Deep Learning Models for Analysis of Non-Destructive Evaluation Data of Reinforced Concrete Bridge Decks: A Survey

Dayakar L. Naik[1], and Sattar Dorafshan[2]

Abstract

Application of deep learning (DL) for automatic condition assessment of bridge infrastructure has been on the rise in the last few years. From the published literature, it is evident that lot of research efforts has been put in identifying the surface defects such as cracks, potholes, spalling etc. using deep learning. However, a concrete bridge deck health is jeopardized by the presence of subsurface defects substantially, however, the task of defect detection using deep learning has not received the proper attention. The goal of this survey paper is to provide a critical review of existing technical knowledge for DL application on NDE data for bridge deck evaluation. The authors reviewed prominent NDE techniques for subsurface defect detection of bridge decks and explored the various DL models proposed to identify these defects. First a brief overview of the working principle of NDE techniques and DL architectures is provided, and then the information about proposed DL models and their efficacy is highlighted. Based on the existing knowledge gaps, various challenges and future prospects associated with application of DL in bridge subsurface inspection are discussed.

Keywords: Bridge Deck; Delamination; Rebar; IRT; Impact Echo and GPR.

[1] Post-Doctoral Research Fellow, Dept. of Civil & Environmental Engineering, University of North Dakota, Grand Forks, ND 58201, email: dayakarnaik.lavadiya@und.edu

[2] Assistant Professor (Corresponding Author), Dept. of Civil & Environmental Engineering, University of North Dakota, Grand Forks, ND 58201, email: sattar.dorafshan@und.edu

1. Introduction

Bridges are the structures that play a pivotal role in nation's roadway network for facilitating connectivity in transportation across waterways, railways, roadways, and other obstacles [1]. According to National Bridge Inventory (NBI) statistics there are more than 619622 bridges in entire United States which are exposed to continuous traffic loads [2]. While 297,908 bridges are classified to be in 'Fair' condition by FHWA, 43,586 are labeled as 'Poor' [3,4]. Since deterioration of bridge structures is inevitable with the progress of time, it is essential to monitor their condition through periodic inspections to ensure safety and serviceability during operation. Although visual inspection has been a common practice employed by bridge management authorities in the past, Non-Destructive Evaluation (NDE) techniques has gained enormous attention in recent two decades with the advent of advanced sensors [5]. NDE technique provides more comprehensive, quantitative and objective condition information of the structural components [6,7]. As mentioned [7], an additional $8 billion annual investment would be needed until 2028 to improve bridge decks' condition to one of good repair. Owing to higher repair costs, there has been lot of emphasis on condition assessment of bridge decks when compared to other bridge components as they deteriorate faster [7,8].

NDE inspections results in both discrete and continuous responses from the sensors [9]. Such responses, referred to as 'signals', often need rigorous data analysis for decision making. However, traditionally, decision making has been very subjective and based on conventional wisdom. In other words, experts with NDE training are required to not only collect the data, but also, they have to have the expertise to confidently judge the quality of the data and to properly interpret them [10]. Adaption of this practice could be cumbersome, infeasible and may yield inconsistent results. Therefore, there is a need to develop and maintain workforce with consistently in data interpreting which currently don't exist. According to Omar et. al. [11] in the context of bridge condition assessment, a unified guidelines and procedures capable of accounting for the uncertainty and complexity in data interpretation is required. To circumvent the issues associated with the current state of NDE practice, a domain adaptative technique such as artificial intelligence (AI) that does not require human expert is necessary.

An increased interest in the structural engineering community to automate the condition assessment of infrastructure has allowed the researchers to adapt the field of artificial intelligence (AI) for bridge inspection. In AI framework, the set of descriptive features characterizing various types of defects are identified and fed as inputs to the AI algorithms. A mapping function or decision function that distinguishes or predicts the target defect is then learnt by the algorithm. While the deterioration of bridge deck could be categorized into two types, namely, surface defects [12] (e.g., cracks, potholes etc.) and subsurface defects (e.g., delamination and rebar corrosion) (see Figure 1), majority of AI application in recent years is noticed for the evaluation of surface detects using visual images [3,13,14]. Both conventional machine learning algorithms and deep learning algorithms were employed on visual images [15–20]. A comprehensive review on application of conventional ML algorithms and deep learning algorithms for surface defect detection could be found elsewhere [3,21,22]. The other set of review performed on application of conventional ML and computer vision techniques for subsurface defect detection can also be found in the literature [3]. Deep learning does not require human expert to identify features associated

with subsurface defects, and can be updated after each repair to augment the model. Keeping in view that application of deep learning for subsurface detection is relatively a new subject and there are no review studies, this paper focused on providing insights about the recent progress achieved in this area.

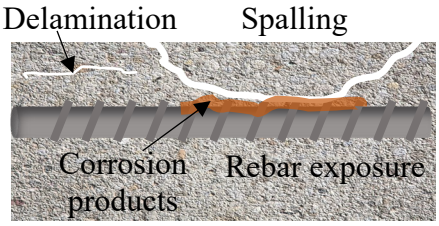

Figure 1. Illustration of subsurface defects in RC bridge.

The goal of this paper is to identify the state-of-the-art NDE techniques for subsurface defect detection in bridges and highlight the potential of deep learning algorithms that could be employed in conjunction with them. The rest of the paper is organized as follows: working principle of the selected NDE techniques is described in Section 2, methodology adapted to conduct the survey is explained in Section 3, an overview of deep learning architectures and their application for subsurface defect detection is reviewed in Section 4, and the challenges and future work recommendations are provided in Section 5.

2. NDE for Bridge Sub-surface Defects

The Strategic Highway Research Program report (SHRP) recognizes 14 potentially effective NDE techniques to detect and characterize subsurface defects and deterioration in reinforced concrete (RC) bridge decks [23]. They are, (1) Impact echo, (2) Ground-penetrating radar, (3) Infrared thermography, (4) Electrical resistivity, (5) Half-cell potential, (6) Microwave moisture technique, (7) Eddy current, (8) Ultrasonic pulse echo, (9) Galvanostatic pulse measurement, (10) Impulse response, (11) Ultrasonic surface waves, (12) Visual inspection, (13) Chain dragging and hammer sounding, and (14) Chloride concentration measurement. In the wake of technological advancements and the field application of current state-of-the-art techniques, this paper only focuses on first five techniques mentioned above that are more widely used in real practice [24–26]. In this section, a brief description about working principle of each technique is provided along with their specific advantages and limitations (see Table 1).

Table 1. Potential sensors and their limitations for bridge subsurface inspection [23]

Sensor	Standard	Potential	Limitations
Impact Echo	ASTM C 1383	Detects delamination, voids, honeycombing, elastic modulus and rebars	Less reliable in the presence of asphalt overlays and requires experienced operator and analyzing expert
GPR	ASTM D 6087	Deck thickness, delamination, corrosive environment and Rebar detection	Presence of moisture content introduces inconsistent results and cannot provide information about mechanical properties of concrete.

