Western University
Scholarship@Western

Electrical and Computer Engineering Publications
Electrical and Computer Engineering Department

2021

A Systematic Review of Convolutional Neural Network-Based Structural Condition Assessment Techniques

Sandeep Sony
Western University

Kyle Dunphy
Western University

Ayan Sadhu
Western University, asadhu@uwo.ca

Miriam A M Capretz
Western University, mcapretz@uwo.ca

Follow this and additional works at: https://ir.lib.uwo.ca/electricalpub

Part of the Civil and Environmental Engineering Commons, Computer Engineering Commons, and the Electrical and Computer Engineering Commons

Citation of this paper:
Sony, Sandeep; Dunphy, Kyle; Sadhu, Ayan; and Capretz, Miriam A M, "A Systematic Review of Convolutional Neural Network-Based Structural Condition Assessment Techniques" (2021). *Electrical and Computer Engineering Publications*. 184.
https://ir.lib.uwo.ca/electricalpub/184

A Systematic Review of Convolutional Neural Network-Based Structural Condition Assessment Techniques

Sandeep Sony[1], Kyle Dunphy[2], Ayan Sadhu[3], and Miriam Capretz[4]

[1]PhD Student, Department of Civil and Environmental Engineering, Western University.

[2]MSc Student, Department of Civil and Environmental Engineering, Western University.

[3]Corresponding Author: Assistant Professor, Department of Civil and Environmental Engineering, Western University, London, Ontario, Canada. Email: asadhu@uwo.ca.

[4]Professor, Department of Electrical and Computer Engineering, Western University.

Abstract

With recent advances in non-contact sensing technology such as cameras, unmanned aerial and ground vehicles, the structural health monitoring (SHM) community has witnessed a prominent growth in deep learning-based condition assessment techniques of structural systems. These deep learning methods rely primarily on convolutional neural networks (CNNs). The CNN networks are trained using a large number of datasets for various types of damage and anomaly detection and post-disaster reconnaissance. The trained networks are then utilized to analyze newer data to detect the type and severity of the damage, enhancing the capabilities of non-contact sensors in developing autonomous SHM systems. In recent years, a broad range of CNN architectures has been developed by researchers to accommodate the extent of lighting and weather conditions, the quality of images, the amount of background and foreground noise, and multiclass damage in the structures. This paper presents a detailed literature review of existing CNN-based techniques in the context of infrastructure monitoring and maintenance. The review is categorized into multiple classes depending on the specific application and development of CNNs applied to data obtained from a wide range of structures. The challenges and limitations of the existing literature are discussed in detail at the end, followed by a brief conclusion on potential future research directions of CNN in structural condition assessment.

Keywords: Structural health monitoring, artificial intelligence, deep learning, CNN, damage detection, anomaly detection, structural condition assessment.

Sony, S., Dunphy, K., Sadhu, A., & Capretz, M. A systematic review of convolutional neural network-based structural condition assessment techniques. Engineering Structures, 226, 111347.

Table 1. List of acronyms.

Acronym	**Description**
AdaBoost	Adaptive Boosting
AE	Auto Encoder
CNN	Convolutional Neural Network
DBN	Deep Belief Network
DBM	Deep Boltzmann Machine
DL	Deep Learning
FCN	Fully Convolutional Network
*k*NN	*k*-nearest Neighbor
ML	Machine Learning
NN	Neural Network
ReLU	Rectified Linear Unit
ResNet	Residual Network
R-CNN	Regional Convolutional Neural Network
RNN	Recurrent Neural Networks
ROC	Receiver Operating Characteristic
SHM	Structural Health Monitoring
SVM	Support Vector Machine
TL	Transfer Learning
VGG	Visual Geometry Group

1. Introduction

Structural health monitoring (SHM) offers emerging and powerful diagnostic tools for damage detection, maintenance, life-cycle cost reduction, and rapid disaster management for structures

(Cawley 2018). Most of these techniques rely on dynamic measurements that require installation of contact sensors such as accelerometers, strain gauges, fiber optic sensors, and ultrasonic wave sensors, which have high installation costs. With the recent development of next-generation sensors (Sony *et al.* 2019; Dabous and Feroz 2020) such as digital and high-speed cameras, unmanned ground vehicles (UGVs), and mobile sensors, there has been a radical shift to non-contact sensing techniques in SHM. They are easier to deploy, less labor-intensive, and more cost-effective, enabling more reliable data acquisition from structures with high-resolution temporal and spatial information (Lattanzi and Miller 2017; Almasri *et al.* 2020). However, unlike traditional contact sensors, non-contact sensors yield images and videos that require significant advances in robotics, image processing, computer vision, and deep learning algorithms, where structural engineers still face several challenges. In recent years, the SHM researchers have explored artificial intelligence techniques to solve these challenges and successfully achieve novel autonomous and intelligent inspection strategies using the non-contact and robotic devices. This research not only accelerates monitoring and maintenance tasks for the infrastructure owners but also allows accurate early-stage defect detection to prevent any catastrophic structural failure in the future. Moreover, the research advancement in this area enables improved structural maintenance with minimal human errors, lower costs, and higher accuracy, providing an end-to-end system to the infrastructure owners. This research has resulted in numerous publications in top-notch structural engineering journals. The main objective of this paper is to provide a systematic review of recent convolutional neural network (a subset of deep learning methods)-based techniques that have been widely developed in the context of non-contact sensing-based SHM.

A non-contact sensor such as a camera, where each pixel is effectively a sensor, can remotely collect a large amount of data from a structure. The challenge is then to interpret these images or videos for decision-making in SHM. Since the last decade, the SHM community has seen significant development in various image-processing algorithms that have enhanced the capabilities of non-contact sensors to undertake structural condition assessment. For example, Jahanshahi *et al.* (2009) reviewed various image processing techniques that were explored for the detection of missing or deformed members, cracks, and corrosion in various structures. A suite of image-based crack acquisition, processing, and interpretation techniques specifically for asphalt pavement was presented by Zakeri *et al.* (2017). Along similar lines, Koch *et al.* (2015) presented a comprehensive summary of various image processing techniques that have been used to identify damage patterns in concrete bridges, tunnels, pipes, and pavement. Recently, Mohan and Poobal (2018) reviewed various image processing techniques for detecting cracks in concrete surfaces and concluded that the direction of the crack was crucial to the ability to detect and quantify the size of cracks.

Overall, existing image processing methods extract features from images using various edges or boundary detection techniques such as the fast Haar transform, Canny filter, Sobel edge detector, morphological detectors, template matching, background subtraction, and texture recognition methods. However, these methods often result in ill-posed problems due to disturbances created by environmental conditions such as light, distortion, weather, shade, and occlusion in outdoor civil structures (Lee *et al.* 2014). The SHM community has recently focused on overcoming these challenges using various computer vision and artificial intelligence (AI) techniques due to their reduced sensitivity to external disturbances and feature selection. Salehi and Burgueno (2018) reviewed a suite of various artificial intelligence (AI) methods that have recently been used in structural engineering. The authors showed the recent trend of AI-assisted research towards pattern recognition and machine learning-based automated data-driven methods. The relative merits and drawbacks of various AI methods were discussed in the context of various structural engineering applications. This paper reviews CNN-based deep learning techniques with a specific focus on the implementation of non-contact sensor-based SHM.

Although AI is a broad area of research covering various engineering disciplines, machine learning (ML) and deep learning (DL) techniques are the two most popular branches of AI that have been heavily explored in SHM research. ML algorithms are trained on a wide variety of data, and the accuracy of the algorithms improves with more data. The purpose of training is to optimize the error along the dimensions of the dataset using optimization functions such as a loss function or objective function and to obtain the best prediction results for test data. However, ML algorithms need features that are obtained from different image processing methods and are fed into different classifiers. Depending on the application, a suitable choice of features and classifiers is essential to identify anomalies from the images.

Ying *et al.* (2013) reviewed various ML-based SHM algorithms for isolating structural damage to steel pipes from environmental factors. Recently, another review paper written by Feng and Feng

(2018) provided an intensive literature review of state-of-the-art computer vision techniques using vision-based displacement sensors that were implemented for SHM. Most of these methods were based on template matching algorithms that extracted displacement time-histories from videos and images. The authors discussed various challenges of displacement extraction from videos obtained from 2D and 3D measurements and from artificial or natural targets, as well as their real-time and preprocessing applications. In particular, Gomes *et al.* (2018) presented a comprehensive review of intelligent computational tools available for damage detection and system identification, with a specific emphasis on composite structures. More recently, state-of-the-art vision-based structural condition assessment techniques using computer vision and ML algorithms were reviewed by Spencer *et al.* (2019). The challenges associated with static and dynamic measurement techniques were discussed, along with future directions of automated and improved decision-making methods for SHM. Overall, it can be concluded from the literature that ML methods rely heavily on feature extraction, followed by the application of suitable classifiers. These methods can manage small anomaly datasets, but may not be adequate for full-scale civil structures such as buildings, bridges, dams, pipelines, and wind turbines where crack patterns are complex and irregular (Yao *et al.* 2014).

Unlike ML, DL-based AI methods automatically extract features and eliminate the need for manual feature extraction. Therefore, DL can differentiate among a large number of classes, and this capability has been recently explored for damage evaluation in structures. DL algorithms are based on vast sets of labeled data and require high computational performance and memory requirements. The term "*deep*" refers to the large number of layers that exist between the raw image input and the final classification output used in a network. Convolutional neural networks (CNNs), which are a popular class of DL methods, have been successfully used since their breakthrough in the 2012 ImageNet challenge due to their ability to extract features automatically. This has enabled automatic and optimized feature extraction to become part of the classifier learning process, which, however, does not compromise its optimality or the accuracy of crack identification. In particular, Bao *et al.* (2019) briefly reviewed improved SHM techniques that explored various data science, computer vision, DL, and ML methods. It was concluded that the application of DL, ML, and computer vision techniques made it possible to extract pertinent data from noisy measurement databases with damage signatures and to analyze them without requiring any predefined classifiers. Zhao *et al.* (2015) and Lei *et al.* (2020) summarized various ML and

DL techniques and their applications that are specific to machine health monitoring. It was concluded that DL techniques were the most effective because they are not restricted to specific machine types and involve minimal human intervention. Recently, Ye *et al.* (2019) provided a general survey and overview of various DL techniques in the context of SHM. Considering the intensity of CNN-based literature in the field of infrastructure monitoring, this paper is intended to provide a systematic review of standalone CNN-based literature that is specific to structural condition assessment.

The key objectives of this review paper are as follows:

1. To review CNN-specific papers that have been recently explored for structural condition assessment, with a specific focus on structural damage and anomaly detection. Similar to the condition monitoring of machines, there has been a significant trend towards using CNN to undertake local damage assessment and anomaly detection in large-scale civil structures. The primary objective of this paper is to conduct a detailed survey of emerging CNN-based SHM papers and to provide a comprehensive review of more than one hundred papers that have been recently published on this topic.
2. To compare existing CNN-based solutions and best practices to address the challenges of infrastructure monitoring and maintenance, which would provide valuable opportunities and guidance to future engineers and researchers to adopt the most relevant CNN architecture depending on their applications.
3. To provide a perspective on CNN-based methods in the domain of SHM that would facilitate valuable feature selection and anomaly detection methodologies in other areas of structural engineering and the broader field of civil engineering.
4. To provide the key challenges of the current literature and identify the potential future research directions of the CNN-based research in structural condition assessment.

This paper is structured as follows. A brief overview of various DL methods and CNN techniques is presented first. Next, the details of various CNN-based condition assessment techniques and their recent applications in structural condition assessment are presented. Different hybrid methods based on CNN are then presented, followed by key conclusions and discussions.

2. Preliminaries of Deep Learning Methods

Non-contact sensing techniques (Sony *et al.* 2019; Dabous and Feroz 2020) and computer vision (Feng and Feng 2018; Spencer *et al.* 2019; Dick *et al.* 2019) have opened up a new era of next-generation autonomous SHM and inspection of large-scale structures. These sensors result in images and videos, requiring AI techniques to analyze complex input-output relationships of the training data and develop predictive models. The trained predictive models are then used for damage classification, localization, and prediction from the new measurement data of a wide range of structures. The objective of this paper is to review CNN-based SHM papers that have been published in the specific context of structural condition assessment. A brief background on DL methods is presented next, followed by a detailed background on CNN techniques.

DL algorithms have an adaptable nature similar to the human brain. These algorithms become more accurate as more training data are provided to them. DL models can simultaneously learn representation and decision rules from the data, like the biological organisms by which they are inspired. DL methods have multiple layers of non-linear transformations. For example, a raw image dataset that is fed through any DL architecture passes through several layers. Each layer, starting with the input layer, improves the identification of the dataset with subsequent layers, and eventually produces a classification or identification at the output layer (Lee *et al.* 2018). The most prominent aspect of DL is that these layers are not designed by engineers, but rather are learned from the data using a general-purpose learning procedure (LeCun *et al.* 2015). The advantage of DL is that it requires minimal user intervention, which has attracted various interdisciplinary researchers to use it for a wide range of applications such as object detection, classification, and segmentation.

In the context of SHM, DL can be used for damage detection in three ways: (a) classification, i.e., labeling an image as damaged or undamaged, (b) localization, i.e., locating the regions where damage exists using bounding boxes and identifying their coordinates, (c) segmentation, i.e., segmenting the pixels of an image into damaged and undamaged pixels (e.g., labeling of all pixels). In the last few years, several methods have been developed, including, but not limited to, the audio signal, time-series, video, and natural language datasets. DL methods (Goodfellow et al. 2016) have several variants such as Auto Encoders (AEs), Deep Belief Networks (DBNs), Deep

Boltzmann Machines (DBMs), Recurrent Neural Networks (RNNs), and Convolutional Neural Networks (CNNs).

