = 75.3% DeepCrack: ODS = 83.4% CFD: ODS = 72.6% CrackTree260: ODS = 32.5%	ODS 2.4% higher than DeepCrack[102]
[100]	DenseNet-121 backbone Global Attention Upsample blocks	Training: Crack500 + DeepCrack + GAPs384	512 × 512	transfer	Crack500: IoU = 62.35%	IoU 4.53% higher than UNet in [101]
[109]	DMA-Net: Deeplabv3-plus (Dilated ResNet101 backbone) + double scale attention decoder	Crack500 DeepCrack FMA dataset	n/a	transfer	Crack500: F1 = 74.4% DeepCrack: F1 = 87.0% FMA: F1 = 75%	F1 14% higher than FPHB on Crack500. F1 0.5% higher than DC on DeepCrack
[111]	ECDFFNet: Enhanced encoder blocks horiz,vert, square conv. + strip mixed conv blocks Hierachical Feature approach	Crack500 DeepCrack CFD	448 × 448 480 × 480 480 × 320	scratch	Crack500: ODS= 78.8% DeepCrack: ODS= 87.2% CFD: ODS= 86.3%	ODS 2.6% higher than DC [95] on Crack500 ODS 3.1% higher than DC [102] on DeepCrack
[112]	KTCAM-Net: Encoder features fused with activation maps produced by Class Activation Mapping. Xception backbone	Crack500 DeepCrack CFD CrackSC	224 × 224	scratch	Crack500: F1 = 75.4% DeepCrack: F1 = 88.6% CFD: F1 = 96.0% CrackSC: F1 = 92.1%	F1 1% higher than DMA-Net on Crack500 F1 1.6% higher than DMA-Net on DeepCrack
[113]	MST-Net: Multiscale inputs + ResNet blocks + Additive attention blocks + Triple attention blocks + Hierarchical Feature approach	DeepCrack YCD CFD	512 × 512 512 × 512 320 × 480	scratch	DeepCrack: mIoU=91.1% CFD: mIoU=78.5% YCD: mIoU=78.7%	mIoU 2.7% higher than ECDFFNet on DeepCrack mIoU 5.1% higher than DC [102] on DeepCrack

Another advantage of YOLO series over Faster R-CNN is that YOLOv3,v4, and v5 are based on COCO pre-trained backbones such as DarkNet53 first proposed by Redmon et al. [143] and CSPDarkNet53 proposed by Wang et al. [147], which are easily scalable by reducing the number of residual convolutions in each block.

Notable improvements for Faster R-CNN have been made by Ahmed in [148] and Xu et al. [149], where a Mask R-CNN [150] was trained and tested on a small dataset of 148 smartphone images to detect and segment cracks. Mask R-CNN extends Faster R-CNN by adding a parallel branch for segmentation in addition to detection.

Tab. 10 lists object detection works with available data, showing by a large amount of high resolution images.

1) EVOLUTIONS ON YOLO-BASED ARCHITECTURES

The evolution of YOLO architecture in recent years has led to highly performing object detection solutions with reduced computational complexity and inference time compared to previous versions.

Wang et al. [156] chose ResNet101 as backbone for YOLOv3, replacing the classical DarkNet53. Combined with an aggressive data augmentation, this approach effectively detected and distinguished potholes filled with water from those without water in top-down images captured from a mobile mapping system with a high resolution camera. Mean average precision reached 89.3%, more than 13% higher than the baseline YOLOv3.

YingChao et al. [157] proposed a modified version of YOLOv3, incorporating in its neck several multi-level attention blocks. This model was trained on the *AUPD* database achieving over 7% mAP improvement compared to the baseline YOLOv3.

Some works have applied YOLOv3 to analyze GPR, B-scan, C-scan, and D-scan images captured by multichannel ground-coupled antenna arrays to detect concealed crack in pavement, rather than the usual RGB images. In [158], Zhen et al. proposed a version of YOLOv3 with four output features for concealed crack detection from single B-scan and C-scan images or their spatial concatenation. This method showed very promising results, outperforming Faster-RCNN and YOLOv3-ResNet50 which was used for concealed crack detection from GPR images by Liu et al. [159].

A significant breakthrough was achieved by YOLOv5 [160], which, despite lacking an official reference paper, has quickly become a standardized architecture within the *Ultralytics* library.

Habeeb et al. [161] improved YOLOv5's performances for pothole detection by using a super resolution image enhancement based on a GAN with a relativistic discriminator.

Fang at al. [162] proposed a lightweight variant of YOLOv5-s called YOLO-LRRDD, whose architecture is

TABLE 10. Resume of available datasets for object detection.

Ref.	Dataset name	Distress	Number of images	Other details	Acquisition	Original resolution	Place
[138]	RDD2018	D00: linear crack, longitudinal, wheel mark part D01: linear crack, longitudinal, construction joint part D10: linear crack, lateral, equal interval D11: linear crack, lateral, construction joint part D20: alligator crack D40: rutting, bump, pothole, separation D43: cross walk blur D44: white line blur	9053 images 15435 instances	First benchmark for this competition	LG Nexus 5X on car dashboard	600 × 600	Japan
[139]	RDD2019	RDD2018 + D50: utility hole	13135 images 30989 instances	RDD2018 augmented with GAN & re-annotation	LG Nexus 5X on car dashboard	600 × 600	Japan
[140]	RDD2020	D00: longitudinal Cracks D10: transverse Cracks D20: alligator cracks D40: potholes	26336 images 31000 instances	RDD2019 augmented with images from India, Czech Republic & recategorized	smartphone app	720 × 720 India 600 × 600 Japan, Czech Republic	Japan India Czech Republic
[141]	RDD2022	same as RDD2020	47420 images 55000 instances	RDD2020 augmented with multi-source images from Norway, USA and China	RDD2020: smartphones Norway: high resolution camera USA: Google street view China: drones & smartphones on motorbikes	720 × 720 600 × 600 512 × 512 3650 × 2044	Japan India Czech Republic USA, China, Norway
[152]	Pothole Image Dataset	potholes	665 images 8000 potholes	includes effects of shadows, moving vehicles illumination variations	Online sources	720 × 720	-
[153]	-	potholes	5676 images	Loc. & classif. with two separate networks 352 × 224 for localization ROIs cropped from 1170 × 1120 images	Camera on car dashboard Near road area cropped and resized to 1170 × 1120	3680 × 2760	South Africa
[147]	UAPD	transverse crack longitudinal crack alligator crack oblique crack repair pothole	3151 images	Largest UAV available dataset	UAV DJI M600 Pro 35.9 mm × 24.0 mm CMOS sensor 300 high resolution images cropped into 512 × 512 regions	7952 × 5304	China
[154]	PANGEA dataset	pothole crack	565 images	Multi-agent data collection system	UAV: DJI Mavic Air 2 4K digital camera	3840 × 2160	China
[155]	-	cracks sealed cracks	10400 images	Redundant annotation Semi-automatic annotation	road measurement vehicle CCD sensor 1mm/pixel top-down view images cropped to 600 × 600	3024 × 1889	Mongolia
[156]	-	Same as RDD2022	2893 images	RDD2022 China drone images + Spain collected drone images	DJI Air 2S	3840 × 2160	China Spain

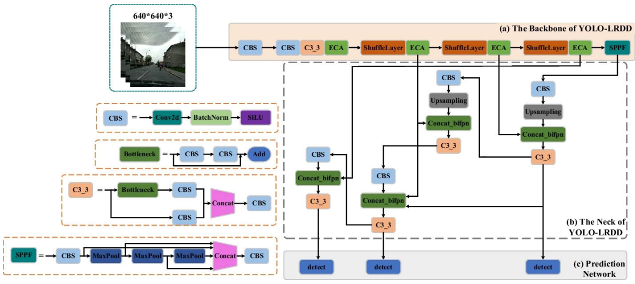

FIGURE 5. YOLO-LRDD architecture proposed in [162].

represented in Fig 5. The majority of bottleneck blocks of YOLOv5 backbone were replaced with ShuffleNet blocks. YOLO-LRDD was trained and tested on a dataset with 13780 images, selected from RDD2020 and supplemented with images collected in China. YOLO-LRDD was revealed to be computationally lighter than YOLOv5-s and achieved a slightly superior detection performances with a 57.6% mAP.