Infrared Thermography	ASTM D 4788	Delamination and corrosion, crack.	Reliability of results depends on Environment; and Cannot provide information about the depth of defects.
Electrical resistivity	ASTM D 3633	Corrosion and chloride penetration	Surface has to be prewetted; the data interpretation is challenging. Automated measurement systems for roads are not available on the market.
Half-cell potential	ASTM C 876	Corrosion	Not suitable for overlays or coated rebar; and moisture content will cause negative shift in potential voltage measurement

2.1 Infrared Thermography (IRT)

Infrared thermography uses thermal information from the Electromagnetic (EM) radiation (in the infrared range – 800nm-1500nm [27]) emitted by the probed material to identify the defects. The surface of the material exhibits a unique thermal signature (or localized contrast in surface temperature) when the inherent flaws are present in the material. This could be attributed to the altered rate of heat transfer in the defect zone on the material (see Figure 2). In the context of RC bridge deck consisting of delamination in the subsurface, the heat transfer is disrupted when compared to surrounding concrete wherein the material is intact. Consequently, the surface above delamination exhibits higher radiation (i.e. quantified as temperature) relative to the surrounding surface when the heat energy strikes the material surface. In practice, sun is used as the source of heat energy to increase the temperature of the object. The emitted radiant energy is then captured by the thermal cameras consisting of sensors which convert the EM waves to temperature. Higher the radiant energy higher is the temperature. According to ASTM D 4788 [28], the ideal time to acquire the data is 4 to 7 hr. after sunrise for noticing the delamination's of 2 and 3 in deep respectively. This ensures that at least a thermal contrast of 0.5°C is achieved within the delaminated area.

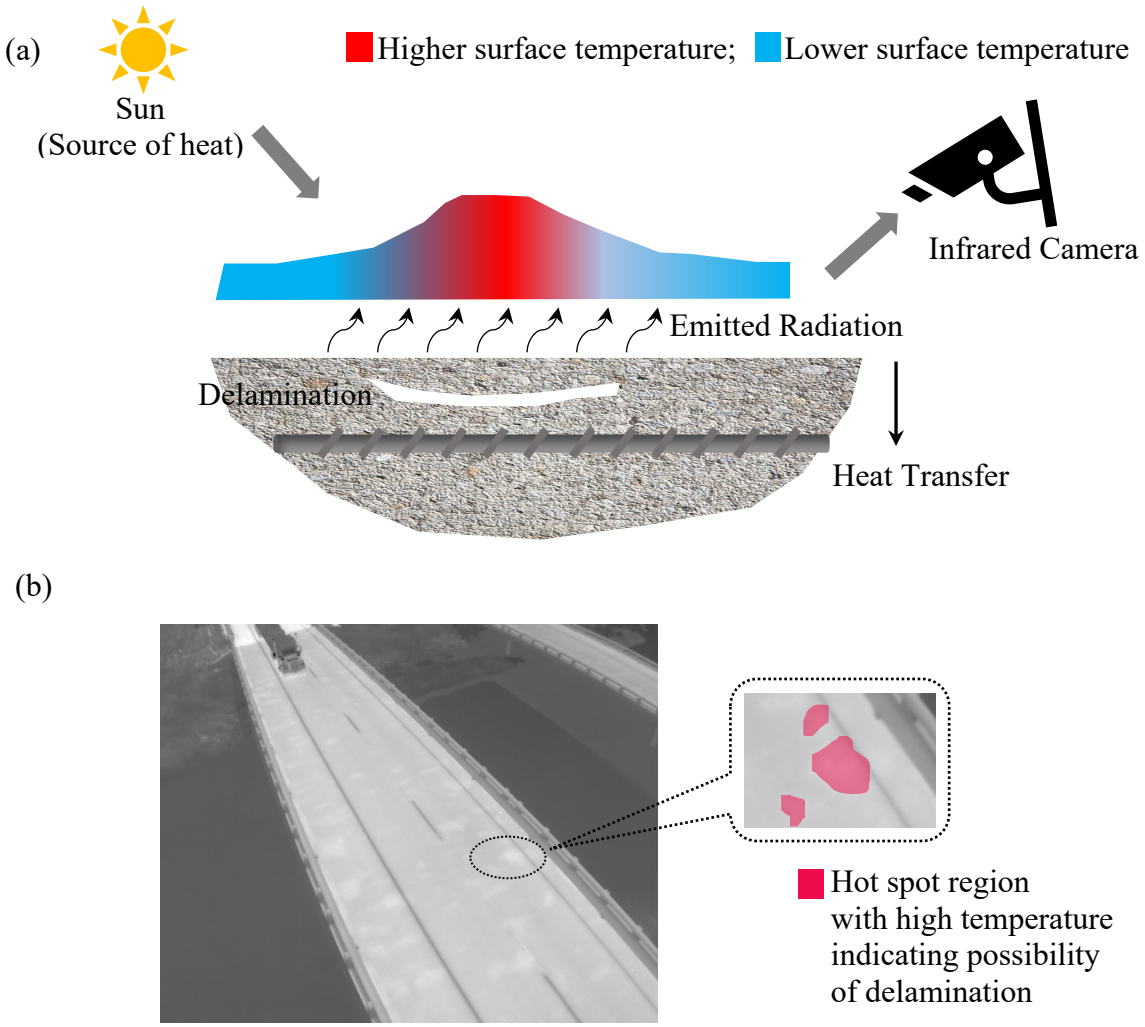

Figure 2. (a) Illustration of working principle of Infrared Thermography, (b) infrared thermography image from SDNET 2021 database [29].

2.2 Impact Echo (IE)

Impact echo technique (ASTM C 1383 [30]) relies on the principle of variation in the stress wave (P-wave) propagation in the solid medium for identifying the presence of subsurface defects. A stress wave, generally induced on the surface of finite solid medium, propagates through the material with a certain speed and reflects back to the surface at a certain frequency referred to as resonance frequency. While the speed of propagation depends on the type of material, the frequency of reflection depends on the interaction of waves with the voids, cracks, dissimilar material interface etc. present in the material. The interaction results in an Echo which can be measured at the surface. Identifying the change in the pattern of the reflected waves' frequency reveals the presence of subsurface defects [31]. A typical impact echo set up employed in practice consists of three components, (1) mechanical impactor (steel balls), (2) a high-fidelity transducer, and (3) a data acquisition system (see Figure 3) [32]. A mechanical impactor generates an impact force and a transducer measures the surface displacement (resulting from reflected waves). The

surface displacement is represented in the form of voltage signal and is displayed in the data acquisition system as a voltage-time waveform data. At this juncture it is important to note that the distance between impactor and transducer is maintained in the range 0.3-0.4T, where T represents the depth of the reflecting interface. This range of distance ensures that the capture of shear waves (S-waves) is avoided. If C_p indicates the velocity of P-wave in concrete, f_T indicates the resonance frequency, then the depth of reflecting interface can be evaluated using Equation (1)

$$T = \frac{\beta C_p}{2 f_T} \tag{1}$$

where $\beta(= 0.96)$ is the correction factor and depends on the Poisson ratio of the material [33]. For determining f_T the waveform data is transformed into the frequency domain by using the Fast-Fourier transform (FFT).