The AE algorithm is used to learn data coding in an unsupervised manner to create a representation for a dataset by dimensionality reduction, ignoring the noise in the dataset (Vincent *et al.* 2008). DBN is a probabilistic generative model composed of multiple layers of stochastic and latent variables. If the number of units in the highest layer is small, DBN performs non-linear dimensionality reduction and can learn short binary codes that enable very fast retrieval of datasets (Hinton *et al.* 2006). DBM is a type of binary pairwise Markov random field with multiple layers of hidden random variables. Similarly to DBN, DBM can learn a complex and abstract internal representation of the input dataset using a limited amount of labeled data (Salakhutdinov and Hinton 2009). RNNs are designed and tested for sequential data, typically for application in dynamic systems such as time-series or speech and language. RNNs are the deepest of all neural networks and can generate memories of arbitrary sequences of input patterns (Funahashi and Nakamura 1993). However, CNNs require less statistical and probabilistic expertise to run and to infer the dataset and results, which makes them a preferred choice for researchers in the SHM community. The next section presents a detailed background on CNN, followed by a systematic literature review of non-contact sensor-based SHM using CNN.

3. Background on Convolutional Neural Networks

CNN is the most popular variant of the DL network. The underlying architecture of CNN is comprised of three layers: (a) convolutional (feature extraction), (b) pooling (dimensionality reduction), and (c) fully-connected layer. The convolutional layer contains a finite number of filters (defined by the kernel or filter size) that convolves with the input data and identify a large number of relevant features from the input image. The pooling layer reduces the dimensions of the resulting features using a down-sampling operation, thereby minimizing the overall computational effort of the network. Depending on the data and the desired accuracy, the system is deepened by repeating the convolution-pooling sequences multiple times. In this way, more high dimensional features are extracted from the input data followed by one or several fully-connected layers that are used for classification. Various C++/Python-based frameworks and platforms (Pouyanfar *et al.* 2018), including *TensorFlow*, *PyTorch*, *Caffe*, *Theano*, and *Keras*, are currently available to execute these tasks.

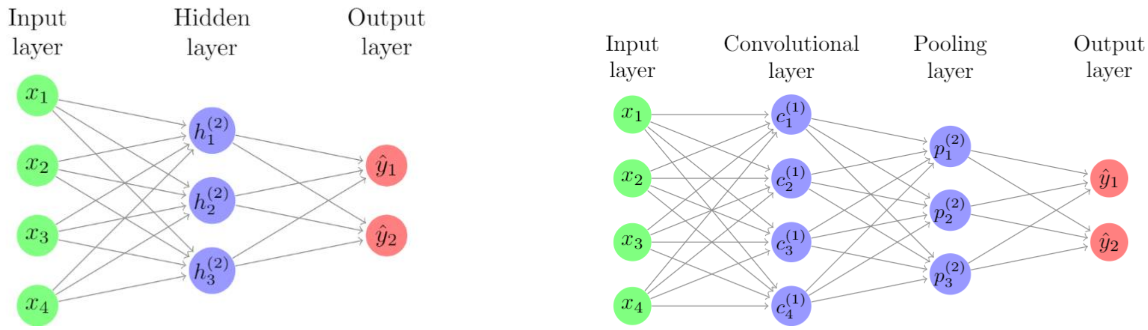
Combined with advances in GPUs and parallel computing, CNNs are a key technology underlying new developments in automated driving and facial recognition. CNNs are trained using a backpropagation algorithm, which combines the chain rule with the principles of dynamic programming. In a traditional neural network (NN), the full connections between the layers lead to time-intensive computations and overfitting of parameters (Abiodun *et al.* 2018). Unlike NN, a CNN convolves by using particular layers and avoids general multiplications, thereby keeping computations faster. CNN passes the input images through many deep layers (Gu *et al.* 2017; Yao *et al.* 2019) such as convolutional, pooling, and activation layers for feature extraction and performs classification using fully connected layers with a non-linear classifier (e.g., a *Softmax* classifier). CNN attempts to extract features by alternating and stacking convolutional kernels and pooling tasks. It tries to find features that best describe the input images with a varying number of deep layers. A rectified linear unit (ReLU) is often used as a non-linear activation function to introduce non-linearity in one or more of these layers on CNN. Auxiliary layers such as *dropout* layers are also used to prevent overfitting on CNN.

Convolutional layers take an input image and convolve it with a filter or kernel, where the size of the kernel matrix is much smaller than the size of the input matrix. The matrix multiplication of convolutional layers reduces the number of weights, which reduces the variance of the model. Convolutions generate invariant local features; at a lower level, filters can be used to detect edges in the image, whereas at a higher level, they can detect more complex shapes and objects that are critical for classifying an image. A convolutional layer is a set of image filters with learnable weights and plays an important role in CNN as a feature extractor.

On the other hand, pooling layers reduce the size of the layer while reducing the number of neurons in networks and extracting the most significant features with fixed-length over sliding windows of the raw input data. The reduction in the number of neurons is carried out by sliding a fixed window across a layer and choosing one value that effectively represents all the units captured by the window. Max-pooling and average-pooling are two common implementations of pooling. In max-pooling, the representative value becomes the largest of all units in the window, whereas, in average-pooling, the representative value becomes the average of all units in the window. A max-pooling layer is mostly used to down-sample the filtered weights from the convolutional layer, reducing computational costs and the probability of overfitting.

A fully connected layer has the shape of a flattened vector and plays an active role as a connector between the two-dimensional convolutional layer and the one-dimensional *Softmax* layer. The *Softmax* layer takes features from the fully connected layer, calculates the probabilities of each class using a normalized exponential function, and outputs the class with the highest probability as the classification result. By passing the images through various layers, a large number of parameters at various layers are optimally tuned and can extract salient features from the training images. In general, the training process varies from a few hours to a couple of days, depending on the network and hardware configurations, the training images, and the learning rate.

Both ordinary NNs and CNNs are feedforward neural networks and are generally trained using backpropagation. The primary difference between NNs and CNNs is the difference in the layers they use to classify images. Figure 1 shows the schematics of a typical NN and CNN architecture. The NN uses hidden layers (denoted as h), whereas CNN uses convolutional (denoted as c) and pooling layers (denoted as p) along with input and output layers. The number of layers depends on the architecture, the data, and the performance required from the model. One of the most critical issues with NNs is overfitting. Large neural nets trained on relatively small datasets can over-fit the training data. Unlike NNs, CNNs are not prone to overfitting due to a reduction in weights and the number of neurons caused by the convolutional layer and pooling layer, respectively. The difference between NN and CNN can be understood using an example of an image. Consider an image of $W * H * 3$ (over three channels, red, blue, and green), where W and H denote the width and height of the image matrix, respectively. An ordinary NN will take the image as the input, pass it through fully connected layers and non-linearities, and finally output a vector of probabilities for each class. The fully connected layer is so named because each of the input neurons n_i is connected to each output neuron n_o. If the number of input neurons is assumed to equal to the number of output neurons, the resulting number of weights becomes considerably large ($n_i * n_o$). In the framework of image classification, it is computationally expensive to train such a network, and it also gives rise to high variance. CNNs are a neural network with a different architecture that significantly reduces the number of weights and, thereby, the variance of the model.



(a) (b)

Figure 1. Schematic of (a) a typical NN and (b) a typical CNN with convolutional and pooling layers.

3.1 CNN Architectures

LeNet (LeCun *et al.* 1998) was originally developed to classify low-resolution images such as handwritten alphanumeric characters. *AlexNet* (Krizhevsky *et al.* 2012), a popular ImageNet CNN model, was developed by researchers from the University of Toronto and used convolutional filters of varying sizes, where the first layer had 11*11 convolution filters. The authors were the first to use rectified linear units (ReLU). Several layers of convolution and max-pooling were used with around 60 million weights, and the model was trained on 2 GPUs. The Visual Geometry Group, *VGGNet* (Simonyan and Zisserman 2014), was developed by researchers from Oxford University and only used 3*3 convolutional filters. Conv-Conv-Conv-pool layers were stacked together, followed by fully connected layers at the end. This research showed how the depth of CNN influences the accuracy of image reconstruction.

GoogleNet (Szegedy *et al.* 2014) was a deeper network, containing 22 layers with more computational efficiency, and did not have any fully connected layers. There were around 5 million parameters in the model. The network was composed of stacked sub-networks called inception modules. It had a naïve inception module that ran convolutional layers in parallel and concatenated the filters together. Moreover, it had a dimensionality reduction inception module that performed 1*1 convolutions, thereby achieving dimensionality reduction. The reduction lowered the computational cost and made the network computationally efficient by stacking multiple inception modules together. *ResNet* (He *et al.* 2015) was deeper than *GoogleNet* with 152 layers, where each layer in the residual block was implemented as a 3*3 convolution.

The development of newer CNN architectures evidenced a trend towards using more and more layers (i.e., a deeper architecture). Using these architectures for structural damage classification is valid only if a large amount of damage data is available. Moreover, the issue of overfitting may arise, and the outcome of high-performing CNNs will not generalize the results for civil engineering applications.

4. Review of CNN-Based SHM Literature

Primarily originated for object recognition, 2D CNN algorithms were mostly explored for 2D images in various SHM applications to detect defects and anomalies autonomously. Moreover, for vibration-based SHM, the researchers attempted to reshape the vibration signal into images by transforming the signal in frequency and time-frequency (TF) domain and used the resulting TF maps as the images in 2D CNN. However, the images involve significant complexity in choosing a large number of labeled data and layers and are not suitable for real-time SHM applications using mobile or handheld devices. To alleviate this problem, 1D CNN was recently introduced such that a time-history of vibration signal can be directly fed into CNN, which requires simple array operations, thereby demanding a shallow architecture with a fewer number of hidden layers (Kiranyaz *et al.* 2019).

Figure 2 shows a flowchart of the state-of-the-art CNN-based SHM literature that leads to significant advancement in this topic in the last few years. The schematic presents the two stages: data acquisition and condition assessment stage. The data acquisition stage is central to understand which type of data is apt for a particular structure. The data preparation precedes the data acquisition stage, depending on the classification or prediction task required from a specific application. Specific CNN architecture is selected next, followed by their further improvement using hyperparameter tuning. Once this step is accomplished, various infrastructure monitoring tasks are achieved in the last stage, demonstrating the novel contributions of the state-of-the-art CNN-based SHM techniques. A detailed systematic review of CNN-based SHM is organized by classifying the current literature into multiple classes, as illustrated below.

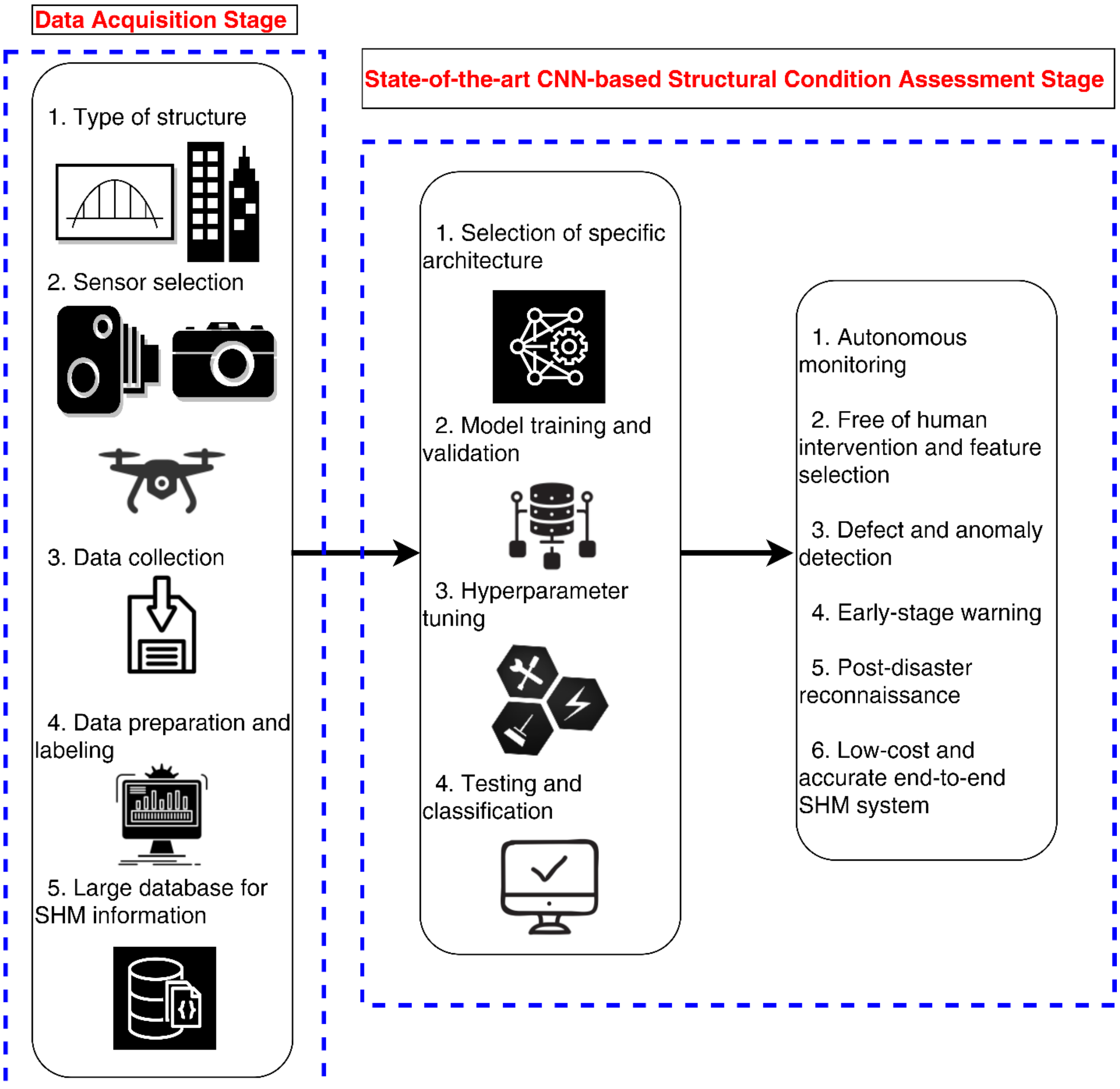


Figure 2. A schematic of the state-of-the-art CNN-based SHM operations.