Ren et al. [163] proposed an enhanced architecture for the neck and head of YOLOv5, replacing the backbone with an additional feature scale output, a Generalized Feature Pyramid Network (FPN) neck, and a decoupled head similar to YOLOX developed by Ge et al. [164]. Trained on Google Street View images to detect various types of crack, patch types, potholes, and other distresses, this model outperformed most newer architectures as YOLOR and YOLOv7 proposed by Wang et al. [165] and Wang et al. [166] respectively.

Roy and Bhaduri [167], to the best of our knowledge, achieved the highest detection accuracy reported in literature on RDD2018 dataset with DenseSPH-YOLOv5, an evolution of TPH-YOLOv5 proposed by Zhu et al. [168]. TPH-YOLOv5 [168], incorporating Swin Transformer blocks and Channel + Spatial attention modules in the neck, was initially designed for detecting different vehicles and pedestrians from drone images. In comparison, DenseSPH-YOLOv5 [167] share a similar neck with TPH but additionally integrates single DenseNet blocks between bottleneck stages in the backbone. This model achieved an mAP of 85.2% for RDD2018, significantly surpassing TPH-YOLOv5 and baseline YOLOv5.

2) REAL-TIME MONITORING SYSTEMS

In the context of object detection, developing scalable models or lightweight architecture is generally more feasible than in pixel wise image analysis. Consequently, several studies were

TABLE 11. Resume of state-of-art deep learning object detection techniques.

Ref.	Distress	Architecture	Dataset details	Input resolution	Data acquisition	Performance	Main novelty	Main disadvantage
[145]	Type I: transverse cracks Type II: longitudinal cracks, poor construction joint Type III: alligator cracks Type IV: Other damage	SSD MobilnetV1-2 SSDLite MobilenetV2 SSD Inception V2 Faster RCNN ResNet50-101 Faster RCNN Inception V2 Faster RCNN Inception-ResNet V2	RDD2018 augmented & recategorized in 4 classes 9493 images	600 × 600	RDD2018	Best performance by Faster R-CNN- -Inception-Resnet-V2: mAP = 54.75%	Benchmark which compares SSD and Faster R-CNN with several backbones on RDD2018	High quality performances are only achieved by Faster R-CNN with heavy backbones
[149]	potholes	Modified VGG-Faster R-CNN YOLOv5-l,m,s ResNet101	Images from Roboflow, MakeML + additional 665 total	varying	smartphone on car windshield	mAP YOLOv5s: 58.9% Faster R-CNN MVGG16: 45.4% Faster R-CNN ResNet50: 64.12%	Proposed Faster R-CNN based on modified VGG16 for pothole detection.	YOLOv5s reaches higher mAP and lower inference time on desktop GPU
[150]	cracks	Mask R-CNN Faster R-CNN YOLOv3	148 images train/test: 90/10 %	800 × 500	smartphone	Mask R-CNN outperforms YOLOv3. Detection performances more sensitive than Faster R-CNN	Compares Mask-RCNN to to YOLOv3 and Faster-RCNN on a small crack dataset.	Detection performances tested on a very small dataset
[152]	potholes	YOLOv4 Tiny YOLOv4 YOLOv5-s	665 pothole images Train/val/test: 70/20/10 %	416 × 416	online sources	mAP: YOLOv4: 77.7 % Tiny YOLOv4: 78.7% YOLOv5-s: 74.8%	Proposes Pothole Image Detection dataset. Compares YOLO performances	Low accuracy for potholes located at long distances
[170]	potholes	YOLOV1-v5 Tiny YOLOv4 SSD MobilenetV2	PID dataset	Tiny YOLOv4: 416 × 416 YOLOv5: 640 × 640	online sources	mAP, FPS Tiny YOLOv4: 80.04%, 31.4 YOLOv5: 95%, 18.5 SSD MobilenetV2 poor performances	Model implementation on a Raspberry Pi OAK-D with OpenVino framework for real-time pothole detection on edge device	Does not propose a novel architecture for real-time detection
[154]	pothole crack	Tiny YOLOv4	565 images	1200 × 900	4K digital camera on a DJI Mavic Air 2 quadcopter	95.70% accuracy 5.52 ms image latency on Nvidia Jetson GPU	Proposed a multi-agent monitoring system based on UAV	Different crack detection techniques did not reach a precision higher than 47 %
[171]	pothole crack yellow lane	Tiny YOLOv3 improved	10000 images	416 × 416	UAV	mAP Tiny YOLOv3 improv: 94% Tiny YOLOv3: 89%	Real-time monitoring system with UAV Images captured by UAV, sent to Jetson TX2 with Wifi.	UAV navigation might be affected by detector performance
[172]	cracks	GAN YOLO-Median Flow Tiny model from YOLO-MF	1595 images 1735 cracks Augmented with 1500 crack images generated by GAN	512 × 512	smartphone	Tiny model acc=98.1%, 29 FPS	Real-time UAV detection system with Median Flow algorithm for crack counting	No comparison with other lightweight architectures
[157]	pothole with water pothole without water	YOLOv3-ResNet101	800 pothole images 1200 images without potholes Train/val/test: 1200/400/400	512 × 512	Mobile Mapping System High definition camera Top-down view	mAP Proposed model: 89.3% YOLOv3-base: 76.0%	Proposed a YOLO-ResNet101 for pothole detection from top-down images	Architecture novelty is only a modification of YOLO anchor sizes
[158]	UAPD categories	YOLOv3 with Multi Level Attention blocks in YOLO neck	UAPD	416 × 416	UAV DJI M600 Pro	mAP YOLOv3-MLAB: 68.75% YOLOv3-base: 61.09%	Proposed a modified YOLOv3 with different types of attention blocks in YOLO neck	Smaller mAP improvement obtained for RDD2020
[159]	concealed cracks	YOLOv3-FDL with multi-scale fusion neck	GPR images B,C-scan, concatenated B,C scan Train/val/test: 2784/928/928	416 × 416	3D radar detection system Multichannel ground-coupled antenna arrays	mAP YOLOv3-FDL DCN: 87.8% YOLOv3-ResNet50: 83.0%	Detection of concealed cracks from B,C-scan GPR images with a 4-feature output YOLOv3 with modified neck	Limited number of images. No data augmentation
[163]	RDD2020 categories	YOLO-LRDD Modified YOLOv5: ShuffleNet layers and ECA blocks backbone	Chinese extension of RDD2020 13780 images	640 × 640	n/a	mAP = 57.6% Accuracy comparable with YOLOv5s-m Lower computational cost than YOLOv5s	Proposed a YOLOv5 with ShuffleNet-ECA backbone and BiFPN neck	Several network undebugged parameters which may affect detection performance
[164]	longitudinal crack transverse crack alligator crack pothole manhole cover longitudinal patch transverse patch	YOLOv5 Improved: Additional output in the backbone Generalized FPN in the neck Decoupled head as YOLOx	2900 images Train/val/test: 1740/580/580	1024 × 1024	Street view Baidu maps	AP=79.8% F1=75% Outperforms YOLOR and YOLOv7	Proposed a modified YOLOv5 with improvements in backbone, neck and head for damage detection from Google street view images	Higher computational cost than YOLOv5s baseline
[168]	RDD2018 categories	DenseSPH-YOLOv5: YOLOv5 with DenseNet blocks in backbone. Swin Transformer and CBAM modules in the neck	RDD2018	600 × 600	LG Nexus 5X on car dashboard	mAP YOLOv5: 71.13% TPH-YOLOv5: 77.62% DenseSPH-YOLOv5: 85.25% Outperforms every other work on RDD2018	Proposed an extension of TPH-YOLOv5 [169] with DenseNet blocks in the backbone	Inference time higher than YOLOv5 on desktop NVIDIA GPU

focused on implementing real-time monitoring systems for road infrastructure.

For instance, Asad et al. [169] analyzed the feasibility of YOLOv5, Tiny YOLOv4 and SSD Mobilenet V2 for real-time pothole detection on the Pothole Image Dataset. Detection was performed with an OAD-K camera equipped with a VPU, a processor optimized for accelerating AI algorithms. Tiny YOLOv4 Darknet was the only model to achieve adequate real-time detection speed on Myriam VPU hardware, reaching 31.76 FPS together with a 80% mAP. Although YOLOv5 achieved 95% mAP, it reached 18 FPS, insufficient for real-time performance. SSD Mobilenet V2 did not achieve sufficient accuracy.