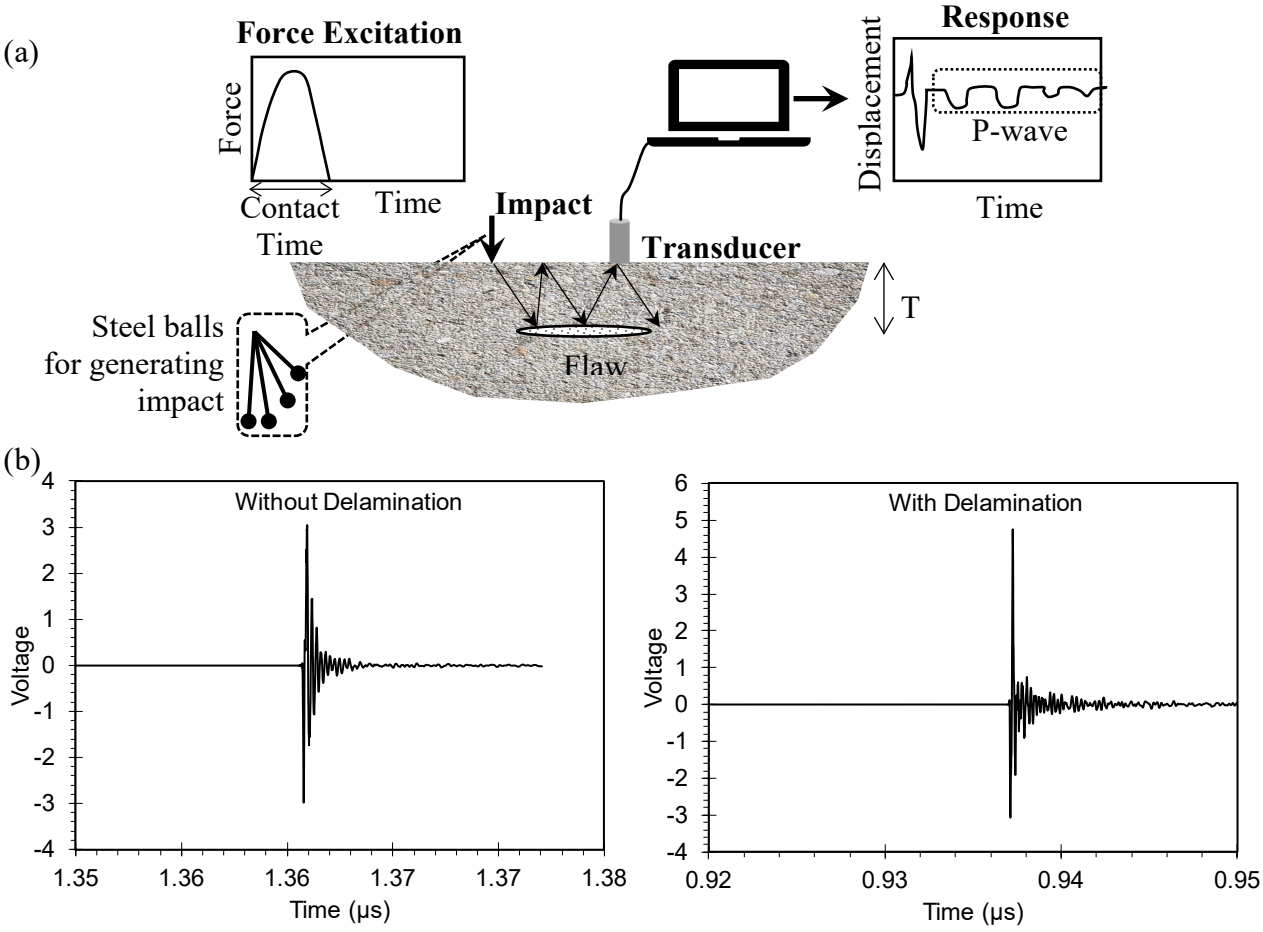

Figure 3. (a) Schematic of impact echo set up [31] and (b) signal of non-defective and defective bridge deck from SDNET database [29].

2.3 Ground Penetrating Radar (GPR)

Ground penetrating radar (GPR) (ASTM D 6087 [34]) utilizes the electromagnetic radiation in the radar range (10-1000 Hz) to infer the discontinuities (e.g. delamination, rebar) in the probed material. It works on the principle of change in the velocity of EM waves with the change in the

dielectric properties of the medium [23]. A typical GPR scenario in operation involves (1) transmitting of EM waves through the transmitter antenna into the probed medium and (2) receiving the waveform signals at the receiver antenna [35]. When the transmitted waves encounter the discontinuity or dissimilar material interface, the reflection-refraction phenomenon arises because of change in the dielectric properties of the medium (see Table 2). The reflected waves are captured by the receiver antenna and the refracted waves traverses further below the interface. The amplitude and phase of received signal is different from the one that is incident by the transmitter antenna. The propagation velocity, v of the EM wave is expressed as (see Equation (2)) [36]

$$v \approx \frac{c}{\sqrt{\varepsilon_r}} \tag{2}$$

where c is the velocity of light in vacuum and ε_r is the relative dielectric permittivity. Based on the two-way travel time, depth of the target can then be determined as (see Equation (3))