4.1 Bridge health monitoring

The bridge infrastructure is critical for transportation and requires continuous monitoring. The critical components of any bridge that are prone to damage are used to acquire data in the form of an acceleration time-history, images, or continuous video streams. Deep learning methods such as CNN, FCN, or R-CNN are used to identify, classify, and quantify the damage. Guo *et al.* (2014) explored a sparse coding-based CNN algorithm with wireless sensors for efficient bridge SHM.

Sparse coding was used as an unsupervised layer for unlabelled data to learn high-level features from acceleration data. Various levels of damage cases were considered for a three-span bridge that was instrumented using wireless sensors. The proposed method was compared with other methods such as logistic regression and decision trees, and the proposed method was shown to outperform other methods with an accuracy of 98%. Gulgec *et al.* (2017) proposed a methodology for structural damage identification using CNN. Numerous undamaged and single-damaged samples of a steel gusset plate connection created in ABAQUS with varying uniformly distributed loads were developed to train, validate, and test the algorithm. Moreover, 50 network configurations with various hyper-parameters were tested over several epochs to determine the optimal CNN parameters.

A multiscale CNN was developed by Narazaki *et al.* (2017) to extract damage to various bridge components from image-based data. Post-processing methodologies such as super-pixel averaging and conditional random field optimization were implemented to enhance the accuracy of the multiscale CNN. The proposed CNN network was developed from a *ResNet* made up of 22 layers that computed the *Softmax* probabilities corresponding to ten scene components. The pixel-wise accuracy was calculated to be only 78.94% for this methodology, suggesting a strong dependence on the quality of super-pixel segmentation with regards to the boundary segmentation of components. An ensemble framework combining a couple of sparse coding algorithms and a CNN was proposed by Fallahian *et al.* (2018) for structural damage assessment under varying temperature effects. Features extracted from the frequency response function of the measured data were fed into a CNN and a couple of sparse coding algorithms to develop the classifier. Stochastic gradient descent was used in CNN to assign weights, and a *Softmax* function as an activation function. The proposed method was validated using a numerical truss bridge and a full-scale bridge. However, there are various types of bridges, and for continuous and autonomous monitoring, the identification of various bridge types is critical along with that of multiple damage types.

Zhao *et al.* (2018) explored CNN for maintenance and inspection of bridges. For bridge classification, an *AlexNet*-based CNN was trained first with more than 3800 images of various bridges. For recognition of bridge components, a *ZF-Net*-based faster R-CNN was trained with 600 bridge images. To detect cracks, a *GoogleNet*-based CNN was trained with 60000 cracked and un-cracked images. Accuracies of 96.6% for bridge classification, 90.45% for bridge component classification, and 99.36% for crack detection during testing were achieved. An image-based approach was proposed by Liang (2018) for holistic post-disaster inspection of reinforced concrete bridges using a DL encompassing system level, a component level, and local damage detection. Algorithmically, the network was made up of a *VGG-16* TL-based NN with Bayesian optimization for classification, a faster R-CNN for component detection, and a fully deep CNN for semantic damage segmentation. In a similar order, Kim *et al.* (2018) explored the application of regions with CNN (R-CNN)-based TL to identify cracks in a concrete bridge that were monitored using a UAV. Data containing 50000 images of 32×32 pixels from *ImageNet* and *Cifar-10* were used to train and classify the data. Max pooling and ReLU layers were used along with the convolutional layer in a sliding window-based CNN. The total length and thickness of cracks were also computed using a planar marker and automatically visualized on the inspection map.

Bao *et al.* (2019) presented computer-vision and DL-based structural anomaly detection to achieve automated SHM. Stacked AE and greedy layer-wise training techniques were used to train the DL networks. The acceleration data from a long-span bridge were first converted into images that were then transformed into grayscale image vectors for training a DNN considering six different anomalies such as missing, minor, outlier, square, drift, and trend data points. Recently, Xu *et al.* (2019) proposed fusion CNN for multilevel and multiscale damage identification in steel box girders without any prior assumptions of crack geometry. The proposed CNN architecture consisted of several layers of convolution, batch normalization, ReLU, max pooling, and *Softmax*, and was implemented using *MatConvNet*. Each image containing one or more cracks, handwriting, and background noise was acquired using a consumer-grade camera that was used for training and validation. The authors showed that fusion CNN worked better than general CNN, with an accuracy of 96.38%. However, its performance was limited to a specific object distance and the focal length of the camera.

Recently, Ni *et al.* (2019) proposed a 1D CNN-based technique in combination with autoencoder data compression for anomaly detection in a long-span suspension bridge. An accuracy of 97.53% was achieved with a compression ratio of 0.1. Similarly, Azmi and Pekcan (2019) proposed a CNN-TL-based SHM technique for damage identification in highly compressed data. A four-story numerical quarter-scale IASC-ASCE SHM model was used for numerical verification, and the proposed model was also validated on experimental studies using the IASC-ASCE SHM benchmark building and the Qatar University Grandstand Simulator. A mean accuracy of 90-100%

was achieved using the proposed model. 1D CNN was also used in a further study by Zhang *et al.* (2019) to detect changes in stiffness and mass. Three structural assemblages, a T-shaped steel beam, a short steel girder bridge, and a long steel girder bridge, were used, and accuracies of 99.79%, 99.36%, and 97.23% were achieved.

4.2 Pavement condition monitoring

Pavements are highly susceptible to damage due to high traffic and extreme weather conditions. The dataset usually consists of images acquired from a dashboard camera or a UAV. Cha *et al.* (2017) introduced a vision-based methodology for detecting cracks in concrete structures using CNN. Using nearly 40,000 images of damaged and undamaged concrete generated from various structures, CNN was tested and validated with more than 97% accuracy. Zhang *et al.* (2017) proposed a pixel-level CNN to detect cracks on 3D pavement surfaces. The proposed CNN, "*CrackNet*", was made up of two fully connected layers, one convolutional layer, one 1 * 1 convolution layer, and one output layer. This network was more efficient than traditional CNNs because of the absence of pooling layers that downsized the output of previous layers. An automated crack-length detection algorithm was proposed for pavement by Tong *et al.* (2017) using a deep CNN. A database of 8000 images of cracked and non-cracked pavement was generated for training, 500 of which were randomly selected to act as the test database. In addition, the images were converted to a grey-scale *.bmp* format so that *k*-means clustering analysis could be used to extract the length and shape of each pavement crack accurately. A five-layer-deep CNN achieved an accuracy of 94.35% with a mean squared error of 0.2377 cm for crack lengths between 0 and 8 cm. In addition, it was concluded that image resolution and lighting conditions had minimal influence on the accuracy of the proposed crack detection method.

Another pavement crack detection approach was investigated by Gopalakrishnan *et al.* (2017, 2018) using TL-based deep CNN. By implementing a truncated *VGG-16* deep CNN pre-trained on the *ImageNet* database, image vectors were extracted to train various classifiers to compare their performance for crack detection. Fan *et al.* (2018) proposed CNN to detect pavement cracks from images acquired by an iPhone from pavements in Beijing, China. Millions of monochromatic and RGB image patches were used. It was demonstrated that the proposed methodology had a precision of approximately 92%, which was better than traditional ML techniques such as local thresholding, *CrackForest*, Canny, minimal path selection, and free-form anisotropy. Similarly, Maeda *et al.* (2018a,b) investigated the capabilities of CNN networks to detect road surface damage from smartphone images. A pavement image dataset of 9,053 images captured using a dashboard-mounted smartphone was annotated using 15,435 bounding boxes to distinguish various damage classes. By analyzing this dataset using two object detection methods, Single-Shot Multibox Detector (SSD) using *Inception V2* and SSD using *MobileNet*, the robustness of these algorithms was investigated. Although the recall value of longitudinal construction joints and rutting, bumps, potholes, and separation was relatively low due to the small size of the training dataset, SSD *MobileNet* detected all damage classes with greater than 75% accuracy.

Fan *et al.* (2019) developed a novel FCN with an adaptive thresholding technique for image-based detection of road cracks. Initially, the FCN classified the images as either positive or negative based on the presence of cracks. The positive images were segmented, and an adaptive threshold technique that minimized the within-cluster sum of squares was used to localize the defects. The study used 40,000 RGB images from training, validation, and testing. The proposed methodology exhibited a precision of 99.92% and 98.70% for classification and pixel-level determination of pavement cracks. In another study, Zhang *et al.* (2018) proposed a novel algorithm to classify sealed and unsealed cracks in asphalt pavement using a TL-based deep CNN. The proposed methodology consisted of three components: (a) the images were initially enhanced to eliminate imbalance from illumination, (b) the images were classified as cracks, sealed cracks, or background images by means of a TL-based DCNN, and (c) fast block-wise segmentation and tensor voting curve detection were used to locate and extract those pixels that were considered cracked or sealed. It was concluded that the proposed method showed superior performance in both the classification and detection of sealed and unsealed pavement cracks.

Another DL algorithm was developed through TL for automated crack detection on concrete surfaces (Kim and Cho (2018)). Initially, a database of 50,000 images was created using the commercial scraper, "*ScrapeBox*", and various data augmentation techniques. By means of TL, a modified network for multiple object detection, "*AlexNet*", was used to train the proposed CNN classifier to identify uncracked pavement, cracks, and single or multiple edges or joints. By defining "crack-like" classes such as edges and joints, the number of false positives was significantly reduced.

4.3 Inspection of underground structures

Underground structures such as sewer pipes and tunnels are inaccessible for inspection. The underground structures are monitored using videos in combination with deep learning techniques. Stentoumis *et al.* (2016) presented CNN-based vision techniques to reconstruct 3D cracks with the aid of a stereo matching and optimization scheme using data acquired from a tunnel by a DSLR camera. A multilevel perceptron CNN was used as a classifier. The proposed method was also compared with various ML techniques such as *k*NN and SVM. The proposed CNN was shown to outperform other methods, with an accuracy of 88.6%. Similarly, Cheng and Wang (2018) evaluated sewer pipe defects through images acquired from closed-circuit television using faster region-based CNN (faster R-CNN). The R-CNN architecture works based on a region proposal network that can generate region proposals with different aspect ratios and scales to differentiate foreground and background noise to localize an anomaly compared to the undamaged section of a region of 3000 images. Doulamis *et al.* (2018) proposed a combined CNN and fuzzy spectral clustering approach for real-time crack detection in tunnels. An autonomous robotic system consisting of a robotic vehicle and a robot arm was used to capture imagery along the tunnel. To analyze complex concrete tunnel images, CNN was first used to capture specific regions of damage, followed by fuzzy clustering to exploit the spatial and orientation coherence of the cracks. It was concluded that the accuracy of crack prediction was relatively low due to limited visibility in the tunnel.

The capabilities of region-based FCN were explored by Xue and Li (2018) for shield tunnel lining defects. The proposed FCN consisted of a backbone convolutional layer and a pooling layer along with a *Softmax* layer and bounding box regression. A dataset containing a total of 4139 images of 3000×3724 pixels each were acquired using a movable tunnel inspection system consisting of several CCD cameras and LEDs as a source of light. The proposed method outperformed *AlexNet* and *GoogleNet* and achieved an accuracy of 96% while performing both object detection and image classification. Recently, Feng *et al.* (2019) developed a TL based on the *Inception-v3* DL algorithm to perform multiple damage type classification for hydro-junction infrastructure. The existing structure of the *Inception-v3* algorithm was modified so that the final layer had five fully connected neurons to increase the accuracy of labeling each damage type. In another study (Kang *et al.* 2020), a basic pursuit-based background filtering algorithm was proposed to improve the visibility of underground objects (e.g., cavities, manholes, and pipes), followed by DCNN using three-dimensional ground-penetrating radar data from urban roads in Korea.

4.4 Building condition assessment

Tall buildings and historical structures pose a challenge for manual inspection and require an accessible way for autonomous monitoring. Chaiyasarn *et al.* (2018) proposed an integrated algorithm combining CNN with classification models such as SVM and random forest for crack detection in historic structures. The data consisted of images from masonry structures containing cracks that were acquired using a digital camera and an unmanned aerial vehicle (UAV). It was shown that CNN with SVM outperformed conventional CNN based on the *Softmax* classifier. Similarly, Yuem *et al.* (2018) used CNN for image classification after post-event (e.g., earthquake, hurricane, tornado, or others) building reconnaissance. The dataset of 90000 colored structural images was used to train the network for scene classification and object detection. All the images were manually labeled using in-house annotation software before the CNN training phase.

To classify various common types of building damage, Perez *et al.* (2019) explored the possibility of detecting common building defects caused by dampness, such as mold, deterioration, and staining through images using CNN. The proposed model was trained using the *VGG-16* (*ResNet-50*) CNN classifier, and class activation mapping was used for object localization. The CNN architecture contained five blocks of convolutional layers with max-pooling for feature extraction. The proposed methodology achieved an overall accuracy of 87.50% and classified multiclass defects using a small dataset. Recently, Jiang and Zhang (2019) used a wall-climbing unmanned aerial system (UAS) to acquire real-time video. The video data were then converted to 1330 crack images, and a CNN was trained. The images were transferred to an Android platform through a wireless data link. An accuracy of 94.48% was achieved using the proposed model.

4.5 Multi-class structural monitoring

Structures experience multiple types of damage, and identifying all of them at once is a faster approach to repair and maintenance. A vision-based multiscale pixel-wise deep CNN network was proposed by Hoskere *et al.* (2017) to detect six types of structural damage. The proposed methodology consisted of two parallel steps: (a) a damage classifier to separate each pixel into predefined classes and (b) a damage segmenter that distinguished damaged pixels from undamaged ones. By implementing 1695 images of over 250 structures, the authors concluded that *ResNet23* and *VGG-19* were the most accurate segmenter and classifier, with accuracies of 88.8% and 71.4%, respectively. Moreover, by combining the segmenter and classifier networks using *Softmax* thresholds, the accuracy across all classes was increased from 71.4% to 86.7%. Lin and Nie (2017) used a CNN with batch normalization to extract and localize structural damage in a simply supported Euler-Bernoulli beam. Numerical simulations were conducted with various damage locations and conditions to generate a dataset of 6,885 measurements. The proposed methodology was compared with a wavelet packet transform approach for both noiseless and noisy single- and multi-damage scenarios. Overall, CNN resulted in superior performance over the wavelet packet transform for single and multiple structural damage sites.