Silva et al. [153] developed a multi-agent monitoring system based on UAVs and the PANGEA platform, utilizing a distributed data collection approach. The system, which detected potholes and cracks, employed Tiny YOLOv4 as a detector, achieving 95% accuracy with a 5.52 ms inference time on a NVIDIA Jetson device. The images were published as a an available benchmark.

Hassan et al. [170] also used UAVs for potholes, cracks and yellow lanes detection. Images were sent via Wifi to a Jetson TX2, which ran an improved Tiny YOLOv3 Darknet detector along with a control algorithm for the drone's position based on the detected yellow lanes. Images were then transmitted to a server with Wifi or 5G technology if pothole or a crack was detected.

Jin et al. [172] proposed a modified version of YOLOv5 for real-time detection of highway road cracks. A 1560 images dataset was collected with a DJI Mavic3 UAV in realistic situations where the crack occupies only few pixels in the image. The proposed modified YOLOv5 included Swin-Transformer blocks in the backbone and a bidirectional feature pyramid network (BIFPN) in the neck. A 90% detection accuracy was obtained along with 43.5 FPS detection speed.

Ma et al. [171] integrated a Median Flow algorithm into a tiny YOLO detector for a real-time crack counting UAV solution. The drone, equipped with a NVIDIA Jetson Nano, executed the Tiny YOLOv3 model accelerated by the TensorRT library, achieving 98.1% accuracy and 29 FPS in a real-time system test. Fig. 6 reports the structure of the UAV monitoring system.

Tab. 11 summarizes the state-of-art of object detection techniques using DL. The evolution of the performances of

TABLE 12. Resume of the main improvements for road condition monitoring with AI algorithms for different data types.

Data type	Algorithm novelties	System novelties
Acoustic	• 2D-CNN with sound spectrogram input image	• Microphone placed inside tyre cavity • Few explored in literature
Vibrational	• DL methods have overcome Traditional ML • More advanced CNN & LSTM architectures. CNN based on 2D images extracted with FFT or wavelet transforms • Data fusion of acceleration, rotation and speed to assess speed dependency problem.	• Larger number of classified distresses • Crowdsourcing solutions.
Vision	• Improvement of crack pixel-wise segmentation performances on existing benchmarks with novel architectures • Improvement of detection performances with novel DL architectures based on YOLO-series	• Distress detection solutions with thermal, 3D laser & GPR images • Real-time monitoring systems with UAVs

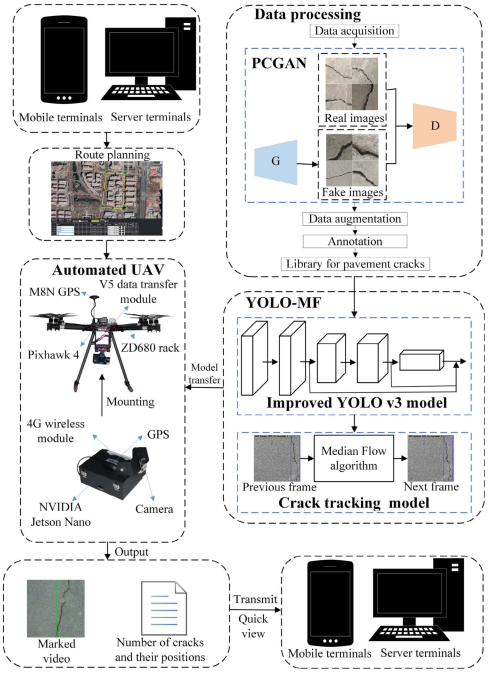

FIGURE 6. Structure of the UAV monitoring system proposed in [171]. Copyright 1558-0016 © 2022 IEEE.

the YOLO architecture is highlighted by the number of recent researches using this architecture.

V. DISCUSSION AND FUTURE WORK

A summary of the most recent advancements in road condition monitoring using AI algorithms, as explored in this paper, is provided in Tab. 12.

Although some studies have achieved promising results, acoustic data-based monitoring remains relatively underexplored in the literature.

For vibrational data-based methods, feature extraction in the frequency and wavelet domain has leveraged the performance advantages of 2D-CNN architectures. In contrast 1D-CNN and RNN architectures could benefit from the fusion of different data types such as acceleration, angular velocity and speed. Additionally, several crowdsourced approaches have been proposed, where data collected by the users is uploaded to a central server or database to train or fine-tune road anomaly classification algorithm.

Vision data-based methods have seen significant improvements in distress classification accuracy due to the constant evolution of DL architectures. For crack pixel-wise segmentation, the high number of available datasets has enabled extensive performance comparisons of proposed models against the state-of-the-art. Novel architectures based on feature pyramidal pooling, attention modules, and innovative convolution structures has led to a visible improvement in segmentation accuracy.

The performances of distress recognition with object detection have also seen notable accuracy improvements, particularly with YOLOv5-based evolutions, which are expanding very rapidly and they are being applied for road condition monitoring more than other standardized architectures such as Faster-RCNN due to their computational scalability. Some studies have successfully explored the feasibility of using different sensor data for detecting pothole, crack, or other distress from RGB images, such as 3D range images from laser, stereo vision data, and GPR images for concealed cracks, along with data fusion approaches. Real-time detection has been primarily explored in UAV monitoring systems, where the drones have hardware with low computational capacity.

The following aspects and challenges remain open or unexplored in the considered research field.

- **Standardization**: There is a lack of of standardization in data collection setups and preprocessing, especially for vibrational data methods. Different works use varying sensors, data acquisition setup, labeling technique, and types of distresses, making results from different studies non-comparable.
- **Few existing available datasets**: The scarcity of publicly available datasets for vibrational data hinders standardization and performance comparison in this context.
- **Data fusion**: The fusion of acceleration data with images or video has been minimally explored and could offer significant performances benefits.

- **Lightweight models**: Lightweight models require lower storage space, shorter training times, and could potentially be used for real-time estimation. The scalability of YOLO detectors offers a promising solution. Additionally, compression techniques based on pruning, quantization, and transfer learning could be highly beneficial.
- **Crowdsourcing systems**: Crowdsourcing systems represent an expanding research field; however, their potentialities for infrastructure monitoring needs further exploration.
- **Distress detailed characterization**: The application of AI techniques to 3D laser, GPR, and LiDAR data could offer solutions for detailed distress shape, dimension, and depth estimation, which are critical parameters for assessing the global infrastructure condition for authorities.

VI. CONCLUSION

In this article, we provide an extensive review of the most recent advances in RCM using AI techniques.

In recent years, the field of AI applied to road condition monitoring has seen a significant increase in published research articles. In this study, we identified key publications from the Scopus and Web of Science (WoS) databases, using the keywords reported in Tab. 1. Notably, only six of the selected papers utilized acoustic data, while 61 papers employed AI techniques on vibrational data, with 51 of these published between 2020 and 2024. The majority of studies, however, focused on image-based approaches for road condition monitoring, comprising 98 papers, 75 of which were published between 2020 and 2024.

Although acoustic data-base approaches have shown some promising results, they remain underexplored.

Vibrational methods have seen significant improvement thanks to the advent of DL architectures, which are replacing traditional ML techniques. These advancements include also new feature extraction methods and the the integration of angular velocity and speed with acceleration data. Additionally, several crowdsourcing studies using participatory sensing were proposed.

However, vision techniques have seen the most significant improvement in detecting road pavement damages, driven by advancements in DL for pixel level crack segmentation and image object detection.

The feasibility of several data types as an alternative or a fusion with RGB images has been explored such as thermal images, 3D range images from laser, stereo vision data and GPR images to detect various types of damages.

Real-time monitoring systems are also expanding, particularly those based on UAVs, where low computational cost DL models are implemented to meet real-time constraint.

REFERENCES

[1] Behalf Eur. Community, (2024). *Status of Progress on Connected, Cooperative and Automated Mobility in Europe*. [Online]. Available: https://research-and-innovation.ec.europa.eu/document/download/1720a5ef-01bf-498e-85f5-c61bb3a7bc31_en?filename=swd_2024_92.pdf

[2] RPB HEALTEC. (2024). *Road Pavements & Bridge Deck Health Monitoring/early Warning Using Advanced Inspection Technologies*. [Online]. Available: https://cordis.europa.eu/project/id/606645/en

[3] (2020). *PAV-DT User Manual V1.0*. [Online]. Available: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5d75d3d1c&appId=PPGM

[4] PAVE-SCAN. *PAVEment SCANning With EGNSS Technology for Accurate Assessment*. Accessed: Oct. 2024. [Online]. Available: https://pave-scan.eu/

[5] (2024). *MOST, Centro Nazionale Per La Mobilità Sostenibile*. [Online]. Available: https://www.centronazionalemost.it/eg/

[6] N. Sholevar, A. Golroo, and S. R. Esfahani, "Machine learning techniques for pavement condition evaluation," *Autom. Construct.*, vol. 136, Apr. 2022, Art. no. 104190.