$$d = \frac{vt}{2} = \frac{ct}{2\sqrt{\varepsilon_r}} \tag{3}$$

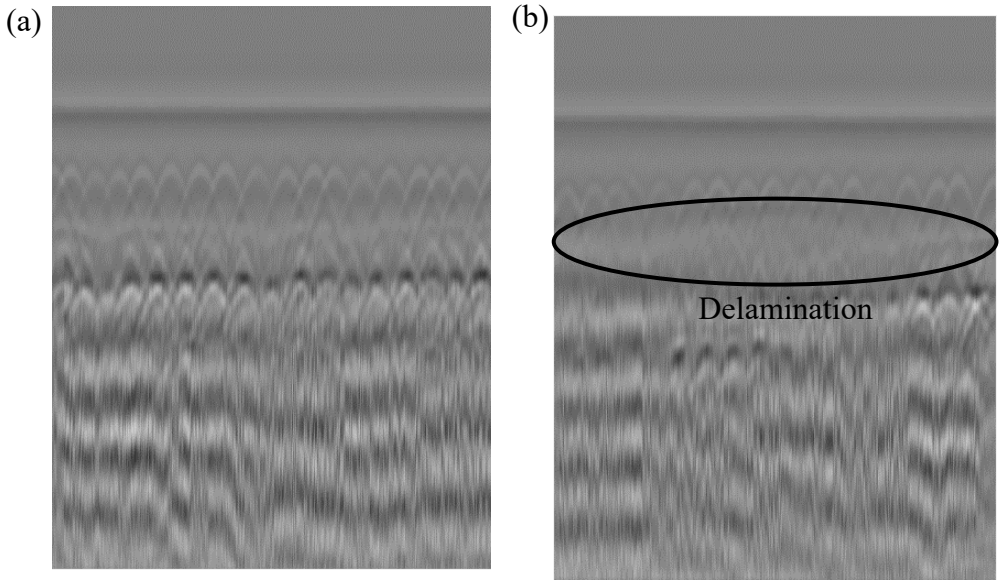

Figure 4. GPR B-Scan signals (a) non-defective and (b) defective

Table 2. Dielectric constant of various materials

Medium	Standard	Medium	Standard
Air	1	Sand	4-6
Water	81	Gravel	4-7
Ice	4	Clay	25-40
Concrete	8-10	Limestone	7
Asphalt	4-8	Glass	4-7

In practice, for instance bridge inspection, the GPR is accommodated in a handheld cart with transmitting and receiving antenna placed at a fixed distance. While the cart is rolled on the surface of the bridge deck, at a slow walking speed, the operation of transmitting and receiving the EM waves takes place. A signal recorded at single point is called trace or A-scan, typically quantifying time vs. amplitude. A series of A-scans recorded along the survey line and put together forms a B-scan (see Figure 4). Stacking multiple B-scans generates a 3D GPR referred to as C-scan. In the case of rebar detection, B-scan data reveals a downward hyperbolic reflection.

2.4 Electrical resistivity

Electrical resistivity technique measures the potential of the generated electric field in the probed material to characterize the corrosive environment (anodic and cathodic areas). A typical set up of this technique, referred to as Wenner set up, consists of four electrode probes spaced equally (see Figure 5(a)) [23]. While a current is applied between the two outer probes, the voltage is measured between the two inner probes. The higher the anodic and cathodic differential is in the probed material, the lower will be the ER of the concrete and higher will be the measured voltage. The ER of an oven dried concrete is around $10^6 \Omega.m$. However, when the pores of the concrete are saturated with the ionic solution the ER decreases (10 to $1000\ \Omega.m$) because of electrolytic conduction. The recorded ER magnitude could then be correlated to the severity of corrosion (see table). The ER (ρ) is evaluated using the Equation (4)

$$\rho = \frac{2\pi a V}{I} \tag{4}$$

where a is the spacing between the probe electrodes, V is the measured voltage, I is the applied current. The correlation between ER magnitude and corrosion rates are given as follows: very high corrosion (less than 5 k$\Omega - cm$); High corrosion (5-10 k$\Omega - cm$); Moderate to low corrosion (10-20 k$\Omega - cm$) and ; low corrosion (greater than 20 k$\Omega - cm$) [23].

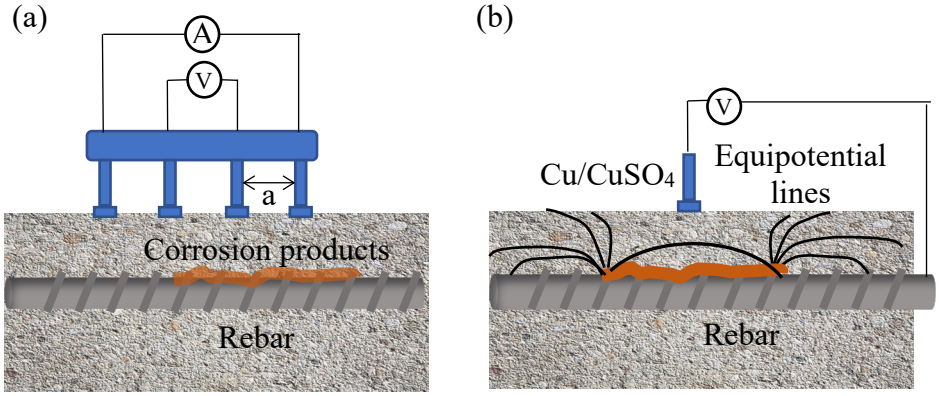

Figure 5. (a) Schematic of Wenner set up and (b) schematic of half-cell set up

2.5 Half-cell potential

Half-cell potential (HCP) technique involves measurement of the electrochemical potential difference between the embedded reinforcement and the standard reference electrode $Cu/CuSO_4$.

According to ASTM C 876 [37] a direct connection is required between the reinforcement and the attachment of the standard electrode to the bare concrete surface. Metal substrate oxidizes when it comes in contact with the electrolytic solution. Subsequently this will result in buildup of positive charge at anodic sites and negative charge on cathodic site. These electrons navigate to the cathode where they form hydroxide ions OH^-. A high impedance voltmeter is employed to measure the potential difference between the reinforcement and the standard electrode (see Figure 5(b)). The higher the potential, the higher the risk of corrosion occurrence. While CSE values greater than −200 mV indicates 90% probability of no corrosion, CSE values less than −350 mV indicates 90% probability of corrosion. For $-350\ \text{mV} \leq \text{CSE} \leq -200\ \text{mV}$, the probability of corrosion is 50% [38].

3. Methodology of the Survey

In this section, the methodology adapted to conduct the survey of state-of-the-art research in automated subsurface detection in RC bridges using deep learning is described. First a search strategy was employed to extract the resources from online database and then the relevant studies were screened by applying an inclusion and exclusion criteria. The resources used in this study mainly included peer-reviewed journal articles and conference articles. Three popular online databases namely 'Google Scholar', 'Web of Science' and 'IEEE Xplore' were employed to obtain the access to the articles. The number of papers found from these databases using specific keywords and their combinations is shown in scientometric Figure 6. Note that the search was restricted to the papers published after year 2002 i.e., last two decades. Since the terminology 'artificial intelligence', 'machine learning', 'deep learning', 'automatic identification' are interchangeably used in structural community for defect detection, all these terms were included to filter publications relevant to this survey paper. The following combinations of keywords are employed:

1. ('concrete Bridge' OR 'RC bridge' OR 'bridge deck') AND ('non-destructive evaluation' OR 'assessment');
2. ('concrete' AND 'bridge') AND ('thermography' OR 'GPR' OR 'impact echo') NOT ('prestress' OR posttension*);
3. ('concrete' AND 'bridge') AND ('machine learning' OR 'automation') AND ('thermography' OR 'GPR' OR 'impact echo') NOT ('prestress' OR posttension*);
4. (('concrete Bridge' OR 'RC bridge' OR 'bridge deck') AND 'deep learning') AND ('NDE' or 'thermography' OR 'GPR' OR 'impact echo') NOT ('prestress' OR posttension*).