Atha and Jahanshahi (2018) evaluated corrosion detection using three proposed CNN architectures, *VGG-15*, *Corrosion5*, and *Corrosion7*. A comparison is presented with the other two state-of-the-art CNN architectures, *VGG-16*, and *ZF-Net*. An approach containing non-overlapping sliding windows was used to isolate the corroded region within each image. The authors investigated the performance of the proposed architecture under various sizes of sliding windows and color spaces. Using two specific properties of CNN (parameter sharing and local connectivity), Khodabandehlou *et al.* (2018) proposed a CNN method that used a reduced number of parameters, hence requiring limited training data for SHM. Behrouzi and Pantoza (2018) used a DL algorithm to identify damage patterns from tagged images of roadways and railways after large seismic events. The authors claimed that the proposed method correctly identified 92% of the roadway images, where 80% of railways were affected by the earthquake. Cha and Kang (2018) carried out damage identification by means of CNN using ultrasonic beacons by geo-tagging a video stream obtained from a UAV. A deep CNN with a sliding window was used as a DL architecture, with ReLU as an activation function and a *Softmax* function as a classifier.

Similarly, Patterson *et al.* (2018) used DL techniques for seismic damage image classification and developed a user-friendly graphic user interface wrapper where *AlexNet* and *ResNet* were used in the pre-trained DL model. Pan *et al.* (2018) evaluated the efficacy of DBN using multiple restricted Boltzmann machines for structural condition assessment to enable timely decision-making for maintenance. A 1D CNN was proposed by Abdeljaber *et al.* (2018) for structural damage detection on an SHM benchmark dataset. Although CNNs are primarily used for 2D signals such as images and videos, the authors used the *tanh* activation function to learn from 1D raw acceleration data and proposed an enhanced adaptive CNN to identify global structural damage in structures. Images acquired using smartphones and UAVs are viable and inexpensive options for acquiring damaged data from structures. Li and Zhao (2018) evaluated CNN for crack detection on a real concrete surface using cropped images taken from a smartphone. A CNN with binary outputs of the cracked or uncracked concrete surface was used to train *GoogleNet*. A total of 60000 images with 256 by 256 pixels each were used to classify cracked concrete surfaces with an accuracy of 99.39%. An application called *Crack Detector* was developed and installed in a smartphone to detect cracks in real-time.

Dorafshan *et al.* (2018a) explored the feasibility of using small off-the-shelf UAVs for inspection of concrete decks and buildings using CNNs. The proposed algorithm was first used to train the model using images acquired from a laboratory-scale bridge deck with a low-resolution camera and achieved an accuracy of 94.7%. The proposed CNN was then used to investigate a building by means of transfer learning (TL) using *AlexNet* with an accuracy of 97.1%. Moreover, Cha *et al.* (2018) proposed an improved visual inspection method using a faster region-based CNN. The proposed method provided robust detection of multi-surface damage types such as concrete cracks, medium and high corrosion of steel, bolt corrosion, and steel delamination using a variable bounding box and was shown to be more efficient than the authors' previous work (Cha *et al.* 2017). Moreover, this technique showed promising results for the autonomous detection of structural defects from quasi-real-time video data. On the other hand, Dorafshan *et al.* (2018b) provided an excellent database for autonomous detection of cracks ranging from 0.06 to 25 mm using CNN on a concrete surface. Spatial- and frequency-domain edge detection methodologies were compared by the same authors (Dorafshan *et al.* 2018c) using DCNN to detect cracks in concrete structures. It was concluded that *AlexNet* could detect smaller cracks (86%) more accurately than Laplacian-of-Gaussian (LoG). Moreover, the authors proposed a hybrid methodology that implemented a CNN to categorize images based on the presence of damage, after which those damaged images were further refined at the pixel level by the LoG edge detection technique.

Hoskere *et al.* (2018) explored FCN with residual network architecture for automated post-earthquake image classification. The FCN was capable of semantic segmentation and classification and was combined with a 3D mesh model of the structure for damage representation in building components. The dataset used to train the FCN included 1000 images of 288 by 288 pixels each and was acquired from post-disaster reconnaissance surveys using a UAV. An accuracy of 91.1% was achieved for damage type identification along with information of structural and non-structural components. Moreover, Rui *et al.* (2019) developed a two-stage CNN to detect and classify defects in narrow overlap welds. Time-series signals from eddy current testing of defective welds were initially converted to 2D diagrams using a continuous wavelet transform. Before the initial data transformation, the 2D diagrams were entered into a two-step CNN network that (a) identified the presence of defects using binary classification and (b) upon detecting defects, further classified them into five defect types. Although both single-step and two-step CNNs had similar accuracy of approximately 97%, the faster computational time of the two-step method made it more efficient.

Recently, Deng *et al.* (2019) implemented a faster R-CNN to detect handwritten scripts and cracks in concrete surfaces. A modified 21-layer *ZF-Net* consisting of three neurons to classify background, cracks, and handwriting was trained using a 20% subset of the authors' generated database of nearly 5000 sub-images. By investigating the influence of handwriting scripts on crack detection, it was concluded that including handwriting scripts as a unique background class significantly increased the accuracy of classifying cracks in concrete surfaces. Furthermore, comparing the proposed methodology with the DL algorithm, '*You Only Look Once' (YoLo) v2*, showed superior performance, with significantly reduced percentages of false positives detected. Dung and Duc Anh (2019) proposed an FCN for segmented vision-based detection and density evaluation of surface cracks in concrete structures. TL was applied as the FCN encoder was based on the *VGG-16* CNN model because this model showed superior performance to *ResNet* and *Inception*. Upon training and validation using 500 images, the FCN was shown to have a max F1 score and average precision of approximately 90%.

Li *et al.* (2019) proposed an FCN to detect four concrete damage classes: cracks, spalling, efflorescence, and holes, from an established smartphone-based image database. The development of the FCN algorithm was based on TL of weights and biases provided by *DenseNet-121* for feature extraction. The algorithm was trained and validated using 2200 images. Compared to *SegNet*, the proposed methodology offered better performance in detecting various types of concrete damage. In another recent study, the authors (Mei and Gul 2020) used a depth-first search algorithm as a preprocessing tool to eliminate isolated pixels, followed by multilevel feature fusion and crack detection using images obtained from a smartphone.

4.6 Inspection of other large-scale structures

Large-scale structures are challenging to monitor, and image-based monitoring techniques provide a powerful tool for effective structural monitoring. CNN was implemented to detect surface defects in rails from photometric stereo images acquired in a dark-field setup by Soukup and Huber-Mork (2014). The setup of various light sources at different oblique angles in the dark-field identified the location of cavities through a scattering of applied light. Comparing traditional model-based approaches to the trained CNN, the authors found a significant reduction in a detection error. Furthermore, regularization methods such as training data augmentation and unsupervised layer-wise pre-training were shown to reduce the probability of overfitting due to the size of the available image dataset. Abdeljaber *et al.* (2017) proposed a nonparametric 1D CNN to extract structural damage from the time-histories of vibration-based responses. In this method, the acceleration at each sensor location was first divided into several frames, each containing a finite number of samples, and then each frame was normalized and fed into a CNN. The probability of damage was then computed to quantify the severity of damage and isolate the damage location. The proposed methodology showed efficient processing of the measured data compared to existing ML techniques, which required significant pre- and post-processing and feature extraction. A laboratory stadium developed in the Qatar University Grandstand Simulator was used to validate the accuracy of the proposed method.

Pan *et al.* (2018) evaluated the efficacy of DBN using multiple restricted Boltzmann machines for structural health assessment to enable timely decision-making for maintenance. Lin *et al.* (2018) compared CNN with SVM for damage assessment in a three-story laboratory model and concluded that DL methods had less noise sensitivity than shallow learning methods. Chen and Jahanshahi (2018) proposed a CNN method with a naïve Bayes data fusion scheme to detect tiny cracks on metallic surfaces from video data for nuclear inspection applications. This methodology was distinct from previous CNNs because it collected image data from multiple video frames to improve crack localization while using a naïve Bayes decision process to reduce false negatives. Through testing and training of approximately 300,000 images extracted from video frames, it was concluded that this methodology achieved an accuracy of 98.3%, showing significant improvement compared to state-of-the-art ML algorithms.

Recently, Dick *et al.* (2019) investigated the use of DL algorithms to inspect critical electric utility infrastructure. Through TL on CNN, images of utility infrastructure from vehicular-mounted cameras were classified into five categories: highways, pine trees, fields, trucks, and power infrastructures. This technique provided automatic detection of vegetation, which was considered a major hazard to power infrastructure. Hoskere *et al.* (2019) proposed deep Bayesian NNs for damage localization in gates of navigation locks. In this proposed research, Monte Carlo dropout was used to increase the accuracy of the trained network and determine the sensitivity of measured strain to damage. Three CNN models were recently tested by Xu *et al.* (2019) to identify cracks in wind turbine blades. In another study (Zhang *et al.* 2020), the authors implemented a faster region-based CNN to detect bolt loosening under different operating conditions such as measurement angle, lighting condition, and vibration condition.

5. Improved CNN methods in SHM

Depending on the complexity of damage and its location in large-scale structures, the SHM community recently implemented several advanced CNN architectures to train these complex models. Some of these newer architectures include fully convolutional networks (FCNs) and transfer learning (TL).

5.1 Fully Convolutional Networks (FCNs)

Yang *et al.* (2018) proposed a novel FCN for pixel-level crack detection. This method consisted of both down-sampling using a *VGG16* network and up-sampling techniques, creating a robust model that could analyze multiscale images. Future improvements to increase performance for the detection of thin cracks, intersections, and border cracks were suggested to increase the accuracy of proposed networks to that of existing state-of-the-art DL algorithms. Hoskere *et al.* (2018) explored FCN with residual network architecture for automated post-earthquake image classification. The FCN was capable of semantic segmentation and classification and was combined with a 3D mesh model of the structure for damage representation in building components. The dataset used for training the FCN included 1000 images of 288 by 288 pixels each and was acquired from post-disaster reconnaissance surveys using a UAV.

The capabilities of region-based FCN were explored by Xue and Li (2018) for shield tunnel lining defects. The proposed FCN consisted of a backbone convolutional layer, a pooling layer, a *Softmax* layer, and bounding box regression. A dataset of 4139 images of 3000×3724 pixels each were acquired using a movable tunnel inspection system consisting of several CCD cameras and LEDs as a source of light. The proposed method outperformed *AlexNet* and *GoogleNet* and achieved an accuracy of 96% while performing both object detection and image classification. Dung and Duc Anh (2019) proposed an FCN for segmented vision-based detection and density evaluation of surface cracks in concrete structures. Fan *et al.* (2019) developed a novel FCN with an adaptive thresholding technique for image-based detection of road cracks. Initially, the FCN classified the images as either positive or negative based on the presence of cracks. These positive images were then segmented, and an adaptive threshold technique that minimized the within-cluster sum of squares was used to localize the defects.

Li *et al.* (2019) proposed an FCN to detect four concrete damage classes: cracks, spalling, efflorescence, and holes, from an established smartphone-based image database. The development of the FCN algorithm was based on TL of weights and biases provided by *DenseNet-121* for feature extraction. The algorithm was trained and validated using 2200 images. Compared to *SegNet*, the proposed methodology offered better performance in detecting various types of concrete damage. An FCN was developed by Rubio *et al.* (2019) to detect delamination and rebar exposure in reinforced concrete bridges. The authors considered a multi-labeled approach for the dataset in which different regions of the images were considered ground truth, uncertain, or penalized depending on the agreement of the various annotators that classified them. This methodology had a mean accuracy of 89.7% and 78.4% for delamination and rebar exposure, meaning that this model could be used as a step towards automating bridge inspection.

5.2 CNN with Transfer Learning

Feng *et al.* (2017) proposed an active learning algorithm for automatic detection and classification of cracks, deposits, and water leakage from concrete structures without requiring time-consuming labelling. The classification and detection of these defects were performed by a deep residual network (*ResNet*). Using the active learning network, the classifiers were continuously retrained with new annotated images, achieving a significant reduction in manual human-based image annotation and labeling. Using a positive-sampling technique, the authors obtained an accuracy of 87.5% for 235,200 image patches. Another pavement crack detection approach was investigated by Gopalakrishnan *et al.* (2017, 2018) using TL-based deep CNN. By implementing a truncated

VGG-16 deep CNN pre-trained on the *ImageNet* database, image vectors were extracted to train various classifiers to compare their performance for crack detection. Kim *et al.* (2018) explored the application of regions with CNN (R-CNN)-based TL to identify cracks in a concrete bridge that was monitored using a UAV. Data containing 50000 images of 32×32 pixels each from *ImageNet* and *Cifar-10* was used to train on the data, followed by classification. Max pooling and ReLU layers were used along with a convolutional layer in the sliding window-based CNN. The total length and thickness of cracks were also computed using a planar marker and were automatically visualized on an inspection map.