[7] N. Ma, J. Fan, W. Wang, J. Wu, Y. Jiang, L. Xie, and R. Fan, "Computer vision for road imaging and pothole detection: A state-of-the-art review of systems and algorithms," *Transp. Saf. Environ.*, vol. 4, no. 4, Dec. 2022, Art. no. tdac026, doi: 10.1093/tse/tdac026.

[8] E. Ranyal, A. Sadhu, and K. Jain, "Road condition monitoring using smart sensing and artificial intelligence: A review," *Sensors*, vol. 22, no. 8, p. 3044, Apr. 2022.

[9] V. F. Vázquez, M. E. Hidalgo, A. M. García-Hoz, A. Camara, F. Terán, A. M. Ruiz-Teran, and S. E. Paje, "Tire/road noise, texture, and vertical accelerations: Surface assessment of an urban road," *Appl. Acoust.*, vol. 160, Mar. 2020, Art. no. 107153.

[10] J. Alonso, J. M. López, I. Pavón, M. Recuero, C. Asensio, G. Arcas, and A. Bravo, "On-board wet road surface identification using tyre/road noise and support vector machines," *Appl. Acoust.*, vol. 76, pp. 407–415, Feb. 2014.

[11] J. Masino, J. Pinay, M. Reischl, and F. Gauterin, "Road surface prediction from acoustical measurements in the tire cavity using support vector machine," *Appl. Acoust.*, vol. 125, pp. 41–48, Oct. 2017.

[12] F. G. Praticò, R. Fedele, V. Naumov, and T. Sauer, "Detection and monitoring of bottom-up cracks in road pavement using a machine-learning approach," *Algorithms*, vol. 13, no. 4, p. 81, Mar. 2020.

[13] S.-K. Lee, J. Yoo, C.-H. Lee, K. An, Y.-S. Yoon, J. Lee, G.-H. Yeom, and S.-U. Hwang, "Road type classification using deep learning for tire-pavement interaction noise data in autonomous driving vehicle," *Appl. Acoust.*, vol. 212, Sep. 2023, Art. no. 109597.

[14] A. Gagliardi, V. Staderini, and S. Saponara, "An embedded system for acoustic data processing and AI-based real-time classification for road surface analysis," *IEEE Access*, vol. 10, pp. 63073–63084, 2022.

[15] X. Li and D. W. Goldberg, "Toward a mobile crowdsensing system for road surface assessment," *Comput., Environ. Urban Syst.*, vol. 69, pp. 51–62, May 2018.

[16] K. Zang, J. Shen, H. Huang, M. Wan, and J. Shi, "Assessing and mapping of road surface roughness based on GPS and accelerometer sensors on bicycle-mounted smartphones," *Sensors*, vol. 18, no. 3, p. 914, Mar. 2018.

[17] A. Vittorio, V. Rosolino, I. Teresa, C. M. Vittoria, P. G. Vincenzo, and D. M. Francesco, "Automated sensing system for monitoring of road surface quality by mobile devices," *Proc.—Social Behav. Sci.*, vol. 111, pp. 242–251, Feb. 2014.

[18] V. K. Nguyen, É. Renault, and R. Milocco, "Environment monitoring for anomaly detection system using smartphones," *Sensors*, vol. 19, no. 18, p. 3834, Sep. 2019.

[19] A. Basavaraju, J. Du, F. Zhou, and J. Ji, "A machine learning approach to road surface anomaly assessment using smartphone sensors," *IEEE Sensors J.*, vol. 20, no. 5, pp. 2635–2647, Mar. 2020.

[20] P. Andrade, I. Silva, G. Signoretti, M. Silva, J. Dias, L. Marques, and D. Costa, "An unsupervised TinyML approach applied for pavement anomalies detection under the Internet of intelligent vehicles," in *Proc. IEEE Int. Workshop Metrology Ind. 4.0 IoT*, vol. 6, 2021, pp. 642–647.

[21] O. A. Egaji, G. Evans, M. G. Griffiths, and G. Islas, "Real-time machine learning-based approach for pothole detection," *Expert Syst. Appl.*, vol. 184, Dec. 2021, Art. no. 115562.

[22] I. Ferjani and S. Ali Alsaif, "How to get best predictions for road monitoring using machine learning techniques," *PeerJ Comput. Sci.*, vol. 8, p. e941, Apr. 2022.

[23] M. R. Carlos, M. E. Aragón, L. C. González, H. J. Escalante, and F. Martínez, "Evaluation of detection approaches for road anomalies based on accelerometer readings—Addressing who's who," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 10, pp. 3334–3343, Oct. 2018.

[24] L. C. González, R. Moreno, H. J. Escalante, F. Martínez, and M. R. Carlos, "Learning roadway surface disruption patterns using the bag of words representation," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 2916–2928, Nov. 2017.

[25] S. Sattar, S. Li, and M. Chapman, "Developing a near real-time road surface anomaly detection approach for road surface monitoring," *Measurement*, vol. 185, Nov. 2021, Art. no. 109990.

[26] R. Du, G. Qiu, K. Gao, L. Hu, and L. Liu, "Abnormal road surface recognition based on smartphone acceleration sensor," *Sensors*, vol. 20, no. 2, p. 451, Jan. 2020.

[27] Z. Zheng, M. Zhou, Y. Chen, M. Huo, L. Sun, S. Zhao, and D. Chen, "A fused method of machine learning and dynamic time warping for road anomalies detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 827–839, Feb. 2022.

[28] N. Chibani, F. Sebbak, W. Cherifi, and K. Belmessous, "Road anomaly detection using a dynamic sliding window technique," *Neural Comput. Appl.*, vol. 34, no. 21, pp. 19015–19033, Nov. 2022.

[29] B. Zhou, W. Zhao, W. Guo, L. Li, D. Zhang, Q. Mao, and Q. Li, "Smartphone-based road manhole cover detection and classification," *Autom. Construct.*, vol. 140, Aug. 2022, Art. no. 104344.

[30] D. Dong and Z. Li, "Smartphone sensing of road surface condition and defect detection," *Sensors*, vol. 21, no. 16, p. 5433, Aug. 2021.

[31] A. Shtayat, S. Moridpour, B. Best, and M. Abuhassan, "Using supervised machine learning algorithms in pavement degradation monitoring," *Int. J. Transp. Sci. Technol.*, vol. 12, no. 2, pp. 628–639, Jun. 2023.

[32] A. Martinelli, M. Meocci, M. Dolfi, V. Branzi, S. Morosi, F. Argenti, L. Berzi, and T. Consumi, "Road surface anomaly assessment using low-cost accelerometers: A machine learning approach," *Sensors*, vol. 22, no. 10, p. 3788, May 2022.

[33] A. Sabapathy and A. Biswas, "Road surface classification using accelerometer and speed data: Evaluation of a convolutional neural network model," *Neural Comput. Appl.*, vol. 35, no. 19, pp. 14183–14194, Jul. 2023.

[34] G. Wang, M. Burrow, and G. Ghataora, "Study of the factors affecting road roughness measurement using smartphones," *J. Infrastruct. Syst.*, vol. 26, no. 3, Sep. 2020, Art. no. 04020020.

[35] D. Luo, J. Lu, and G. Guo, "Road anomaly detection through deep learning approaches," *IEEE Access*, vol. 8, pp. 117390–117404, 2020.

[36] G. Baldini, R. Giuliani, and F. Geib, "On the application of time frequency convolutional neural networks to road Anomalies' identification with accelerometers and gyroscopes," *Sensors*, vol. 20, no. 22, p. 6425, Nov. 2020.

[37] C. Chen, H. Seo, and Y. Zhao, "A novel pavement transverse cracks detection model using WT-CNN and STFT-CNN for smartphone data analysis," *Int. J. Pavement Eng.*, vol. 23, no. 12, pp. 4372–4384, Oct. 2022.