In the first combination, a generic search involving non-destructive evaluation of concrete bridge decks is performed. In the second combination the search was restricted to specific NDE techniques that is of interest in sub-surface defect evaluation i.e., thermography, impact echo and GPR. In the third combination the search was further restricted to the use of machine learning or automated computer vision techniques for identifying the subsurface defects. In the fourth combination the terms 'machine learning' and 'automation' was replaced with 'deep learning' to explicitly find the papers that implemented state-of-the-art deep learning architectures.

The trend of number of publications employing above mentioned combination of keywords is shown in Figure 6. From Figure 6(a), for combination 1, it is noticed that the number of

publications increased every year. While a total of 3,161 publications were found from year 2002 to 2022, interestingly 40% of them were published in last four years. From Figure 6(b), for combination 2, it is observed that the number of publications pertaining to the use of impact echo, GPR and thermography techniques increased by two times since 2016. While a total of 307 publications were identified from year 2002 to 2022, a total of 53% of these publications were produced in last six years. From Figure 6(c), for combination 3, it is noticed that the machine learning or automation for detecting subsurface defects gained attention from year 2011. In total, 34 publications were identified. However, the number of publications were found to increase exponentially only from year 2017. From Figure 6(d), for combination 4, it is evident that deep learning received attention of researchers since 2018. In total, 15 publications were produced. All these are reviewed in this study to provide comprehensive information about the application of deep learning in RC bridge subsurface detection.

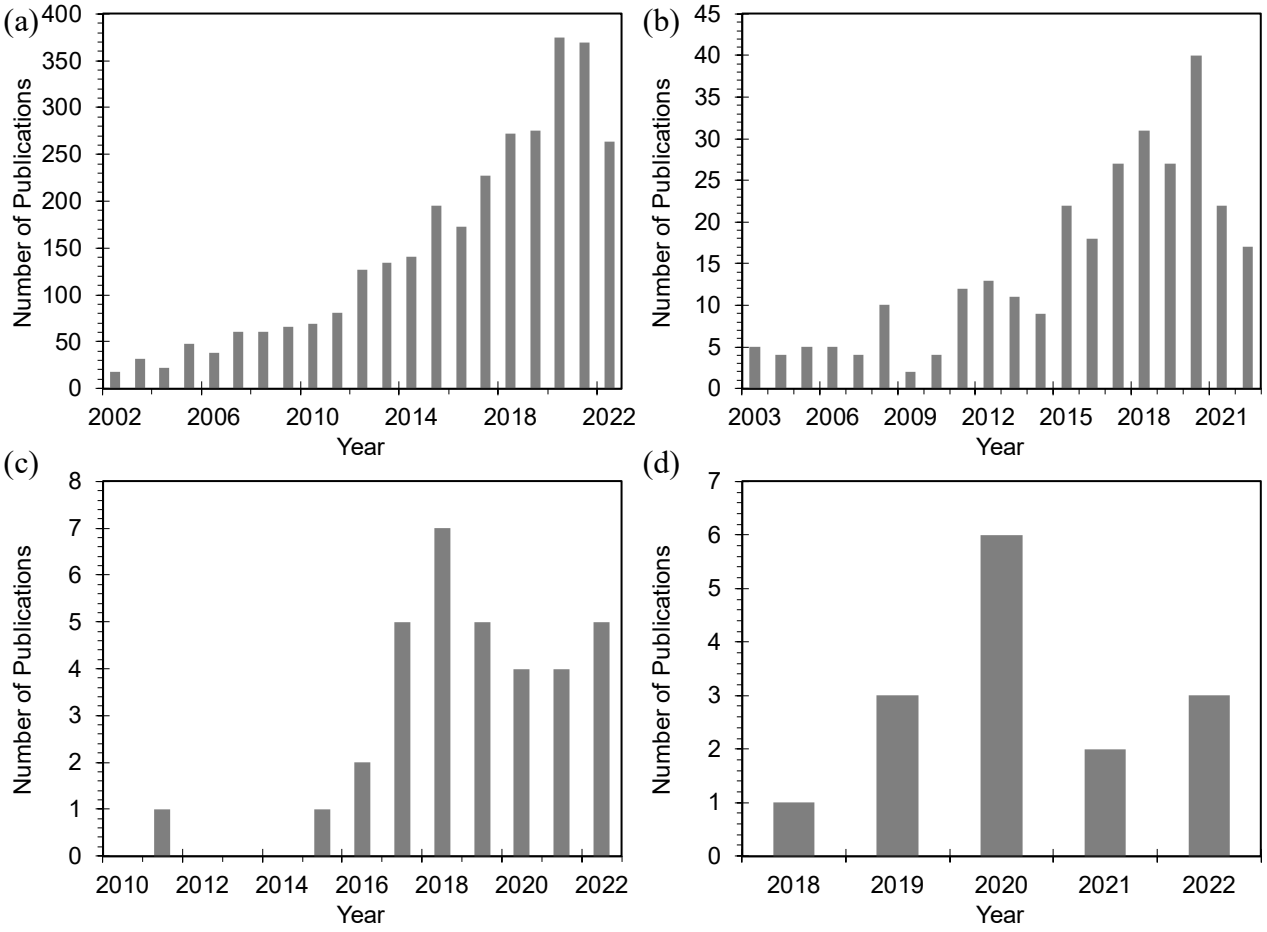

Figure 6. Trend of Publications from 2002-2022. (a) Combination 1 (number of publications 3100), (b) combination 2 (number of publications = 307 after filtering for IE, IRT and GPR), (c) Combination 3 (number of publications = 34 after filtering for machine learning algorithms in conjunction with IE, IRT and GPR), and (4) Combination 4 (number of publications = 15 after filtering for deep learning applications in IE, IRT and GPR).

4. Deep Learning (DL)

Deep learning is a sub-field of machine learning that utilizes the hierarchical architectures to learn the high-level abstractions present in the data [39]. DL is inspired by the mechanism of information processing carried out in the brain. In recent decade they gained popularity in various scientific disciplines such as cybersecurity, natural language processing, bioinformatics, robotics and control, medical information processing etc. Specifically, for performing wide range of complex cognitive tasks such as classification, clustering, dimensionality reduction, regression, video recommendation, spam detection etc [40,41]. One of the major reasons for the paradigm shift to DL is its ability to extract the discriminative features from the data. DL is also referred to as universal learner as it possesses the ability to learn approximate function in all domains. They are robust in terms of extracting relevant features automatically and inherent variations in the data. DL architectures could be generalized to various data types and applications by accounting for the concept of transfer learning. Furthermore, DL models are highly scalable. According to Alzubaidi et. al. [41] deep learning is a solution (1) when human experts are unavailable or fail to explain decision making, (2) the dataset is so huge that the interpretation is beyond the scope of human reasoning, (3) domain where adaption is important.