In another recent study, Gao and Mosalam (2018) developed a *Structural ImageNet* to detect various types of post-disaster damage using a modified TL-based *VGG-16* network. The robustness of detecting four pre-defined features: (1) component type, (2) spalling condition, (3) damage level, and (4) damage type was investigated using feature extraction and fine-tuning of the TL technique. Parametric studies were conducted to determine the optimal image size to reduce computational complexity while retaining valuable information. Moreover, complexities in the four-class damage-level features resulted in decreased accuracy (68%) and increased overfitting (23%), suggesting that this model may be a baseline for future research into *Structural ImageNet*. Zhang *et al.* (2018) proposed a novel algorithm to classify sealed and unsealed cracks in asphalt pavement using a TL-based deep CNN. The proposed methodology consisted of three components: (a) the images were initially enhanced to eliminate imbalance with illumination, (b) images were classified as unsealed cracks, sealed cracks, or background images by means of a TL-based DCNN, and (c) fast block-wise segmentation and tensor voting curve detection were used to locate and extract those pixels that were considered cracked or sealed. It was concluded that the proposed method showed superior performance for both the classification and detection of sealed and unsealed pavement cracks compared to other image processing methods. Another DL algorithm was developed through TL for the automated detection of cracks on a concrete surface (Kim and Cho 2018). Initially, a database of 50,000 images was created using the commercial scraper, "*ScrapeBox*", and various data augmentation techniques. By means of TL, a modified network for multiple object detection, "*AlexNet*", was used to train the proposed CNN classifier to identify non-cracks, cracks, and single or multiple edges or joints. By defining "crack-like" classes such as edges and joints, the number of false positives was significantly reduced.

Recently, Feng *et al.* (2019) developed a TL based on the *Inception-v3* DL algorithm to detect multiple damage classifications for hydro-junction infrastructure. The existing structure of the *Inception-v3* algorithm was modified so that the final layer had five fully connected neurons to increase the accuracy of labeling each damage type. Kim and Sim (2019) addressed the automation of operational modal analysis by developing a faster R-CNN for automated extraction of peaks from frequency-domain image data. Faster R-CNNs such as the *VGGNet* and *ZF-Net* implemented in this study used region proposal networks (RPNs) to generate rectangular object regions through the shared convolutional features of fast R-CNN networks. The network was trained using 15,596 peaks extracted from a multiple-degree-of-freedom numerical model. Upon comparison with time domain-based methods for peak extraction, it was found that the proposed method had superior performance to F1 scores and computational time.

6. Comprehensive Summary of the Reviewed Literature

As shown in Sections 4-5, structural condition assessment involves major tasks such as system identification, damage identification, crack, and anomaly detection. The accuracy of these tasks strongly depends on sensor placement and presence of sensor faults, fluctuations in environmental and operational conditions, the suitability of appropriate features and feature extraction methods such as time-, frequency-, time-frequency methods (Qarib and Adeli 2016; Sadhu *et al.* 2019; Barbosh *et al.* 2020; Kankanamge *et al.* 2020), image processing (Mohan and Poobal 2018) and other ML techniques (Sun *et al.* 2020). Therefore, the conventional ML-based SHM strategies strongly rely on expert knowledge to design the most appropriate features for a given data of critical infrastructure. Unlike the traditional approaches, CNN undertakes similar tasks without requiring any feature selection stage. It relies on a large database of training data and builds a deep network with a suite of network and training parameters, implicitly performing both feature extraction and pattern classification. At one end, 1D CNN (Kiranyaz *et al.* 2020) uses structured information such as vibration or time-series data to perform global damage detection. On the other hand, 2D CNN has been explored to analyze unstructured data such as actual images or derived TF images (e.g., spectrograms or scalograms) of time-series to undertake local damage identification. Overall, CNN has achieved significant popularity in the SHM literature due to its requirement of having minimum knowledge of the best-suited features of a dataset. Table 2 finally provides a summary of the literature reviewed in Sections 4 and 5 with a systematic presentation of the specific application and data used for structural condition assessment.

Table 2: Summary of CNN-based structural condition assessment literature.

<table>
<tr><th>Reference</th><th>Application</th><th>CNN architecture</th><th>Specifics of data</th></tr>
<tr><td colspan="4"><u>Bridge health monitoring</u>
Merits:
1. A wide variety of data types includes sequential/time-series and visual-based images and videos, where both 1D and 2D CNNs have been equally effective.
2. The application of CNNs enables the identification of both global and local structural damage.
Drawbacks:
1. The sparse coding algorithm is often needed as a preprocessor for feature extraction in combination with CNNs to overcome the challenge of data labeling.
2. Vision-based data collection of independent bridge components is a challenging task; CNNs are used to train the classification based on scene segmentation and bridge component identification from a large-scale image.</td></tr>
<tr><td>Guo et al. (2014)</td><td>Global condition assessment</td><td>Inclusion of sparse coding in CNN</td><td>Acceleration time-histories</td></tr>
<tr><td>Gulgec et al. (2017)</td><td>Anomaly detection in steel gusset plate</td><td>CNN</td><td>Simulated strain measurements</td></tr>
<tr><td>Narazaki et al. (2017)</td><td>Global and component-level damage assessment</td><td>Multiscale CNN developed from a ResNet</td><td>Images of scene components</td></tr>
<tr><td>Fallalian et al. (2018)</td><td>Global condition assessment</td><td>Integration of coupled sparse coding in DNN</td><td>Simulated and experimental acceleration data</td></tr>
<tr><td>Zhao et al. (2018)</td><td>Component-level damage assessment</td><td>AlexNet, ZF-Net, and GoogleNet</td><td>Cracked and un-cracked images</td></tr>
</table>

Liang (2018)	Global and component-level damage assessment	*VGG16*, R-CNN, and fully deep CNN through semantic segmentation with Bayesian optimization	Cracked and un-cracked images of reinforced concrete bridges
Kim *et al*. (2018)	Component-level damage assessment	R-CNN-based TL (*ImageNet* and *Cifar10*)	Images from UAV
Bao *et al*. (2019)	Anomaly detection	DNN-stacked AE and greedy layer-wise training techniques	Acceleration data
Xu *et al*. (2019)	Damage assessment in steel box girders	FCNN implemented with *MatConvNet*	Images acquired from a consumer-grade camera
Rubio *et al*. (2019)	Component-level damage assessment	FCNs	Images
Ni *et al*. (2019)	Anomaly detection with data compression	1D CNN	Acceleration data
Azimi and Pekcan (2019)	Damage identification	CNN with TL	Acceleration data
Zhang *et al*. (2019)	Damage identification with changes in stiffness and mass	1D CNN	Acceleration data
Pavement condition monitoring			

Merits: 1. The image datasets can be acquired under varying environmental conditions. The data acquired is suitable for multiclass problems (e.g., identification of cracks, their sizes, and locations). 2. The crack length identification is carried out efficiently by increasing the subsampling between the convolution layers and creating a deep CNN. **Drawbacks:** 1. In the presence of noise and complicated cracks, the CNNs are supplemented with additional preprocessing such as bilateral filtering and adaptive thresholding. 2. The datasets often result in imbalance measurements. 3. In case of similar crack identification, such as open crack and sealed crack under noise is tackled using a special treatment such as TL and tensor voting-based crack detection.			
Cha *et al*. (2017)	Concrete surface	CNN with sliding window technique	Images from DSLR camera
Zhang *et al*. (2017)	Automated pavement crack detection	*CrackNet* in the absence of pooling layer	3D asphalt images
Tong *et al*. (2017)	Crack length detection	Deep CNN	Cracked and un-cracked RGB images
Gopalakrishnan *et al*. (2017,2018)	Pavement defects	*VGG16*, DCNN	Images acquired using UAV
Fan *et al*. (2018)	Crack size estimation	CNN	Monochromatic and RGB images from iPhone
Maeda *et al*. (2018a,b)	Anomaly detection on the road surface	CNN integrated with two object detection methods	Images acquired from a dashboard-mounted smartphone in a vehicle
Fan *et al*. (2019)	Road inspection	FCN with adaptive threshold technique	RGB images

<table>
<tr><td>Zhang et al. (2018)</td><td>Asphalt pavement</td><td>TL-based deep CNN</td><td>Images</td></tr>
<tr><td>Kim and Cho (2018)</td><td>Crack inspection in an onsite environment</td><td>TL integrated with AlexNet</td><td>Images and videos acquired from UAVs</td></tr>
<tr><td colspan="4"><u>Inspection of underground structures</u>

Merits:
1. Underground structures such as sewer and water pipes, tunnels, and heavy infrastructures such as hydropower dams are difficult to inspect due to their depth, and thickness using the traditional vibration-based SHM methods.
2. For extremely large, inaccessible structures such as hydro structures, UAVs with real-time kinematic global positioning system can be used for data collection and defect identification.
3. In the presence of sequential data such as radar data, CNNs perform better with de-noised signals.

Drawbacks:
1. Data acquisition from structures such as tunnels and sewer pipe require different approaches. For example, images from tunnels can be acquired using DSLR cameras and robotic vehicles; however, for sewer pipe, images are obtained from pre-installed closed-circuit cameras.
2. CNNs are also required to be combined with unsupervised clustering to refine the detected crack regions from noisy images exploiting spatial and orientation coherency in the presence of inadequate lighting conditions.
3. If the dataset is small, TL is applied for the enhancement of CNN damage classification performance.</td></tr>
<tr><td>Stentoumis et al. (2016)</td><td>Highway and railway tunnels</td><td>CNN connected with multilevel perceptron to build a 3D crack model</td><td>Images from DSLR camera</td></tr>
<tr><td>Cheng and Wang (2018)</td><td>Sewer pipe defects</td><td>Faster region-based CNN</td><td>Images acquired from closed-circuit television</td></tr>
<tr><td>Doulamis et al. (2018)</td><td>Tunnel inspection</td><td>CNN combined with fuzzy spectral clustering</td><td>Images obtained from a robotic vehicle</td></tr>
</table>

<table>
<tr><td>Xue and Li (2018)</td><td>Tunnel lining</td><td>Region-based FCN with Softmax layer and bounding box regression</td><td>Images from CCD camera</td></tr>
<tr><td>Feng et al. (2019)</td><td>Hydro infrastructure</td><td>Inception-V3 and TL</td><td>Images from a high-definition camera</td></tr>
<tr><td>Kang et al. (2020)</td><td>Underground cavity detection</td><td>CNN with a basic pursuit-based background algorithm</td><td>3D ground penetration radar data</td></tr>
<tr><td colspan="4"><u>Building condition assessment</u>
Merits:
1. Buildings are tall spatial structures that require condition assessment on internal and external components. The evaluation of external components, e.g., assessment of post-disaster nonstructural damages, is now possible with vision-based CNN methods. The datasets can be easily acquired using an inexpensive digital handheld camera, smartphones, and UAVs.
2. In many studies, apart from the crack or defect detection, the Class Activation Mapping layer is added to CNNs for object identification. The object localization is highly beneficial for the identification of damage in structural and nonstructural components.
Drawbacks:
1. CNNs are often reinforced with an additional 3D image stitching technique to analyze the structure in the 3D coordinate system.
2. The training database is often not enough; CNNs are required to pre-trained on benchmark models such as VGG16 or CrackNet.</td></tr>
<tr><td>Chaiyasarn et al. (2018)</td><td>Global condition assessment in historical masonry structures</td><td>CNN with SVM and random forest</td><td>Images from digital camera and UAV</td></tr>
</table>

Yuem *et al*. (2018)	Post-disaster building reconnaissance	CNN with in-house automation software to label images	Scene classification and object detection for damage classification
Perez *et al*. (2019)	Surface-level defects caused by mold, stain, and deterioration	*VGG16* and class activation mapping	Images acquired using a mobile phone and hand-held camera along with copyrighted images from Internet
Jiang and Zhang	Crack detection	CNN	Unmanned aerial system to acquire video and images
Multi-class structural monitoring **Merits:** 1. Offer autonomous monitoring systems and eliminate manual inspections that are time-consuming, labor-intensive, subjective, and often unsafe. 2. Allow rapid decision making for post-disaster damage assessment. 3. The proposed techniques are mostly insensitive to the measurement noise. **Drawbacks:** 1. Need further improvement to develop more robust multi-type damage classification techniques. 2. Significantly more layers would be required to distinguish between different types of complexities in structures, damage conditions, and background effects. 3. Few of these methods are heavily dependent on the results of the FE model as the real condition data are scarce. 4. Proper labeling of multiclass damages is always a challenge.			
Hoskere *et al*. (2017)	Post-earthquake multiclass structural inspection	Multiscale pixel-wise deep CNN	Various images of concrete and steel surfaces
Lin and Nie (2017)	Numerical simulation using a simply supported beam	CNN	Time-series data

Atha and Jahashahi (2018)	Corrosion detection on a metallic surface	*VGG15*, *Corrosion5*, and *Corrosion7* with non-overlapping sliding windows	Colour images
Khodabandehlou *et al*. (2018)	Vibration-based condition assessment	2D CNN	Acceleration time-histories
Behrouzi and Pantoza (2018)	Post-earthquake inspection	DL network	Tagged images of roadways and railways
Kang and Cha (2018)	Structural inspection where using GPS is not feasible	Deep CNN with sliding window	Geo-tagging of a video stream from a UAV
Patterson *et al*. (2018)	Seismic damage classification	*AlexNet* and *RestNet*	GUI wrapper
Abdeljaber *et al*. (2018)	SHM benchmark data	1-D adaptive CNN with (hyperbolic tangent) *tanh* activation function	Acceleration data
Li and Zhao (2018)	Concrete surface	*GoogleNet* (an app, *Crack Detector*, was developed)	Cropped images are taken from a smartphone
Dorafshan *et al*. (2018)	Component-level damage assessment in bridges and buildings	TL and *AlexNet* DCNN	Imaged from off-the-shelf UAV
Cha *et al*. (2018)	Multi-surface damages	Faster-R-CNN	Quasi-real-time video data

Dorafshan *et al.* (2018b, 2018c)	Concrete surface	CNN with LoG edge detection	Benchmark database with cracks ranging from 0.06 to 25 mm
Yang *et al*. (2018)	Pixel-level crack detection	FCN via *VGG16*	Multiscale images
Hoskere *et al*. (2018)	Post-earthquake inspection	FCN	Reconnaissance survey from a UAV
Rui *et al*. (2019)	Defective welds	Wavelet-assisted CNN with binary classification	Time-series data of eddy current
Deng *et al*. (2019)	Concrete surface	Faster R-CNN, *ZF-Net*, and *YoLo v2*	Images with handwritten scripts and cracks
Dung and Duc Anh (2019)	Surface cracks in concrete structures	*VGG16*	Images and video of crack data
Li *et al*. (2019)	Multiple concrete damage types	*DenseNet-121*-based FCN	Smartphone-based images
Mei and Gul (2020)	Pixel-level crack detection	DNN with depth-first search-based preprocessing	Smartphone-based images

<u>Inspection of other large-scale structures</u>

Merits:

1. Many algorithms showed robustness in different environmental conditions.

Drawbacks:

1. Noise interference could contaminate the data in large-scale structures; deeper neural networks could be used to solve this issue.
2. A large number of training data is needed to achieve data convergence and prevent overfitting.