[38] E. A. Martinez-Ríos, R. Bustamante-Bello, and S. A. Navarro-Tuch, "Generalized Morse wavelets parameter selection and transfer learning for pavement transverse cracking detection," *Eng. Appl. Artif. Intell.*, vol. 123, Aug. 2023, Art. no. 106355.

[39] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[40] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <1MB model size," 2016, *arXiv:1602.07360*.

[41] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[42] B. Varona, A. Monteserin, and A. Teyseyre, "A deep learning approach to automatic road surface monitoring and pothole detection," *Pers. Ubiquitous Comput.*, vol. 24, no. 4, pp. 519–534, Aug. 2020, doi: 0.1007/s00779-019-01234-z.

[43] F. M. Bianchi, S. Scardapane, S. Løkse, and R. Jenssen, "Reservoir computing approaches for representation and classification of multivariate time series," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 5, pp. 2169–2179, May 2021.

[44] A. Salman and A. N. Mian, "Deep learning based speed bumps detection and characterization using smartphone sensors," *Pervas. Mobile Comput.*, vol. 92, May 2023, Art. no. 101805, doi: 10.1016/j.pmcj.2023.101805.

[45] B. D. Setiawan, U. I. Serdült, and V. Kryssanov, "Smartphone sensor data augmentation for automatic road surface assessment using a small training dataset," in *Proc. IEEE Int. Conf. Big Data Smart Comput. (BigComp)*, Jan. 2021, pp. 239–245.

[46] B. D. Setiawan, M. Kovacs, U. Serdült, and V. Kryssanov, "Semantic segmentation on smartphone motion sensor data for road surface monitoring," *Proc. Comput. Sci.*, vol. 204, pp. 346–353, May 2022.

[47] J. Menegazzo and A. von Wangenheim, "Road surface type classification based on inertial sensors and machine learning," *Computing*, vol. 103, no. 10, pp. 2143–2170, Oct. 2021.

[48] J. Menegazzo. (2024). *Road Surface Type Classification*. [Online]. Available: https://www.kaggle.com/code/jefmenegazzo/road-surface-type-classification

[49] J. Menegazzo and A. von Wangenheim, "Speed bump detection through inertial sensors and deep learning in a multi-contextual analysis," *Social Netw. Comput. Sci.*, vol. 4, no. 1, p. 18, Oct. 2022.

[50] N. Hnoohom, S. Mekruksavanich, and A. Jitpattanakul, "A comprehensive evaluation of state-of-the-art deep learning models for road surface type classification," *Intell. Autom. Soft Comput.*, vol. 37, no. 2, pp. 1275–1291, 2023.

[51] P. Singh, A. E. Kamal, A. Bansal, and S. Kumar, "Road pothole detection from smartphone sensor data using improved LSTM," *Multimedia Tools Appl.*, vol. 83, no. 9, pp. 26009–26030, Aug. 2023.

[52] I. Khan and Z. Ahmed, "ML and DL classifications of route conditions using accelerometers and gyroscope sensors," in *Proc. 3rd Int. Conf. Artif. Intell. (ICAI)*, Feb. 2023, pp. 242–249.

[53] I. Siddiqui, S. Mazhar, N. Hassan, and W. Sultani, "Fine-grained road quality monitoring using deep learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 10, pp. 10691–10701, Oct. 2023.

[54] E. Raslan, M. F. Alrahmawy, Y. A. Mohammed, and A. S. Tolba, "IoT for measuring road network quality index," *Neural Comput. Appl.*, vol. 35, no. 3, pp. 2927–2944, Jan. 2023, doi: 10.1007/s00521-022-07736-x.

[55] M. Hijji, R. Iqbal, A. K. Pandey, F. Doctor, C. Karyotis, W. Rajeh, A. Alshehri, and F. Aradah, "6G connected vehicle framework to support intelligent road maintenance using deep learning data fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 7, pp. 7726–7735, Jul. 2023.

[56] A. Aboah and Y. Adu-Gyamfi, "Smartphone-based pavement roughness estimation using deep learning with entity embedding," *Adv. Data Sci. Adapt. Anal.*, vol. 12, Jul. 2020, Art. no. 2050007.

[57] J. Jeong, H. Jo, and G. Ditzler, "Convolutional neural networks for pavement roughness assessment using calibration-free vehicle dynamics," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 35, no. 11, pp. 1209–1229, Nov. 2020.

[58] F. Tyan, Y.-F. Hong, S.-H. Tu, and W. Jeng, "Generation of random road profiles," *J. Adv. Eng.*, vol. 4, no. 2, pp. 1373–1378, 2009.

[59] C. Liu, D. Wu, Y. Li, and Y. Du, "Large-scale pavement roughness measurements with vehicle crowdsourced data using semi-supervised learning," *Transp. Res. C, Emerg. Technol.*, vol. 125, Apr. 2021, Art. no. 103048.

[60] J.-H. Jeong and H. Jo, "Toward real-world implementation of deep learning for smartphone-crowdsourced pavement condition assessment," *IEEE Internet Things J.*, vol. 11, no. 4, pp. 6328–6337, Feb. 2024.

[61] P. Li, G. Hu, H. Xia, and R. Guo, "Efficient method based on recurrent neural networks for pavement evenness detection," *Measurement*, vol. 212, May 2023, Art. no. 112676.

[62] F. Ozoglu and T. Gökgöz, "Detection of road potholes by applying convolutional neural network method based on road vibration data," *Sensors*, vol. 23, no. 22, p. 9023, Nov. 2023.

[63] A. K. Pandey, R. Iqbal, T. Maniak, C. Karyotis, S. Akuma, and V. Palade, "Convolution neural networks for pothole detection of critical road infrastructure," *Comput. Electr. Eng.*, vol. 99, Apr. 2022, Art. no. 107725.

[64] H. Xin, Y. Ye, X. Na, H. Hu, G. Wang, C. Wu, and S. Hu, "Sustainable road pothole detection: A crowdsourcing based multi-sensors fusion approach," *Sustainability*, vol. 15, no. 8, p. 6610, Apr. 2023.

[65] K. Bansal, K. Mittal, G. Ahuja, A. Singh, and S. S. Gill, "DeepBus: Machine learning based real time pothole detection system for smart transportation using IoT," *Internet Technol. Lett.*, vol. 3, no. 3, p. e156, May 2020.

[66] S. Tiwari, R. Bhandari, and B. Raman, "RoadCare: A deep-learning based approach to quantifying road surface quality," in *Proc. 3rd ACM SIGCAS Conf. Comput. Sustain. Societies*. New York, NY, USA: ACM, Jun. 2020, pp. 231–242.

[67] A. Mihoub, M. Krichen, M. Alswailim, S. Mahfoudhi, and R. B. H. Salah, ''Road scanner: A road state scanning approach based on machine learning techniques,'' *Appl. Sci.*, vol. 13, no. 2, p. 683, Jan. 2023.

[68] C. Wu, Z. Wang, S. Hu, J. Lepine, X. Na, D. Ainalis, and M. Stettler, ''An automated machine-learning approach for road pothole detection using smartphone sensor data,'' *Sensors*, vol. 20, no. 19, p. 5564, Sep. 2020.

[69] S. Seid, M. Zennaro, M. Libse, and E. Pietrosemoli, ''Mobile crowdsensing based road surface monitoring using smartphone vibration sensor and lorawan,'' in *Proc. 1st Workshop Experiences Des. Implement. Frugal Smart Objects*. New York, NY, USA: ACM, Sep. 2020, pp. 36–41, doi: 10.1145/3410670.3410858.

[70] R. Leizerovych, G. Kondratenko, I. Sidenko, and Y. Kondratenko, ''IoT-complex for monitoring and analysis of motor highway condition using artificial neural networks,'' in *Proc. IEEE 11th Int. Conf. Dependable Syst., Services Technol. (DESSERT)*, Oct. 2020, pp. 207–212.

[71] R. Bustamante-Bello, A. García-Barba, L. A. Arce-Saenz, L. A. Curiel-Ramirez, J. Izquierdo-Reyes, and R. A. Ramirez-Mendoza, ''Visualizing street pavement anomalies through fog computing V2I networks and machine learning,'' *Sensors*, vol. 22, no. 2, p. 456, Jan. 2022.

[72] M. Bhosale, L. Guo, G. Comert, and Y. Jia, ''On-board smartphone-based road hazard detection with cloud-based fusion,'' *Vehicles*, vol. 5, no. 2, pp. 565–582, May 2023.