In the context of bridge condition assessment, specifically surface defects, DL was found to outperform the conventional ML techniques. For instance, in the case of concrete crack detection, Dorafshan et.al. [42] has shown that the best edge detector exhibited accuracy in the range of 53-79%, whereas deep convolutional neural network exhibited an accuracy of 86%. Gopalakrishnan et. al. [43]employed transfer learning in conjunction with ML algorithms such as neural networks (NN), support vector machine (SVM), random forest (RF), logistic regression (LR), and extremely randomized tree (ERT) on FHWA/LTTP database, and has demonstrated that accuracy improved in the range of 77% to 88%. Some [44] employed DL on roadway image database and concluded that DL has significant potential in automated pavement crack detection and classification. A comprehensive review of various DL algorithms for surface defect detection could be found elsewhere [45].

When compared to conventional machine learning algorithms, any deep learning architecture essentially comprises of several layers (referred to as computational units) incorporated to extract and transform the abstract information from the provided data to map the given input and the corresponding output. The most commonly used deep learning architectures can broadly be divided into four categories [46,47], namely (1) Convolutional Neural Networks (CNN), (2) Autoencoders, (3) Restricted Boltzmann Machine (RBM), and (4) Long Short-Term Memory (LSTM). In this section a general overview of these architectures is provided. At this juncture it is important to note that application of Restricted Boltzmann Machine and Autoencoder for subsurface defect detection is not found in the literature. However, their descriptions are provided for the sake of completeness.

4.1 Types of DL Architectures

4.1.1 Convolutional Neural Networks

CNNs are the most widely used architecture in diverse computer vision applications. Similar to traditional neural networks, CNNs also adapt forward propagation and back propagation algorithm to train all the parameters (weights and biases) of the network. A typical CNN architecture consists of three layers, (1) convolutional layers, (2) pooling layers, and (3) fully connected layers (see Figure 7) [41]. Each layer plays a different role. Convolutional layer utilizes kernel weights to convolve the image and generate various feature maps. Pooling layer is followed by convolutional layer and aims at subsampling or reducing the spatial dimensions of feature maps. Fully Connected layers are preceded by last pooling layer and converts the 2D feature map into 1D feature vector. Fully connected layers function similar to a neural network followed by output neurons. At this juncture it is important to note that there are various types of pooling strategies. Among them, 'average' and 'max' pooling are commonly employed. 'Max' pooling was found to lead to faster convergence during training operation. More details about various pooling operations and their influence on performance of CNN can be found elsewhere [41]. Examples of popular CNN architecture include AlexNet, LeNet, ResNet, Inception, Image Net, VGG and GoogLeNet (see Section 4.2). Primarily these architectures differ in their configuration by varying the depth of the network. For instance, AlexNet configuration consists of five convolutional layers and three fully connected layers, while GoogLeNet consists of 21 convolutional layers and one fully connected layer.

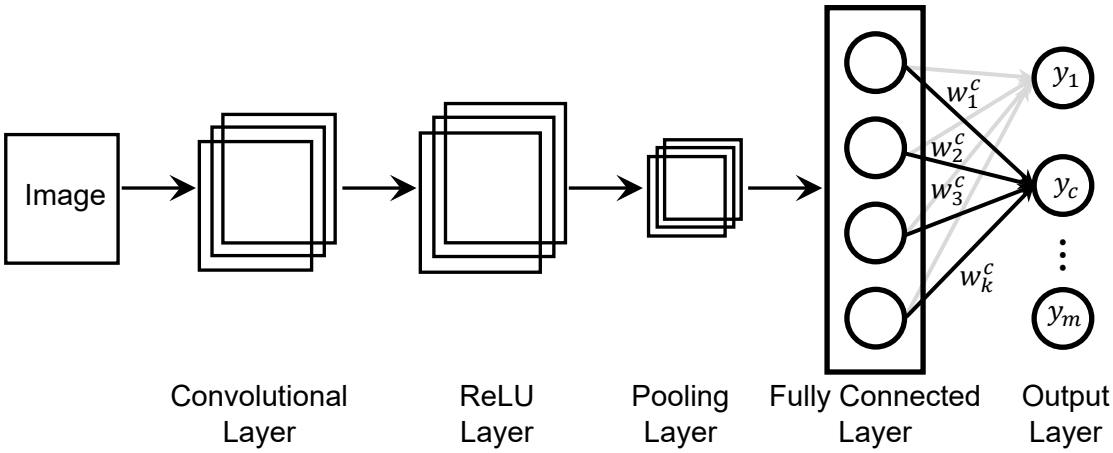

Figure 7. A typical CNN architecture

4.1.2 Restricted Boltzmann Machines

RBMs are probabilistic graphical neural network models that learns the hidden patterns in the input data [46]. In other words, the latent variables are extracted which are dimensionally reduced form of input data. RBMs are generative models. They have the simplest network possible and consists of two layers, namely visible layer and hidden layer. Every neuron in visible layer is connected to all the neurons in the hidden layer by weights and bias. RBM's objective is to find the joint probability distribution between visible and hidden layers that maximizes the log-likelihood function. For the purpose of learning the parameters (i.e., weights and bias) of the network RBM

uses the contrast divergence. A detailed discussion on training RBM could be found elsewhere. Employing basic RBM module three deep models were devised, (1) Deep Belief Networks (DBN), (2) Deep Boltzmann Machines (DBM) and (3) Deep Energy Models (DEM) (see Figure 8). While DBNs have undirected connections between the top two layers and directed connections in the bottom layers, DBMs have undirected connections between all layers of the network. Unlike DBN and DBM, DEMs only have a single layer of stochastic hidden units and rest being deterministic.