Soukoup and Huber-Mork (2014)	Metal surface of rails	Unsupervised layer-wise pre-training.	Photometric stereo images
Abdeljaber *et al*. (2017)	Laboratory study	One-dimensional CNN	Acceleration time-histories
Feng *et al*. (2017)	Less time-consuming labelling operation	*ResNet* with active learning	Image dataset
Pan *et al*. (2018)	Experimental study	Deep Bayesian NN using multiple restricted Boltzmann machines	Acceleration data
Lin *et al*. (2018)	Laboratory studies	Comparison of CNN with SVM and other shallow learning methods	Acceleration data
Chen and Jahanshahi (2018)	Nuclear power plant	CNN with a naïve Bayes data fusion	Video data
Dick *et al*. (2019)	Electrical utility infrastructure	TL and CNN	Images from a vehicle-mounted camera
Hoskere *et al*. (2019)	Navigation infrastructure	Deep Bayesian NN	Finite element model-based simulated data and measured strain data
Xu *et al*. (2019)	Wind turbine blade	Three CNN models	Images from UAVs
Kim and Sim (2019)	Operational modal analysis	*VGGNet* and *ZF-Net*	Frequency peaks from simulated data.

Zhang *et al*. (2020)	Detection of bolt loosening using experimental study	Region-based CNN	Webcam data

7. Challenges for CNN Implementation in Structural Condition Assessment

With increasing computational capabilities in the era of big data, high-performance computing, parallel processing, and cloud computing, CNN techniques have witnessed significant developments in remote and autonomous SHM of critical civil infrastructure. 2D CNN has brought a radical shift in SHM using non-contact sensors and robotic devices. Whereas, 1D CNN, which is free of major matrix operations, has resulted in efficient classification and clustering of vibration-based SHM data, enabling its capabilities in low power real-time applications (e.g., smartphone or handheld device). The CNN techniques offer new advantages and opportunities that are systematically reviewed in this paper based on the ongoing research published in top-notch journals and conference papers. At one end, the state-of-the-art research offers remote and autonomous SHM systems for cost-effective and accurate structural inspection. On the other hand, it allows feature-free early-stage warning or post-disaster reconnaissance for the infrastructure owners and stakeholders, enhancing an end-to-end SHM system. However, the existing CNN-based literature presents several challenges that must be addressed in the upcoming years before this approach can be positioned as a generalized strategy for monitoring and maintenance of a wide range of infrastructure. The identified real-world challenges are illustrated below:

i) Data imbalance issue in large-scale infrastructure: CNN implicitly adopts a deep network depending on the complexity of the data. Unlike systems in other engineering domains, civil infrastructure is large in size and composed of decades of design life. Due to such size and life-span, structural condition data obtained from limited sparse measurements have a wide variety of damage states (Sun *et al.* 2020), causing data imbalance issue in SHM. Although the researchers have proposed various data augmentation techniques to alleviate the over-fitting caused by the data imbalance, it remains a significant challenge to the SHM community (Gopalakrishnan *et al.* 2017, 2018; Liang 2018; Kim *et al.* 2018; Zhang *et al.* 2018), unlike in other engineering domain.

Moreover, acquiring a large number of images with a wide variety of historical damage events forms another hindrance to developing a training database, which limits the applicability of CNN in structural condition assessment.

ii) Data variety and lack of expandability in SHM: SHM data has a wide variety depending on the type of infrastructure and sensors, quality of the database and background noise, level of damage and sensor locations, presence of outlier and bias, environmental and operating conditions. Therefore, the existing literature of data-driven condition assessment approaches has primarily focussed on finding the most appropriate CNN architecture (Yuem *et al*. 2018) required for specific data of interest. For example, it may not be necessary that the training data of a steel and concrete bridge of the same length subjected to similar operational and environmental loads will have identical CNN architecture. The scalability and expandability of CNN architecture across various infrastructure is still a challenge.

iii) Cost of implementation to the infrastructure owners: Depending on the complexity in the data and existing conditions of a critical infrastructure, a deep and complex network is often needed to train a large database of SHM data. Such implementation of network demands high-performance workstations, cloud computing, parallel processing, graphic processing units and massive storage. Therefore, CNN is associated with high operating costs to analyze big data of infrastructure monitoring and maintenance for the decision-makers.

iv) Amplification of error in the network due to poorly measured data: False positives are often triggered due to varying image background caused by environmental effects (e.g., shadow, texture, light, rain, fog, and other adverse weather conditions), changes in color (e.g., material deterioration), and the presence of unwanted objects (e.g., debris, people, and vehicles). These noisy training data may lead to inaccurate damage detection in public infrastructures such as bridges, pavements, potholes, and pipelines (Azimi and Pekcan 2019; Kang *et al.* 2020). In particular, the impact of weather and lighting conditions, background noise, and the distance of the camera from the structures have still not been investigated in the context of multiclass crack detection.

The false positives may be removed using the traditional image processing or time-series based anomaly detection techniques during the data preparation stage. Having a well-processed data will enable CNN to produce higher accuracy and precision-recall value. The SHM community has advanced in the use of DL algorithms; however, data preparation and the amount of data usage without increasing the complexity of the network architecture is an open area of research. Moreover, the optimal network architecture and the configurations of input images and categories are still topics of active research in SHM.

v) Multiclass damage detection as a black box operation: There is often a lack of robustness in detecting multiple damage types (e.g., identification of cracks due to fatigue, delamination, voids, spalling, corrosion, etc.), requiring CNN architecture to be significantly deep to classify various components (Khodabandehlou *et al.* 2019). Any data-driven CNN network involves a scientific selection of the structure of layers as well as an optimal number of layers (Sandler *et al.* 2019; Tan and Le 2019) to achieve the best accuracy without resulting in overfitting, which still forms a black box to the majority of the structural engineers and infrastructure owners. Apart from the system architecture, the black-box nature of neural networks or CNN *per se* appears due to the traditional interpretability of the results. The matrices used for most of the networks are the accuracy and ROC curves, however, in a situation like structural damage detection and localization, only accuracy as a measure of performance of the CNN model may lead to catastrophic failures. Considering "false-negative rate" along with accuracy will improve the damage diagnosis model and also remove any situation where the CNN model ignores the possibility of damage. Moreover, improved visualization techniques of layer-wise classification results will eliminate the black-box nature of CNN for complex SHM applications.

8. Future Research Directions

i) Next-generation infrastructure monitoring and maintenance using big data: Smart and autonomous monitoring systems of future urban cities will result in internet-of-things (IoT)-enhanced big data for large-scale structures. This data will include either time-series measurements obtained from long-term embedded sensors within the structures or a large number of images obtained from sophisticated vision measurement systems such as drones and robots (Spencer *et al.* 2019). Such big data will enable a large and wide range of databases for CNN methods for robust structural condition assessment, and eliminate data imbalance issue.

ii) Real-time CNN implementation for remote and autonomous SHM systems: 1D CNN (Kiranyaz *et al.* 2020) has shown capabilities of utilizing a shallow architecture for structured SHM data such as time-series (e.g., vibration measurement). This results in less computationally intensive tasks on CNN, which can be implemented in mobile or handheld devices that are low cost and low powered in nature. Future application of 1D CNN will enable real-time indirect SHM for bridges using smart-phones installed in passing vehicles. There is a need to develop efficient strategies to accelerate the training and validation process and reduce the cost of deployment of CNN algorithms in SHM.

iii) Transfer learning-enabled efficient CNN using SHM data across various infrastructure: Improved CNN integrated with TL and Active Learning (Bull *et al*. 2018, 2019), and population-based SHM technique (Worden *et al.* 2015) may offer attractive solutions where statistically similar datasets of identical structure can be leveraged to replace the requirement for large training datasets from existing structures. CNN methods trained in one domain may be transferred into other domains, especially when the previous domain lacks training data. TL is a new development that uses knowledge from a source domain to target a domain that might be related but different, making existing pre-trained models more useful in the context of limited available datasets and relaxing the prerequisite for larger training datasets. The primary use of TL in CNN would be to use the parameters in a well-trained model in the source domain and to assist in generating limited training datasets in the target domain. The application of TL has a promising future while using the well-established benchmarks models for training the model and feature extraction, and improving the fully-connected classification layer for damage diagnosis.

iv) Field implementation: At present, there exist very few civil engineering image databases that have representative images of the damage to train the CNN architectures. Many images are obtained in a laboratory setting. Very few studies quantify the influence of measurement noise (wind, light, and angle) or mechanical vibrations from UAVs on the ability to capture damage using CNNs accurately. More controlled field measurements and shared case studies will allow SHM researchers to check the robustness and efficacy of the new algorithms. It is also expected that the SHM community will see a significant revolution of large databases in the near future that will allow the researchers to validate the new algorithms for a broad range of images.

v) Improved visualization of big SHM data: Building information modeling and mixed reality such as virtual reality and augmented reality has huge potential to allow structural engineers to manage and visualize long-term SHM data (Napolitano *et al.* 2018; Boddupalli *et al.* 2019, Singh and Sadhu 2020). These visualization tools integrated with the data storage capabilities of cloud computing, high-performance computing, and parallel processing will allow systematic interpretation of long-term SHM data.

vi) Multidisciplinary research in SHM: Although CNN and its architectures stem from Computer Science and Data Analytics, domain expertise in structural engineering and SHM is still of paramount importance to select appropriate features and classes specific to any SHM applications. On the other hand, the selection of a suitable number of hidden layers (i.e., depth of the network), structure of the network, and various hyper-parameters such as the number of epochs, batch size, and iterations vary with the data and should be carefully selected by the AI experts. Therefore, multidisciplinary research amongst the researchers from structural engineering, computer science, and big data analytics will be essential to achieve optimal performance.

vii) The potential use of video data in SHM: The majority of current approaches are limited to static images and do not apply to video data. Future research should be directed to acquiring high-definition videos and processing them as a sequential dataset of static images using RNNs.

Finally, figure 3 shows a summary of potential future research directions that will enhance the deployment of CNN in many SHM applications in upcoming years. Three critical components include balanced and real-time data collection and its visualization, development of laboratory and field measurements, and use of various forms of data type, such as time-series data and video data.

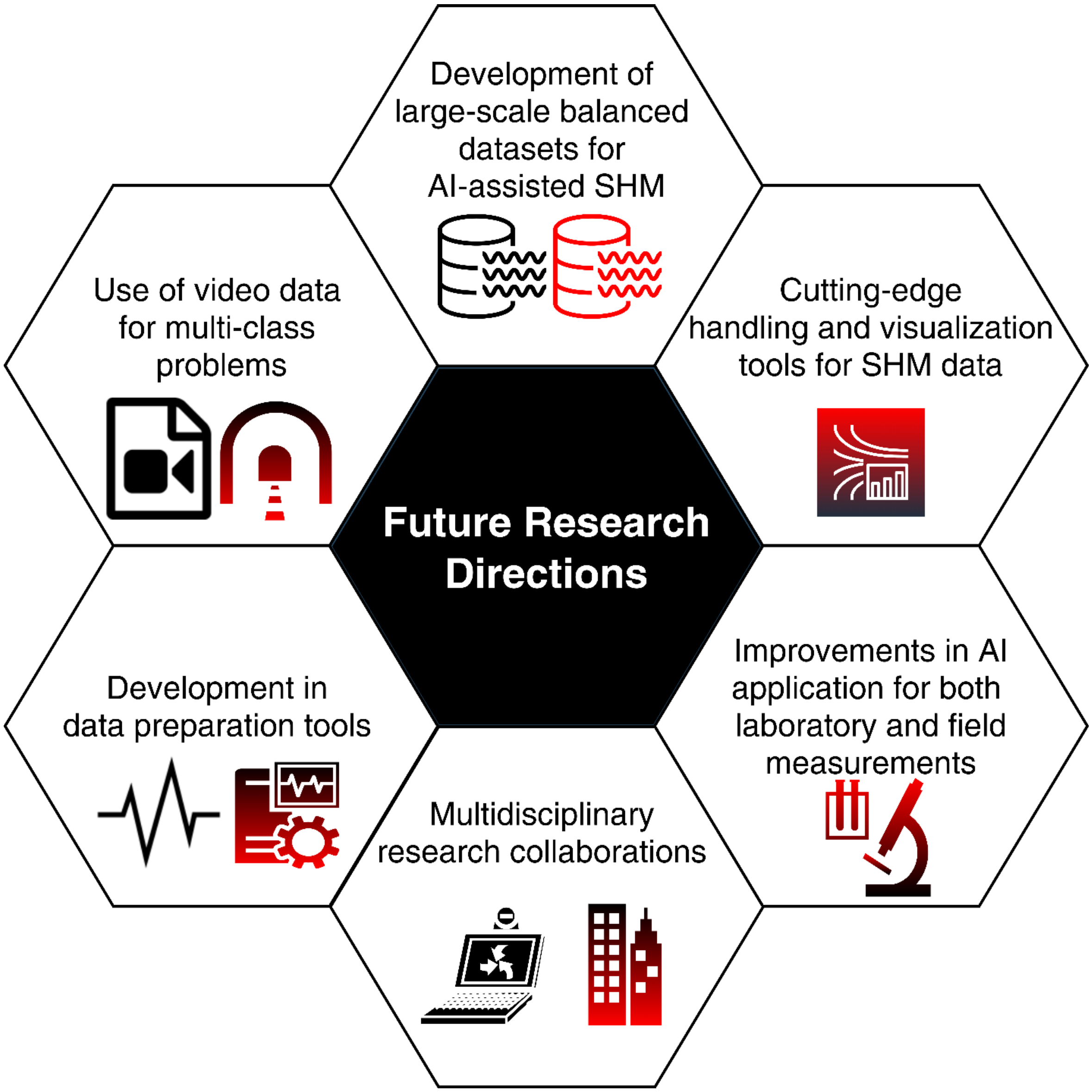


Figure 3. A schematic of the potential future research directions of CNN-based SHM research.