[73] N. Sabor and M. AbdelRaheem, ''CMNN-RADC: A crowdsensing convolutional-based mixer neural network road anomalies detector and classifier,'' *Internet Things*, vol. 22, Jul. 2023, Art. no. 100771.

[74] A. Ramesh, D. Nikam, V. N. Balachandran, L. Guo, R. Wang, L. Hu, G. Comert, and Y. Jia, ''Cloud-based collaborative road-damage monitoring with deep learning and smartphones,'' *Sustainability*, vol. 14, no. 14, p. 8682, Jul. 2022.

[75] W. S. Qureshi, S. I. Hassan, S. McKeever, D. Power, B. Mulry, K. Feighan, and D. O'Sullivan, ''An exploration of recent intelligent image analysis techniques for visual pavement surface condition assessment,'' *Sensors*, vol. 22, no. 22, p. 9019, Nov. 2022.

[76] M. Eisenbach, R. Stricker, D. Seichter, K. Amende, K. Debes, M. Sesselmann, D. Ebersbach, U. Stoeckert, and H.-M. Gross, ''How to get pavement distress detection ready for deep learning? A systematic approach,'' in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 2039–2047.

[77] 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 *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8.

[78] C. Zhang, E. Nateghinia, L. F. Miranda-Moreno, and L. Sun, ''Pavement distress detection using convolutional neural network (CNN): A case study in montreal, Canada,'' *Int. J. Transp. Sci. Technol.*, vol. 11, no. 2, pp. 298–309, Jun. 2022.

[79] P.-J. Chun, T. Yamane, and Y. Tsuzuki, ''Automatic detection of cracks in asphalt pavement using deep learning to overcome weaknesses in images and GIS visualization,'' *Appl. Sci.*, vol. 11, no. 3, p. 892, Jan. 2021.

[80] E. Eslami and H.-B. Yun, ''Attention-based multi-scale convolutional neural network (A+MCNN) for multi-class classification in road images,'' *Sensors*, vol. 21, no. 15, p. 5137, Jul. 2021.

[81] Aparna, Y. Bhatia, R. Rai, V. Gupta, N. Aggarwal, and A. Akula, ''Convolutional neural networks based potholes detection using thermal imaging,'' *J. King Saud Univ.—Comput. Inf. Sci.*, vol. 34, no. 3, pp. 578–588, Mar. 2022.

[82] C. Chen, S. Chandra, Y. Han, and H. Seo, ''Deep learning-based thermal image analysis for pavement defect detection and classification considering complex pavement conditions,'' *Remote Sens.*, vol. 14, no. 1, p. 106, Dec. 2021.

[83] M. Tan and Q. V. Le, ''EfficientNet: Rethinking model scaling for convolutional neural networks,'' 2019, *arXiv:1905.11946*.

[84] C. Chen, S. Chandra, and H. Seo, ''Automatic pavement defect detection and classification using RGB-thermal images based on hierarchical residual attention network,'' *Sensors*, vol. 22, no. 15, p. 5781, Aug. 2022.

[85] F. Liu, J. Liu, and L. Wang, ''Deep learning and infrared thermography for asphalt pavement crack severity classification,'' *Autom. Construct.*, vol. 140, Aug. 2022, Art. no. 104383.

[86] S. Zhou and W. Song, ''Deep learning–based roadway crack classification with heterogeneous image data fusion,'' *Struct. Health Monitor.*, vol. 20, no. 3, pp. 1274–1293, May 2021.

[87] L. Pei, Z. Sun, L. Xiao, W. Li, J. Sun, and H. Zhang, ''Virtual generation of pavement crack images based on improved deep convolutional generative adversarial network,'' *Eng. Appl. Artif. Intell.*, vol. 104, Sep. 2021, Art. no. 104376.

[88] B. Xu and C. Liu, ''Pavement crack detection algorithm based on generative adversarial network and convolutional neural network under small samples,'' *Measurement*, vol. 196, Jun. 2022, Art. no. 111219.

[89] Y. Que, Y. Dai, X. Ji, A. Kwan Leung, Z. Chen, Z. Jiang, and Y. Tang, ''Automatic classification of asphalt pavement cracks using a novel integrated generative adversarial networks and improved VGG model,'' *Eng. Struct.*, vol. 277, Feb. 2023, Art. no. 115406.

[90] R. Bibi, Y. Saeed, A. Zeb, T. M. Ghazal, T. Rahman, R. A. Said, S. Abbas, M. Ahmad, and M. A. Khan, ''Edge AI-based automated detection and classification of road anomalies in VANET using deep learning,'' *Comput. Intell. Neurosci.*, vol. 2021, no. 1, Jan. 2021, Art. no. 6262194.

[91] Y. Hou, Q. Li, Q. Han, B. Peng, L. Wang, X. Gu, and D. Wang, ''MobileCrack: Object classification in asphalt pavements using an adaptive lightweight deep learning,'' *J. Transp. Eng., B, Pavements*, vol. 147, no. 1, Mar. 2021, Art. no. 04020092, doi: 10.1061/jpeodx.0000245.

[92] S. Jana, A. I. Middya, and S. Roy, ''Participatory sensing based urban road condition classification using transfer learning,'' *Mobile Netw. Appl.*, vol. 29, no. 1, pp. 42–58, Feb. 2024.

[93] S. Patra, A. I. Middya, and S. Roy, ''PotSpot: Participatory sensing based monitoring system for pothole detection using deep learning,'' *Multimedia Tools Appl.*, p. 25171, Apr. 2021.

[94] Q. Zou, Z. Zhang, Q. Li, X. Qi, Q. Wang, and S. Wang, ''DeepCrack: Learning hierarchical convolutional features for crack detection,'' *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1498–1512, Mar. 2019.

[95] M. D. Jenkins, T. A. Carr, M. I. Iglesias, T. Buggy, and G. Morison, ''A deep convolutional neural network for semantic pixel-wise segmentation of road and pavement surface cracks,'' in *Proc. 26th Eur. Signal Process. Conf. (EUSIPCO)*, Sep. 2018, pp. 2120–2124.

[96] J. Long, E. Shelhamer, and T. Darrell, ''Fully convolutional networks for semantic segmentation,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2015, pp. 3431–3440.

[97] F. Yang, L. Zhang, S. Yu, D. Prokhorov, X. Mei, and H. Ling, ''Feature pyramid and hierarchical boosting network for pavement crack detection,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 4, pp. 1525–1535, Apr. 2020.

[98] C. Han, T. Ma, J. Huyan, X. Huang, and Y. Zhang, ''CrackW-Net: A novel pavement crack image segmentation convolutional neural network,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 22135–22144, Nov. 2022.

[99] W. Wang and C. Su, ''Convolutional neural network-based pavement crack segmentation using pyramid attention network,'' *IEEE Access*, vol. 8, pp. 206548–206558, 2020.

[100] S. L. H. Lau, E. K. P. Chong, X. Yang, and X. Wang, ''Automated pavement crack segmentation using U-Net-based convolutional neural network,'' *IEEE Access*, vol. 8, pp. 114892–114899, 2020.

[101] Y. Liu, J. Yao, X. Lu, R. Xie, and L. Li, ''DeepCrack: A deep hierarchical feature learning architecture for crack segmentation,'' *Neurocomputing*, vol. 338, pp. 139–153, Apr. 2019.

[102] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, ''Rethinking atrous convolution for semantic image segmentation,'' 2017, *arXiv:1706.05587*.

[103] Z. Qu, C. Cao, L. Liu, and D.-Y. Zhou, ''A deeply supervised convolutional neural network for pavement crack detection with multiscale feature fusion,'' *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4890–4899, Sep. 2022.

[104] W. Song, G. Jia, H. Zhu, D. Jia, and L. Gao, ''Automated pavement crack damage detection using deep multiscale convolutional features,'' *J. Adv. Transp.*, vol. 2020, p. 6412562, Jan. 2020.

[105] Z. Fan, C. Li, Y. Chen, J. Wei, G. Loprencipe, X. Chen, and P. Di Mascio, ''Automatic crack detection on road pavements using encoder–decoder architecture,'' *Materials*, vol. 13, no. 13, p. 2960, Jul. 2020.

[106] Z. Qu, W. Chen, S.-Y. Wang, T.-M. Yi, and L. Liu, ''A crack detection algorithm for concrete pavement based on attention mechanism and multi-features fusion,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 8, pp. 11710–11719, Aug. 2022.