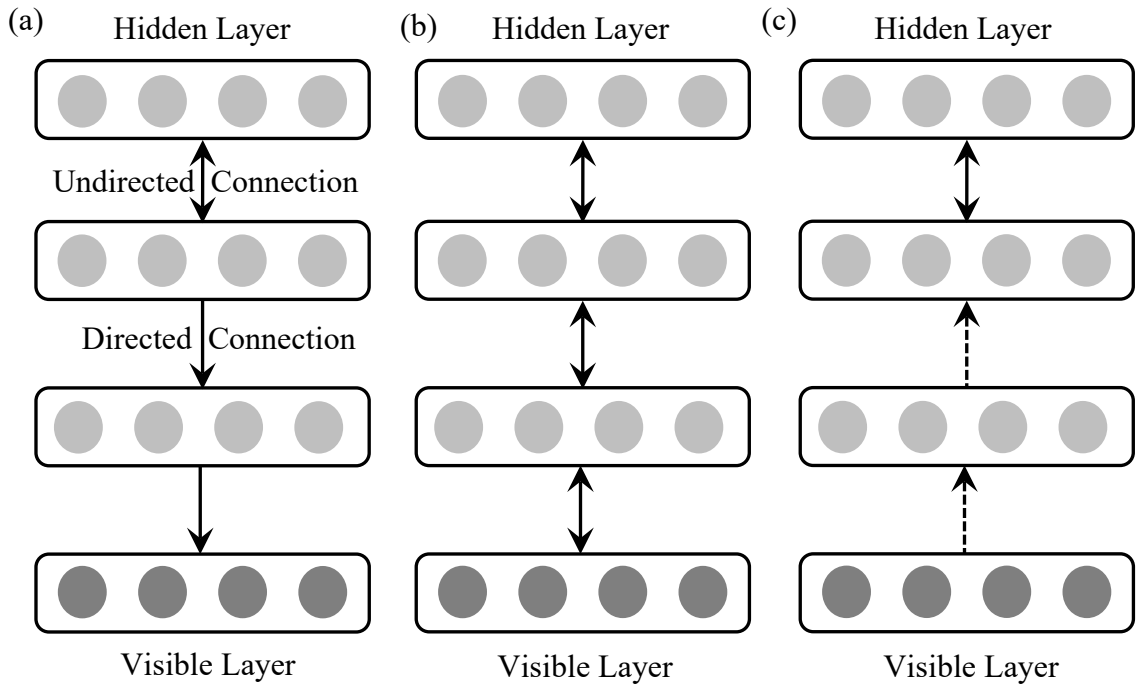

Figure 8. Schematic of Various Boltzmann Machines. (a) Deep Belief Networks, (b) Deep Boltzmann Machine, and (c) Deep Energy Model. (Note: the dashed arrows represent the connection of the deterministic hidden units)

4.1.3 Autoencoders

Similar to RBM, autoencoder is also an unsupervised learning algorithm that learns the latent variables or representation of the input data. The fundamental difference between RBM and autoencoder is that the RBMs are probabilistic and generative models whereas autoencoders are non-probabilistic and non-generative models [39]. Unlike principal component analysis (PCA) which yields linear representation of input data, autoencoders possess the ability to produce nonlinear representations. Autoencoders aims at reconstruction of its own inputs (Figure 9). In other words, the training process does not involve learning a mapping between input variables and target variables. A typical autoencoder architecture consists of encoder and decoder block (see Figure 9). While encoder segment comprises of network weights that extracts the latent variables of input, decoder segment comprises of weights that reconstructs the input data. Some of the well-known variants of autoencoders include sparse AE [48], Denoising AE [49], Variational AE [50], Convolutional AE [51] and Contractive AE [52].

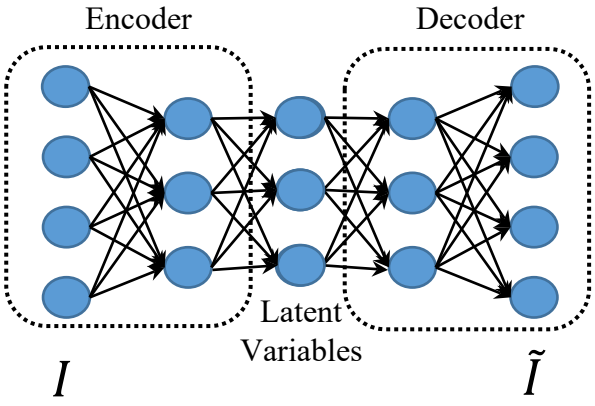

Figure 9. Typical architecture of autoencoder.

4.1.4 Long-Short Term Memory

LSTMs are invariant recurrent neural networks (RNN) applicable for sequential prediction problems that consists of data which relies on the context and the earlier state [46]. LSTM was introduced to overcome the vanishing/exploding gradient problem that persists in RNNs when learning long-term dependencies. In other words, RNNs fail to carry information if the sequence is long. In the context of the sensor data that involves time varying signals, RNNs may lose information that appears at the beginning of the signal. A typical LSTM consists of three parts referred to as forget gate, input gate and output gate which regulates the flow of information. Note that the gates are composed of sigmoid activation layer and pointwise multiplication operation. While forget gate identifies the relevance of holding information from the prior steps, the input gate determines the need for addition of information from the current step. The output gate concludes the next hidden state [20].

4.2 An Overview of State-of-the-art Architectures

The list of popular DL architectures that are being applied in various fields is composed in Table. However, in this subsection, only few of them are chosen that is of interest in the current context and their brief description is provided. Following are the architectures: (1) AlexNet, (2) GoogleNet (3) DenseNet, (4) ResNet (5) Xception, (6) MobileNet, and (7) DeepLab.

4.2.1 AlexNet

AlexNet was first proposed by Krizhevesky et. al. [53] for image recognition and classification. The idea behind devising this architecture was to extract more robust features that could generalize the learning ability of the algorithm. For this purpose, the depth of the layers were increased from five in LeNet to seven in AlexNet. Furthermore, to avoid problems of overfitting and saturation of activation, drop out scheme and ReLU activation functions were employed. When compared to LeNet, the size of the kernels adapted in AlexNet were 5×5 and 11×11. While first five layers are convolutional layers, last three layers are fully connected layers. For subsampling ‘Average Pooling’ was adapted and for the output ‘Softmax’ function was used.

4.2.2 GoogleNet

GoogleNet [54], also referred to as Inception -V1, was developed by Google with the aim of achieving an architecture that is computationally inexpensive and highly accurate. This architecture proposed use of different sizes of kernels within convolutional layer so that the information in the image at diverse spatial resolutions can be captured. GoogleNet also implemented sparse connections to resolve the redundant information problem. The backbone behind reducing computational cost was ignoring the irrelevant channels. By introducing global average pooling (GAP) layer in the place of fully connected (FC) layer the number of parameters were dropped from 40 to 5 million and the computational cost was decreased.

4.2.3 ResNet

ResNet [55] was emerged to resolve the problem of vanishing gradient during back propagation. This was achieved by introducing the concept of bypass pathway also referred to as residual block. Instead of training the network to map the inputs and outputs, ResNet architecture learns residual mapping. For instance, if $h(\boldsymbol{x})$ is the initial mapping, ResNet fits $f(\boldsymbol{x}) = h(\boldsymbol{x}) - \boldsymbol{x}$**.** If a network was to learn an identity function mapping that maps input to input, then $f(\boldsymbol{x})$ in ResNet becomes zero. The authors argue that learning a function equating to zero is relatively simple.