9. Conclusions

Civil Structures are composed of several material types, and often, therefore, subject to a wide range of damage categories. Such diversity applies to not only the majority of civil structures, but also railway infrastructure, pipelines, power generation plants, transmissions lines, and towers. Moreover, there is a prevalence among these structures to be highly susceptible to damages due to natural disasters and life-span fatigue due to ageing or normal operational conditions. Also, post-disaster inspections are often time-consuming, unsafe, and labour-intensive, making it difficult for human beings to accomplish these tasks efficiently. This paper systematically reviews the recent development of CNN-based SHM research that has been directed to solve these challenges. The state-of-the-art CNN-based architectures and newer SHM technologies have allowed the infrastructure owners to accurately and autonomously detect and localize multiple damage types in various structures using next-generation sensors such as cameras, drones and robots. In conclusion, future research will focus on developing the real-time implementation of CNN algorithms, open-source databases for civil structures, generalized application of CNN techniques using TL, and reducing classification imbalances that occur in large-scale infrastructure.

Acknowledgments

The authors would like to thank the Natural Sciences and Engineering Research Council (NSERC) of Canada to provide financial support to conduct this research through NSERC Discovery Grant and Western University's NSERC Accelerator Grant of the corresponding author.

References

Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M. and Inman, D. J. (2017), "Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks", *Journal of Sound and Vibration*, 388, 157 – 170.

Abdeljaber, O., Avci, O., Kiranyaz, M. S., Boashash, B., Sodano, H., & Inman, D. J. (2018). "1-D CNNs for structural damage detection: Verification on a structural health monitoring benchmark data". *Neuro-computing*, 275, 1308–1317. https://doi.org/10.1016/j.neucom.2017.09.069

Abiodun, O. I., Jantan, A., Omolara, A. E., Dada, K. V., Mohamed, N. A., and Arshad, H. (2018), "State-of-the-art in artificial neural network applications: A survey", *Heliyon*, 4, e00938.

Almasri, N., Sadhu, A., and Ray Chaudhury, S. (2020). "Towards compressed sensing of structural monitoring data using discrete cosine transform.", *Journal of Computing in Civil Engineering*, 34(1), 04019041.

Aria, A., Droguett, E. L., Azarm, S. & Modarres, M. (2019). "Estimating damage size and remaining useful life in degraded size and remaining useful life in degraded structures using deep learning-based multi-source data fusion". *Structural Health Monitoring*, https://doi.org/10.1177 /1475921719890616.

Atha, D. J., & Jahanshahi, M. R. (2018). Evaluation of deep learning approaches based on convolutional neural networks for corrosion detection. *Structural Health Monitoring*, 17(5), 1110–1128. https://doi.org/10.1177/1475921717737051.

Azimi, M., and Pekcan, G. (2019). "Structural health monitoring using extremely compressed data through deep learning". *Computer-Aided Civil and Infrastructure Engineering*, 1-18. https://doi.org/10.1111/mice.12528.

Bao, Y., Chen, Z., Wei, S., Xu, Y., Tang, Z. and Li, H. (2019), "The state of the art of data science and engineering in structural health monitoring", *Engineering*, 5(2), 234-242.

Bao, Y., Tang, Z., Li, H., & Zhang, Y. (2019). Computer vision and deep learning-based data anomaly detection method for structural health monitoring. Structural Health Monitoring, 18(2), 401–421. https://doi.org/10.1177/1475921718757405.

Barbosh, M., Singh, P., and Sadhu, A. (2020), "Empirical mode decomposition and its variants: a review with applications in structural health monitoring", *Smart Materials and Structures*, 29(9).

Behrouzi, A., & Pantoja, M. (2018). "Software Toolset To Enable Image Classification of Earthquake Damage To Above-Ground Infrastructure". *16th European Conference on Earthquake Engineering*, Thessaloniki, 18-21 June 2018, 1–10.

Boddupalli, C., Sadhu, A., Azar, E. and Pattyson, S. (2019). "Improved visualization of infrastructure monitoring data using building information modeling". *Structure and Infrastructure Engineering*, 15(9), 1247 – 1263.

Bull, L., Worden, K., Manson, G., and Dervilis, N. (2018), "Active learning for semi-supervised structural health monitoring", *Journal of Sound and Vibration*, 437, 373-388.

Bull, L. A., Rogers, T. J., Wickramarachchi, C., Cross, E. J., Worden, K. and Dervilis, N. (2019), "Probabilistic active learning: An online framework for structural health monitoring", *Mechanical Systems and Signal Processing*, 134, 1 – 20.

Cawley, P., (2018), "Structural health monitoring: closing the gap between research and industrial deployment", *Structural Health Monitoring*, 17(5), 1225 – 1244.

Cha, Y. J., Choi, W. and Buyukozturk, O. (2017), "Deep learning-based crack damage detection using convolutional neural networks", *Computer-Aided Civil and Infrastructure Engineering*, 32, 361 – 378.

Cha, Y. J., Choi, W., Suh, G. and Mahmoudkhani, S. (2018), "Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types", *Computer-Aided Civil and Infrastructure Engineering*, 33, 731 – 747.

Chaiyasarn, K., Sharma, M., Ali, L., Khan, W., & Poovarodom, N. (2018). Crack detection in historical structures based on Convolutional Neural Network. International Journal of GEOMATE, 15(51), 240–251. https://doi.org/10.21660/2018.51.35376.

Chen, F. C. and Jahanshahi, M. R. (2018), "NB-CNN: Deep learning-based crack detection using convolutional neural network and naïve Bayes data fusion", *IEEE Transactions on Industrial Electronics*, 65 (5), 4392 – 4400.

Cheng, J. C. P., & Wang, M. (2018). Automated detection of sewer pipe defects in closed-circuit television images using deep learning techniques. Automation in Construction, 95, 155–171. https://doi.org/10.1016/j.autcon.2018.08.006

Dabous, S., and Feroz, S. (2020), "Condition monitoring of bridges with noncontact testing technologies", *Automation in Construction*, 116, 103224.

Deng, J., Lu, Y. and Lee, V. C.-S. (2019), “Concrete crack detection with handwriting script interferences using faster region-based convolutional neural network”, *Computer-Aided Civil and Infrastructure Engineering*, 1 – 17.

Dick, K., Russell, L., Dosso, Y. S. and Kwamena, F. (2019), “Deep learning for critical infrastructure resilience”, *Journal of Infrastructure Systems*, 25(2).

Dorafshan, S., Thomas, R. J., Coopmans, C., & Maguire, M. (2018a). Deep Learning Neural Networks for sUAS-Assisted Structural Inspections: Feasibility and Application. 2018 International Conference on Unmanned Aircraft Systems, ICUAS 2018, (June), 874–882. https://doi.org/10.1109/ICUAS.2018.8453409.

Dorafshan, S., Thomas, R. J. and Maguire, M. (2018b), “SDNET2018: An annotated image dataset for non-contact concrete crack detection using deep convolutional neural networks”, *Data in Brief*, 21, 1664 – 1668.

Dorafshan, S., Thomas, R. J. and Maguire, M. (2018c), “Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete”, *Construction and Building Materials*, 186, 1061 – 1045.

Doulamis, A., Doulamis, N., Protopapadakis, E., & Voulodimos, A. (2018). Combined convolutional neural networks and fuzzy spectral clustering for real-time crack detection in tunnels. Proceedings - International Conference on Image Processing, ICIP, 4153–4157. https://doi.org/10.1109/ICIP.2018.8451758

Dung, C. V., and Duc Anh, L. (2019), “Autonomous concrete crack detection using deep fully convolutional neural network”, *Automation in Construction*, 99, 52 – 58.

Fallahian, M., Khoshnoudian, F., & Meruane, V. (2018). Ensemble classification method for structural damage assessment under varying temperatures. *Structural Health Monitoring*, 17(4), 747–762. https://doi.org/10.1177/1475921717717311

Fan, Z., Wu, Y., Lu, J. and Li, W. (2018), “Automatic pavement crack detection based on structural prediction with the convolutional neural network”, *arXiv preprint arXiv:1802.02208.*

Fan, R., Bocus, M. J., Zhu, Y., Jiao, J., Wang, L., Ma, F., Cheng, S., and Liu, M. (2019), “Road crack detection using deep convolutional neural network and adaptive thresholding”, *arXiv preprint arXiv: 1904.08582.*

Feng, C., Liu M.-Y., Kao C.-C. and Lee, T.-Y. (2017), “Deep active learning for civil infrastructure defect detection and classification”, *Computing in Civil Engineering 2017: Information Modeling and Data Analytics*, 298 – 306.

Feng, C., Zhang, H., Wang, S., Li, Y., Wang, H. and Yan, F. (2019), “Structural damage detection using deep convolutional neural network and transfer learning”, *KSCE Journal of Civil Engineering*, 23(10), 4493 – 4502.

Feng, D., & Feng, M. Q. (2018). Computer vision for SHM of civil infrastructure: From dynamic response measurement to damage detection – A review. Engineering Structures, 156, 105–117. https://doi.org/10.1016/j.engstruct.2017.11.018

Funahashi, K-I, and Nakamura, Y. (1993). Approximation of dynamical systems by a continuous-time recurrent neural network, *Neural Network*, 6(6), 801-806.

Gao, Y. and Mosalam, K. M. (2018), “Deep transfer learning for image-based structural damage recognition”, *Computer-Aided Civil and Infrastructure Engineering*, 33, 748 – 768.

Gomes, G. F., Mendéz, Y. A. D., da Silva Lopes Alexandrino, P., da Cunha, S. S., & Ancelotti, A. C. (2018). The use of intelligent computational tools for damage detection and identification with an emphasis on composites – A review. Composite Structures. https://doi.org/10.1016/j.compstruct.2018.05.002.

Goodfellow, I., Bengio, Y. and Courville, A. (2016), “Deep Learning”, *Adaptive Computation and Machine Learning Series*, MIT Press.

Gopalakrishnan, K., Khaitan, S. K., Choudhary, A. and Agrawal, A. (2017), “Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection”, *Construction and Building Materials*, 157, 322 – 330.

Gopalakrishnan, K., Gholami, H., & Agrawal, A. (2018). Crack Damage Detection in Unmanned Aerial Vehicle Images of Civil Infrastructure Using Pre-Trained Deep Learning Model. *International Journal for Traffic and Transport Engineering*, 8(1), 1–14. https://doi.org/10.7708/ijtte.2018.8(1).01.

Gu et al. (2017), “Recent advances in convolutional neural networks”, *Pattern Recognition*.

Gulgec, N. S., Takac, M. and Pakzad, S. N. (2017), “Structural damage detection using convolutional neural networks”, *Conference Proceedings of the Society for Experimental Mechanics Series*, 331 – 337.

Guo, J., Xie, X., Bie, R. and Sun, L. (2014), “Structural health monitoring by using a sparse coding-based deep learning algorithm with wireless sensor networks”, *Pers Ubiquit Comput*, 18, 1977 – 1987.

Hinton, G. E., Osindero, S., and Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. Neural Computation, 18(7), 1527–1554.

Hoskere, V., Narazaki, Y., Hoang, T. A., and Spencer Jr., B. F. (2017), “Vision-based structural inspection using multi-scale deep convolutional neural networks”, *3rd Huixian International Forum on Earthquake Engineering for Young Researchers*, August 2017.

Hoskere, V., Narazaki, Y., Hoang, T. A., & Spencer, B. F. (2018). “Towards Automated Post-Earthquake Inspections with Deep Learning-based Condition-Aware Models”. The 7th World Conference on Structural Control and Monitoring, 7WCSCM, July 22-25, 2018, Qingdao, China. Retrieved from http://arxiv.org/abs/1809.09195.

Hoskere, V., Eick, B., Spencer, B. F., Smith, M. D., & Foltz, S. D. (2019). “Deep Bayesian neural networks for damage quantification in miter gates of navigation locks”, *Structural Health Monitoring*, 1-30.

He, K., Zhang, X., Ren, S. & Sun, J. (2015). Deep Residual Learning for Image Recognition, arxiv:1512.03385.

Jahanshahi, M. R., Kelly, J. S., Masri, S. F. and Sukhatme, G. S. (2009), “A survey and evaluation of promising approaches for automatic image-based defect detection of bridge structures”, *Structure and Infrastructure Engineering*, 5(6), 255-486.

Jiang, S., and Zhang, J. (2019), "Real-time crack assessment using deep neural networks with the wall-climbing unmanned aerial system", *Computer-Aided Civil and Infrastructure Engineering*. https://doi.org/10.1111/mice.12519.

Kankanamge, Y., Hu, Y. and Shao, X. (2020), "Application of wavelet transform in structural health monitoring." *Earthquake Engineering and Engineering Vibration*, 19, 515-532.

Kang, D. and Cha, Y.-J. (2018). Damage detection with an autonomous UAV using deep learning. In Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018 (Vol. 1059804, p. 4). https://doi.org/10.1117/12.2295961.

Kang, M-S., Kim, N., Lee, J. J, and An, Y-K. (2020). Deep learning-based automated underground cavity detection using three-dimensional ground penetration radar. *Structural Health Monitoring*, 2018, 19(1), 173-185.

Khodabandehlou, H., Pekcan, G., and Fadali, M. S. (2018), "Vibration-based structural condition assessment using convolution neural networks", *Structural Control and Health Monitoring*, 26, e2308.

Kim, B. and Cho, S. (2018), "Automated vision-based detection of cracks on concrete surface using a deep-learning technique", *Sensors*, 18(10), 1 -18.