[107] S.-H. Gao, M.-M. Cheng, K. Zhao, X.-Y. Zhang, M.-H. Yang, and P. Torr, ''Res2Net: A new multi-scale backbone architecture,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 2, pp. 652–662, Feb. 2021.

[108] X. Sun, Y. Xie, L. Jiang, Y. Cao, and B. Liu, ''DMA-Net: DeepLab with multi-scale attention for pavement crack segmentation,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 18392–18403, Oct. 2022.

[109] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," 2014, *arXiv:1409.0575*.

[110] Q. Zhou, Z. Qu, S.-Y. Wang, and K.-H. Bao, "A method of potentially promising network for crack detection with enhanced convolution and dynamic feature fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 10, pp. 18736–18745, Oct. 2022.

[111] Z. Al-Huda, B. Peng, R. N. A. Algburi, M. A. Al-antari, R. Al-Jarazi, and D. Zhai, "A hybrid deep learning pavement crack semantic segmentation," *Eng. Appl. Artif. Intell.*, vol. 122, Jun. 2023, Art. no. 106142.

[112] L. Yang, S. Bai, Y. Liu, and H. Yu, "Multi-scale triple-attention network for pixelwise crack segmentation," *Autom. Construct.*, vol. 150, Jun. 2023, Art. no. 104853.

[113] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, "ENet: A deep neural network architecture for real-time semantic segmentation," 2016, *arXiv:1606.02147*.

[114] W. Wang and C. Su, "Deep learning-based real-time crack segmentation for pavement images," *KSCE J. Civil Eng.*, vol. 25, no. 12, pp. 4495–4506, Dec. 2021, doi: 10.1007/s12205-021-0474-2.

[115] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "BiSeNet: Bilateral segmentation network for real-time semantic segmentation," 2018, *arXiv:1808.00897*.

[116] S. Jégou, M. Drozdzal, D. Vazquez, A. Romero, and Y. Bengio, "The one hundred layers tiramisu: Fully convolutional DenseNets for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1175–1183.

[117] P. Yong and N. Wang, "RIIAnet: A real-time segmentation network integrated with multi-type features of different depths for pavement cracks," *Appl. Sci.*, vol. 12, no. 14, p. 7066, Jul. 2022.

[118] G. Doğan and B. Ergen, "A new mobile convolutional neural network-based approach for pixel-wise road surface crack detection," *Measurement*, vol. 195, May 2022, Art. no. 111119.

[119] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," 2018, *arXiv:1801.04381*.

[120] J. Liu, X. Yang, S. Lau, X. Wang, S. Luo, V. C. Lee, and L. Ding, "Automated pavement crack detection and segmentation based on two-step convolutional neural network," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 35, no. 11, pp. 1291–1305, Nov. 2020.

[121] N. H. T. Nguyen, S. Perry, D. Bone, H. T. Le, and T. T. Nguyen, "Two-stage convolutional neural network for road crack detection and segmentation," *Expert Syst. Appl.*, vol. 186, Dec. 2021, Art. no. 115718.

[122] Q. Yang and X. Ji, "Automatic pixel-level crack detection for civil infrastructure using Unet++ and deep transfer learning," *IEEE Sensors J.*, vol. 21, no. 17, pp. 19165–19175, Sep. 2021.

[123] 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," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 32, no. 10, pp. 805–819, Oct. 2017.

[124] Y. Fei, K. C. P. Wang, A. Zhang, C. Chen, J. Q. Li, Y. Liu, G. Yang, and B. Li, "Pixel-level cracking detection on 3D asphalt pavement images through deep-learning-based CrackNet-V," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 1, pp. 273–284, Jan. 2020.

[125] Y. Tang, A. A. Zhang, L. Luo, G. Wang, and E. Yang, "Pixel-level pavement crack segmentation with encoder–decoder network," *Measurement*, vol. 184, Nov. 2021, Art. no. 109914.

[126] Z. Hong, F. Yang, H. Pan, R. Zhou, Y. Zhang, Y. Han, J. Wang, S. Yang, P. Chen, X. Tong, and J. Liu, "Highway crack segmentation from unmanned aerial vehicle images using deep learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[127] B. Wang, "AerialCrackDataset: Towards object detection with dataset," Key Lab. Optoelectron. Imag. Technol. Syst., School Optoelectron., Beijing Inst. Technol., Beijing, China, Tech. Rep., 2017.

[128] J. Guan, X. Yang, L. Ding, X. Cheng, V. C. S. Lee, and C. Jin, "Automated pixel-level pavement distress detection based on stereo vision and deep learning," *Autom. Construct.*, vol. 129, Sep. 2021, Art. no. 103788.

[129] W. Tang, S. Huang, Q. Zhao, R. Li, and L. Huangfu, "An iteratively optimized patch label inference network for automatic pavement distress detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 7, pp. 8652–8661, Jul. 2022.

[130] Q. Mei and M. Gül, "A cost effective solution for pavement crack inspection using cameras and deep neural networks," *Construct. Building Mater.*, vol. 256, Sep. 2020, Art. no. 119397.

[131] S. Shim, J. Kim, S.-W. Lee, and G.-C. Cho, "Road damage detection using super-resolution and semi-supervised learning with generative adversarial network," *Autom. Construct.*, vol. 135, Mar. 2022, Art. no. 104139.

[132] H. Oliveira and P. L. Correia, "Automatic road crack detection and characterization," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 1, pp. 155–168, Mar. 2013.

[133] R. Amhaz, S. Chambon, J. Idier, and V. Baltazart, "Automatic crack detection on two-dimensional pavement images: An algorithm based on minimal path selection," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 10, pp. 2718–2729, Oct. 2016.

[134] X. Yang, H. Li, Y. Yu, X. Luo, T. Huang, and X. Yang, "Automatic pixel-level crack detection and measurement using fully convolutional network," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 12, pp. 1090–1109, Dec. 2018.

[135] K. Liu, X. Han, and B. M. Chen, "Deep learning based automatic crack detection and segmentation for unmanned aerial vehicle inspections," in *Proc. IEEE Int. Conf. Robot. Biomimetics (ROBIO)*, Dec. 2019, pp. 381–387.

[136] F. Guo, Y. Qian, J. Liu, and H. Yu, "Pavement crack detection based on transformer network," *Autom. Construct.*, vol. 145, Jan. 2023, Art. no. 104646.

[137] H. Maeda, Y. Sekimoto, T. Seto, T. Kashiyama, and H. Omata, "Road damage detection and classification using deep neural networks with smartphone images," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 33, no. 12, pp. 1127–1141, Dec. 2018.

[138] H. Maeda, T. Kashiyama, Y. Sekimoto, T. Seto, and H. Omata, "Generative adversarial network for road damage detection," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 36, no. 1, pp. 47–60, Jan. 2021.

[139] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, and Y. Sekimoto, "RDD2020: An annotated image dataset for automatic road damage detection using deep learning," *Data Brief*, vol. 36, Jun. 2021, Art. no. 107133.

[140] D. Arya, H. Maeda, S. K. Ghosh, D. Toshniwal, and Y. Sekimoto, "RDD2022: A multi-national image dataset for automatic road damage detection," 2022, *arXiv:2209.08538*.

[141] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Computer Vision—ECCV 2016* (Lecture Notes in Computer Science), vol. 9905, 2016, p. 21.

[142] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, Red Hook, NY, USA, C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, Eds. Curran Associates, 2015, pp. 1–14.

[143] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[144] M.-T. Cao, Q.-V. Tran, N.-M. Nguyen, and K.-T. Chang, "Survey on performance of deep learning models for detecting road damages using multiple dashcam image resources," *Adv. Eng. Inf.*, vol. 46, Oct. 2020, Art. no. 101182.

[145] Y. Du, N. Pan, Z. Xu, F. Deng, Y. Shen, and H. Kang, "Pavement distress detection and classification based on YOLO network," *Int. J. Pavement Eng.*, vol. 22, no. 13, pp. 1659–1672, Nov. 2021.

[146] J. Zhu, J. Zhong, T. Ma, X. Huang, W. Zhang, and Y. Zhou, "Pavement distress detection using convolutional neural networks with images captured via UAV," *Autom. Construct.*, vol. 133, Jan. 2022, Art. no. 103991.

[147] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A new backbone that can enhance learning capability of CNN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 1571–1580.

[148] K. R. Ahmed, "Smart pothole detection using deep learning based on dilated convolution," *Sensors*, vol. 21, no. 24, p. 8406, Dec. 2021.