Kim, I. H., Jeon, H., Baek, S. C., Hong, W. H., & Jung, H. J. (2018). Application of crack identification techniques for an aging concrete bridge inspection using an unmanned aerial vehicle. Sensors (Switzerland), 18(6), 1–14. https://doi.org/10.3390/s18061881

Kim, H. and Sim, S.-H. (2019), "Automated peak picking using region-based convolutional neural network for operational modal analysis", *Structural Control and Health Monitoring*, 1 – 16.

Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., and Inman, D. J. (2019), "1D convolutional neural networks and applications – a survey", *Mechanical Systems and Signal processing.*

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6), 84–90.

Koch, C., Georgieva, K., Kasireddy, V., Akinci, B., & Fieguth, P. (2015). A review on computer vision-based defect detection and condition assessment of concrete and asphalt civil infrastructure. *Advanced Engineering Informatics*, 29(2), 196–210. https://doi.org/10.1016/j.aei.2015.01.008

Lattanzi, D. and Miller, G. (2017), "Review of robotic infrastructure inspection systems", *Journal of Infrastructure Systems*, 23(3), 04017004.

Lecun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444.

Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., and Nandi, A. K. (2020). "Application of machine learning to machine fault diagnosis: A review and roadmap". *Mechanical Systems and Signal processing*, 138, 106587.

Lee, S., Ha, J., Zokhiroba, M., Moon, H., & Lee, J. (2018). "Background information of deep learning for structural engineering. *Arch Computat Methods Eng*, 25, 121-129.

Lee, Y. S., Vakakis, A. F., Mcfarland, D. M., & Bergman, L. A. (2014). Crack detection and characterization techniques-An overview. Structural Control and Health Monitoring, 21(June), 742–760. https://doi.org/10.1002/stc.

Li, S. and Zhao, X. (2018), "Convolutional neural networks-based crack detection for real concrete surface", Proceedings of the SPIE.

Li, S., Zhao, X. and Zhou G. (2019), "Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network", *Computer-Aided Civil and Infrastructure Engineering*, 34, 616 - 634.

Liang, X. (2018), 'Image-based post-disaster inspection of reinforced concrete bridge systems using deep learning with Bayesian optimization", *Computer-Aided Civil and Infrastructure Engineering*, 34, 415 – 430.

Lin, Y.-Z., Nie, Z.-H. and Ma, H.-W. (2017), "Structural damage detection with automatic feature-extraction through deep learning", *Computer-Aided Civil and Infrastructure Engineering,* 32, 1025 – 1046.

Lin, Z., Pan, H., Gui, G., & Yan, C. (2018). Data-driven structural diagnosis and conditional assessment: from shallow to deep learning. In Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018 (Vol. 1059814, p. 38). https://doi.org/10.1117/12.229696

Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T. and Omata, H. (2018a), "Road damage detection using deep neural networks with images capture through a smartphone", *arXiv preprint arXiv:1801.09454.*

Maeda, H., Sekimoto, Y., Seto, T., Kashiyama, T., & Omata, H. (2018b). Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images. Computer-Aided Civil and Infrastructure Engineering, 33(12), 1127–1141. https://doi.org/10.1111/mice.12387

Makantasis, K., Protopapadakis, E., Doulamis, A., Doulamis, N., & Loupos, C. (2015). Deep Convolutional Neural Networks for efficient vision-based tunnel inspection. In Proceedings - 2015 IEEE 11th International Conference on Intelligent Computer Communication and Processing, ICCP 2015 (pp. 335–342). https://doi.org/10.1109/ICCP.2015.7312681.

Mei, Q. and Gul, M. (2020), "Multi-level feature fusion in densely connected deep-learning architecture and depth-first search for crack segmentation on images collected with smartphones", *Structural Health Monitoring*¸ https://doi.org/10.1177/1475921719896813

Mohan, A. and Poobal, S. (2018), "Crack detection using image processing: A critical review and analysis", *Alexandria Engineering Journal*¸ 57, 787-798.

Napolitano, R., Blyth, A., and Glisic, B. (2018), "Virtual environments for visualizing structural health monitoring sensor networks, data, and metadata.", *Sensors*, 18(1).

Narazaki, Y., Hoskere, V., Hoang, T. A. and Spencer Jr., B. F. (2017), "Automated vision-based bridge component extraction using multiscale convolutional neural networks", *3^{rd} Huixian International Forum on Earthquake Engineering for Young Researchers*, August 2017.

Ni, F., Zhang, J., and Noori, M. N. (2019). "Deep learning for data anomaly detection and data compression of a long-span suspension bridge", *Computer-Aided Civil and Infrastructure Engineering*. https://doi.org/10.1111/mice.12528

Pan, H., Gui, G., Lin, Z., & Yan, C. (2018). Deep BBN Learning for Health Assessment toward Decision-Making on Structures under Uncertainties. KSCE Journal of Civil Engineering, 22(3), 928–940. https://doi.org/10.1007/s12205-018-1301-2.

Pathirage, C. S. N., Li, J., Li, L., Hao, H., Liu, W., & Ni, P. (2018). Structural damage identification based on auto-encoder neural networks and deep learning. Engineering Structures, 172, 13–28. https://doi.org/10.1016/j.engstruct.2018.05.109.

Pathirage, C. S. N., Li, J., Li, L., Hao, H., Liu, W., & Wang, R. (2019). Development and application of deep learning-based sparse autoencoder framework for structural damage identification. Structural Health Monitoring, 18(1), 103–122. https://doi.org/10.1177/1475921718800363.

Patterson, B., Leone, G., Pantoja, M., & Behrouzi, A. (2018). Deep Learning for Automated Image Classification of Seismic Damage To Built Infrastructure. *In 11th U.S. National Conference on Earthquake Engineering* (Vol. 93407).

Perez, H., Tah, J. H. M., & Mosavi, A. (2019). Deep Learning for Detecting Building Defects Using Convolutional Neural Networks. Sensors, 19(August), 1–23. https://doi.org/10.20944/preprints201908.0068.v1.

Pouyanfar, S., Sadiq, S., Yan, Y., Tian, H., Tao, Y., Reyes, M. P., Shyu, M., Chen, S.C., Iyenger, S. (2018). "A survey on deep learning: algorithms, techniques, and applications". *ACM Computing Survey*, 51(5).

Qarib, H. and Adeli, H. (2016). "A comparative study of signal processing methods for structural health monitoring". *Journal of Vibroengineering*, 18(4), 2186-2204.

Rubio, J. J., Kashiwa, T., Laiteerapong, T., Deng, W., Nagai, K., Escalera, S., Nakayama, K., Matuso, Y. and Prendinger, H. (2019), "Multi-class structural damage segmentation using fully convolutional networks", *Computers in Industry*, 112, 1 – 10.

Rui, M., Yuntian, G., Liang, G., Zihang, J. and Jie, Z. (2019), "Online defect recognition of narrow overlap weld based on two-stage recognition model combining continuous wavelet transform and convolutional neural network", *Computers in Industry*, 112, 1 – 11.

Sadhu, A., Narasimhan, S., and Antoni, J. (2019), "A review of output-only structural mode identification literature employimg blind source separation", *Mechanical Systems and Signal Processing*, 94, 415 – 431.

Sandler, M., Howard, M., Zhu, M., Zhmoginov, A., and Chen, L. (2019). "MobileNetV2: Inverted residuals and linear bottlenecks", *The IEEE Conference on Computer Vision and Pattern Recognition*, 4510 – 4520.

Salakhutdinov, R. and Hinton, G. E. (2009). Deep Boltzmann machines, *Proceedings of the 12^{th} International Conference on Artificial Intelligence and Statistics (AISTATS)*, 5, 445-448.

Salehi, H., & Burgueño, R. (2018). "Emerging artificial intelligence methods in structural engineering". *Engineering Structures*, 171(May), 170–189. https://doi.org/10.1016/j.engstruct.2018.05.084.

Silva, M., Santos, A., Santos, R., Figueiredo, E., Sales, C. & Costa, J. C. (2019). Deep principal component analysis: An enhanced approach for structural damage identification. *Structural Health Monitoring*, 18(5), 1444-1463.

Simonyan, K. & Zisserman, A. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*, abs/1409.1556.

Singh, P. and Sadhu, A. (2020), "System identification-enhanced visulaization tool for infrastructure monitoring and maintenance", *Frontiers in Built Environment*, 6, 76.

Sony, S., Laventure, S. and Sadhu, A. (2019), "A literature review of next-generation smart sensing technology in structural health monitoring", *Structural Control & Health Monitoring*, 26, e2321.

Soukup, D. and Huber-Mork, R. (2014), "Convolutional neural networks for steel surface defect detection from photometric stereo images", *International Symposium on Visual Computing (ISVC 2014): Advances in Visual Computing*, 668 – 677.

Spencer, B. F., Hoskere, V., & Narazaki, Y. (2019). Advances in Computer Vision-Based Civil Infrastructure Inspection and Monitoring. *Engineering*, 5(2), 199–222. https://doi.org/10.1016/j.eng.2018.11.030

Stentoumis, C., Protopapadakis, E., Doulamis, A., & Doulamis, N. (2016). A holistic approach for inspection of civil infrastructures based on computer vision techniques. In International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives (Vol. 41, pp. 131–138). https://doi.org/10.5194/isprsarchives-XLI-B5-131-2016.

Sun, L., Shang, Z., Xia, Y., Bhowmick, S., Nagarajaiah, S. (2020). "Review of bridge structural health monitoring aided by big data and artificial intelligence: from condition assessment to damage detection.", *Journal of Structural Engineering*, 146(5), 04020073.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V. & Rabinovich, A. (2014). Going Deeper with Convolutions, arxiv:1409.4842.

Tan, M., and Le, Q. V. (2019). "EfficientNet: Rethinking model scaling for convolutional neural network", *ICML*.

Teng, S., Chen, G., Gong, P., Liu, G. and Cui, F. (2019), "Structural damage detection using convolutional neural network combining strain energy and dynamic response", *Mechanica*.

Tong, Z., Gao, J., Han, Z. and Wang, Z. (2017), "Recognition of asphalt pavement crack length using deep convolutional neural networks", *Road Materials and Pavement Design*, 19(6), 1334 – 1349.

Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P.-A. (2008). Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th International Conference on Machine Learning* - ICML 08.

Worden, K., Cross, E. J., Dervilis, N., Papatheou, E., Antoniadou, I., (2015). Structural health monitoring: from structures to systems-of-systems. *IFAC-PapersOnline*, 48(21), 1-17.

Xu, D., Wen, C., & Liu, J. (2019). Wind turbine blade surface inspection based on deep learning and UAV-taken images. *Journal of Renewable and Sustainable Energy*, 11, 053305.

Xu, Y., Bao, Y., Chen, J., Zuo, W., & Li, H. (2019). Surface fatigue crack identification in steel box girder of bridges by a deep fusion convolutional neural network based on consumer-grade camera images. *Structural Health Monitoring*, 18(3), 653–674. https://doi.org/10.1177/1475921718764873

Xue, Y., & Li, Y. (2018). A Fast Detection Method via Region-Based Fully Convolutional Neural Networks for Shield Tunnel Lining Defects. *Computer-Aided Civil and Infrastructure Engineering*, 33(8), 638–654. https://doi.org/10.1111/mice.12367

Yang, X., Li, H., Yu, Y., Luo, X., Huang, T. and Yang, X. (2018), "Automated pixel-level crack detection and measurement using fully convolutional network", *Computer-Aided Civil and Infrastructure Engineering*, 33, 1090 – 1109.

Yao, G., Tao, L., and Zhong, Jiandin (2019), "A review of convolutional-neural-network-based action recognition", *Pattern Recognition Letters*, 118, 14-22.

Yao, Y., Tung, S., and Glisic, B. (2014), "Crack detection and characterization techniques – An Overview", *Structural Control and Health Monitoring*, 21, 1387-1413.

Ye, X. W., T. Jin, and C. B. Yun (2019). "A review on deep learning-based structural health monitoring of civil infrastructures." *Smart Structures and Systems*, 24(5), 567-585.

Yeum, C. M., Dyke, S. J., & Ramirez, J. (2018). Visual data classification in post-event building reconnaissance. *Engineering Structures*, 155, 16–24. https://doi.org/10.1016/j.engstruct.2017.10.057

Ying, Y., Garrett Jr., J. H., Oppenheim, I. J., Soibelman, L., Harley, J. B., Shi, J. and Jin, Y. (2013), "Toward data-driven structural health monitoring: application of machine learning and signal processing to damage detection", *Journal of Computing in Civil Engineering*, 27(6), 667-680.

Zakeri, H., Nejad, F. M. and Fahimifar, A. (2017), "Image-based techniques for crack detection, classification and quantification in asphalt pavement: a review", *Archives of Computational Methods in Engineering*¸24(4), 935-977.

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

Zhang, K., Cheng, H. D., and Zhang, B. (2018), "Unified approach to pavement crack and sealed crack detection using pre-classification based on transfer learning", *Journal of Computing in Civil Engineering*, 32(2).

Zhang, Y., Miyamori, Y., Mikami, S., and Saito, T. (2019). "Vibration-based structural state identification by a 1-dimensional convolutional neural network", *Computer-Aided Civil and Infrastructure Engineering*, 34(9), 822-839.

Zhang, Y., Sun, X., Loh, K. J., Su, W., Xue, Z. and Zhao, X. (2020). "Autonomous bolt loosening detection using deep learning", *Structural Health Monitoring*, 19(1), 105-122.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P. and Gao, R. X. (2015), “Deep learning and its applications to machine health monitoring: a survey”, *Journal of Latex Class Files*, 14(8), 1-14.

Zhao, X., & Li, S. (2018). Convolutional neural networks-based crack detection for real concrete surface (p. 143). https://doi.org/10.1117/12.2296536.

Zhao, X., Li, S., Su, H., Zhou, L. & Loh, K. J. (2018). “Image-based comprehensive maintenance and inspection method for bridges using deep learning”, SMASIS2018- 8268. Proceedings of the ASME 2018 Conference on Smart Materials, Adaptive Structures and Intelligent Systems SMASIS2018 September 10-12, 2018, San Antonio, TX, USA, 1–7.