[149] X. Xu, M. Zhao, P. Shi, R. Ren, X. He, X. Wei, and H. Yang, "Crack detection and comparison study based on faster R-CNN and mask R-CNN," *Sensors*, vol. 22, no. 3, p. 1215, Feb. 2022.

[150] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[151] S.-S. Park, V.-T. Tran, and D.-E. Lee, "Application of various YOLO models for computer vision-based real-time pothole detection," *Appl. Sci.*, vol. 11, no. 23, p. 11229, Nov. 2021.

[152] H. Chen, M. Yao, and Q. Gu, "Pothole detection using location-aware convolutional neural networks," *Int. J. Mach. Learn. Cybern.*, vol. 11, no. 4, pp. 899–911, Apr. 2020.

[153] L. A. Silva, H. S. S. Blas, D. P. García, A. S. Mendes, and G. V. González, "An architectural multi-agent system for a pavement monitoring system with pothole recognition in UAV images," *Sensors*, vol. 20, no. 21, p. 6205, Oct. 2020.

[154] N. Yang, Y. Li, and R. Ma, "An efficient method for detecting asphalt pavement cracks and sealed cracks based on a deep data-driven model," *Appl. Sci.*, vol. 12, no. 19, p. 10089, Oct. 2022.

[155] L. A. Silva, V. R. Q. Leithardt, V. F. L. Batista, G. V. González, and J. F. De P. Santana, "Automated road damage detection using UAV images and deep learning techniques," *IEEE Access*, vol. 11, pp. 62918–62931, 2023.

[156] D. Wang, Z. Liu, X. Gu, W. Wu, Y. Chen, and L. Wang, "Automatic detection of pothole distress in asphalt pavement using improved convolutional neural networks," *Remote Sens.*, vol. 14, no. 16, p. 3892, Aug. 2022.

[157] Y. Zhang, Z. Zuo, X. Xu, J. Wu, J. Zhu, H. Zhang, J. Wang, and Y. Tian, "Road damage detection using UAV images based on multi-level attention mechanism," *Autom. Construct.*, vol. 144, Dec. 2022, Art. no. 104613.

[158] Z. Liu, X. Gu, J. Chen, D. Wang, Y. Chen, and L. Wang, "Automatic recognition of pavement cracks from combined GPR B-scan and C-scan images using multiscale feature fusion deep neural networks," *Autom. Construct.*, vol. 146, Feb. 2023, Art. no. 104698.

[159] Z. Liu, X. Gu, H. Yang, L. Wang, Y. Chen, and D. Wang, "Novel YOLOv3 model with structure and hyperparameter optimization for detection of pavement concealed cracks in GPR images," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 22258–22268, Nov. 2022.

[160] G. Jocher. (2020). *Ultralytics YOLOv5*. [Online]. Available: https://github.com/ultralytics/yolov5

[161] H. Salaudeen and E. Çelebi, "Pothole detection using image enhancement GAN and object detection network," *Electronics*, vol. 11, no. 12, p. 1882, Jun. 2022.

[162] F. Wan, C. Sun, H. He, G. Lei, L. Xu, and T. Xiao, "YOLO-LRDD: A lightweight method for road damage detection based on improved YOLOv5s," *EURASIP J. Adv. Signal Process.*, vol. 2022, no. 1, p. 98, Oct. 2022, doi: 10.1186/s13634-022-00931-x.

[163] M. Ren, X. Zhang, X. Chen, B. Zhou, and Z. Feng, "YOLOv5s-M: A deep learning network model for road pavement damage detection from urban street-view imagery," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 120, Jun. 2023, Art. no. 103335.

[164] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO series in 2021," 2021, *arXiv:2107.08430*.

[165] C. Y. Wang, I. H. Yeh, and H. Y. M. Liao, "You only learn one representation: Unified network for multiple tasks," *J. Inf. Sci. Eng.*, vol. 39, no. 2, pp. 691–709, 2021.

[166] C.-Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 7464–7475.

[167] A. M. Roy and J. Bhaduri, "DenseSPH-YOLOv5: An automated damage detection model based on DenseNet and swin-transformer prediction head-enabled YOLOv5 with attention mechanism," *Adv. Eng. Inf.*, vol. 56, Apr. 2023, Art. no. 102007.

[168] X. Zhu, S. Lyu, X. Wang, and Q. Zhao, "TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 2778–2788.

[169] M. H. Asad, S. Khaliq, M. H. Yousaf, M. O. Ullah, and A. Ahmad, "Pothole detection using deep learning: A real-time and AI-on-the-edge perspective," *Adv. Civil Eng.*, vol. 2022, no. 1, Apr. 2022, Art. no. 9221211.

[170] S.-A. Hassan, T. Rahim, and S.-Y. Shin, "An improved deep convolutional neural network-based autonomous road inspection scheme using unmanned aerial vehicles," *Electronics*, vol. 10, no. 22, p. 2764, Nov. 2021.

[171] D. Ma, H. Fang, N. Wang, C. Zhang, J. Dong, and H. Hu, "Automatic detection and counting system for pavement cracks based on PCGAN and YOLO-MF," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 22166–22178, Nov. 2022.

[172] J. Xing, Y. Liu, and G.-Z. Zhang, "Improved YOLOV5-based UAV pavement crack detection," *IEEE Sensors J.*, vol. 23, no. 14, pp. 15901–15909, Jul. 2023.

LORENZO MANONI (Member, IEEE) received the B.Sc. and M.Sc. degrees in electronics engineering and the Ph.D. degree in information engineering from the Università Politecnica delle Marche, Ancona, Italy, in 2015, 2018, and 2022, respectively. He was a Research Fellow with the Department of Information Engineering (DII), in 2022. He is currently a Researcher with DII. His current research interests include embedded systems, machine learning, deep learning, artificial intelligence applications, convolutional neural networks, signal processing, algorithms analysis and design, and bio-signal analysis.

SIMONE ORCIONI (Senior Member, IEEE) received the Laurea and Ph.D. degrees in electronics engineering from the Università Politecnica delle Marche, Italy, in 1992 and 1995, respectively.

In 2000, he became an Assistant Professor, teaching courses in analog and digital electronics and publishing a text book. In 2017, he was an Adjunct Professor with the Ubiquitous Computing Laboratory (UC-Laboratory), HTWG Konstanz, where he is currently a Guest Researcher. Since 2021, he has been an Associate Professor with the Department of Information Engineering, Università Politecnica delle Marche, where he is also the President of the Unified Council of the Electronic Engineering Study Course. He has published more than 50 articles in journals and more than 100 in international conference proceedings and international book chapters. He was a Guest Editor of *EURASIP Journal on Embedded Systems*, *Frontiers in Energy Research*, and *Sensors* (MDPI); a reviewer of 18 international journals; a program committee member of seven international conferences; the program chair of three international conferences; an editor of four international books; and an inventor in two patents. He has been confirmed in the World's Top 2% Scientists by Stanford University in both the "2023 Annual Influence Ranking" and "Lifetime Scientific Influence Ranking." He has been working in statistical device modeling and simulation, analog circuit design, cyber-physical system simulation, and linear and nonlinear system identification. His current research interests include nonlinear digital signal processing and electronics for renewable energies. He is an EURASIP Member.

MASSIMO CONTI (Member, IEEE) received the Graduate degree in electronics engineering from the University of Ancona, Italy, in 1987. He is currently an Associate Professor with the Dipartimento di Ingegneria dell'Informazione (DII), Università Politecnica delle Marche (UNIVPM), Ancona, Italy. He is the co-author of more than 250 papers on international books, journals, or conferences. His Scopus: 190 publications, 1378 citations, and H-index: 18. He was the Coordinator of European and national research projects. He is also an editor of 11 international books. His research interests include microelectronics is mainly devoted to system level design of low power integrated circuits, electronic smart systems for ambient assisted living, design of energy harvesting systems, battery management systems, V2G vehicle to grid connection, smart grids, state of health estimation, analysis of cell mismatch on battery pack performances, design of BMS and battery life tracing for reuse, recycle and end-of-life, smart mobility, and NFC for food traceability. He is also a lead guest editor of special issue of international journals. He is also the General Chairperson of ten international conferences. For more information visit the link (www.univpm.it/massimo.conti).

Open Access funding provided by 'Università Politecnica delle Marche' within the CRUI CARE Agreement