4.2.4 DenseNet

Similar to ResNet, DenseNet [56] was also emerged to resolve the problem of vanishing gradient during back propagation. However, DenseNet employed cross-layer connectivity i.e., each layer is connected to all the layers in the network through the feed-forward approach. With this the feature maps extracted in previous layers were incorporated as input in to the succeeding layers. DenseNet addressed one of the main limitations of ResNet i.e., the residual blocks fail to preserve features, and consequently might limit the representation power of the network.

4.2.5 Xception

Xception [57] is a deep convolutional neural network architecture that involves Depthwise Separable Convolutions. It is also known as “extreme” version of an Inception module. Depthwise separable convolution consists of two operations, namely, pointwise convolution and depthwise convolution. Unlike typical convolution operation in CNN architecture, Xception is computationally effective since it involves fewer convolution operations. There is one more difference between Inception and Xception. The presence or absence of a non-linearity after the first operation. In Inception model, both operations are followed by a ReLU non-linearity, however Xception doen't introduce any non-linearity.

4.2.6 MobileNet

MobileNet [58] architecture also employs Depthwise Separable Convolutions and is designed to be used in mobile applications. Every convolution layer is followed by a batch normalization and a ReLU. Two global hyperparameters are introduced to effectively reduce the computational cost further, namely, width multiplier (α) and resolution multiplier (ρ). While width multiplier is introduced to control the channel depth, resolution multiplier is introduced to control the input

image resolution of the network. Both α and ρ varies from 0 to 1, wherein the accuracy of model reduces as their magnitude drops from 1 to 0.

4.2.7 DeepLab

DeepLab [59] is a semantic segmentation architecture. It involves analysis and classification of each pixel into multiple class labels. Unlike all the above-mentioned architectures that were devised to classify images into specific classes, semantic segmentation automatically identifies distinct objects present within the image using the similarity of their characteristics. A typical semantic segmentation architecture consists of encoder network followed by a decoder network. While encoder comprises of pre-trained classification network like VGG [60] to extract the features, the task of the decoder is to semantically project the discriminative features onto the pixel space to get a dense classification. Unlike other segmentation algorithms such as FCN [61], SegNet [62] and U-Net [63], the DeepLab applies atrous convolution for up-sample. Atrous convolutions can capture information from a larger effective field of view while using the same number of parameters and computational complexity. In other words, the spatial resolution of feature maps is significantly reduced when deconvolution is performed with typical convolution operation.

5. Application of DL in NDE of Bridge Deck Subsurface

In this section the application of DL algorithms in conjunction with NDE techniques to identify subsurface defects is reviewed (see Table 3).

5.1 IRT

In this subsection, the performance of various deep learning model architectures employed in conjunction with IRT for delamination detection in RC bridge decks is provided. Based on the published literature, thresholding and semantic segmentation technique was employed to detect delamination. List of architectures implemented includes, Xception, DenseNet, Autoencoder and CNN. Zhang et. al. [64] performed automatic detection and quantification of delamination for concrete bridge decks. Infrared images were collected from four real bridge decks (McLaughlin et.al. [65]). Deep learning architecture was proposed and pixel-wise labeling was performed. Note that image labeling toolbox in MATLAB® was employed for annotating pixels. Xception blocks were used as backbone layers in Deep LabV3+ encoder-decoder architecture to build a stronger network. In total 261 images with delamination and 239 images without delamination was considered. As performance evaluation metrics accuracy and mean intersection of union were evaluated. The proposed model achieved accuracy of 99.36%, 97.96% and 97.83%; and mean intersection union of 0.98, 0.96 and 0.95 for training, validation and testing data respectively. Cheng et.al. [66] proposed CNN-based deep architecture, DenseNet, for segmenting delamination through thermography. The thermal images of concrete decks fabricated in the laboratory with different depths of delamination were acquired to train the DenseNet. Threshold of 0.5°C was chosen as the cut-ff to capture thermal contrast between delaminated and non-delaminated part. Crop and translation features were incorporated to include data augmentation. In total 960, 1200 and 240 images were employed for training, testing and validation purpose. While good performance was observed in terms of performance metrics precision (84.3%) and recall (85.02%), the unbalanced class was found to cause model bias. Also, increase in training data variety

enhanced the generalization of model. However, the hyper-parameter tuning and incorporation of additional data augmentation features needs further investigation. Additionally, the models were not tested or verified on data from real bridges which typically yields to reduced performance compared to the laboratory data.

Omar et. al. [27]carried out passive thermography of a full-scale deteriorated concrete deck bridge on-field in Quebec, Canada. FLIR T650sc thermal camera was employed to acquire the data and Flir software was used to enhance the image resolution. A mosaicked thermogram was created using stitching algorithm and then k-means clustering algorithm was employed to create a condition amp delineating the delamination portion on the bridge deck. While the proposed methodology detected the delamination at different survey times and environmental conditions, the depth or thickness of such defect was not indicated. Furthermore, the number of clusters must be set in advance. Mc Laughlin et al [65,67] employed MobileNetV2 architecture as an encoder to extract the features. Then DeepLab V3 is implemented to perform pixel-wise segmentation in infrared images. The mean intersection of Union of 82.7% was reported.

Sandra et. al. [68] investigated the performance of five CNN architectures VGG16, ResNet 18, ResNet 50, Xception and MobileNetV2. Based on case studies, highest recall value of 96.5% and F1-score of 85.2% was achieved by MobileNetV2. Considering the precision i.e., minimizing the false positive, VGG performed better and considering the recall MobileNetV2 and ResNet 50 performed better. Rubio et.al. [69] demonstrated the use of VGG-based fully convolutional neural network to perform semantic segmentation of thermal images. A mean accuracy of 89.7% and weighted F1-score of 81.9% was reported. Montaggiioli et. al. [70] detected corrosion and delamination in bridges using IR images. An automated image processing algorithm incorporating local thermal response of the structure was developed. Thermal images were transformed to grayscale images and local intensity weighting was applied to enhance contrast and steepness of gradient. Otsu thresholding was applied to remove background and conventional Canny edge detector was employed to detect damage. While temperature range of 16.2°C- 18.7°C highlighted defects under deck, the range 18.7°C-21.2°C highlighted the defects in the bean.

Table 3. Summary of various DL algorithms for detection of subsurface defects in bridge decks.

NDE	Ref	Dataset	DL architecture	Achievement	Limitation
Impact Echo		Lab (2464 instances) 2212 for training 252 for testing	1D-CNN 2D-CNN AlexNet GoogleNet, ResNet biLSTM	A: 0.88; F1-Score: 0.75 A: 0.86; F1-Score: 0.72 A: 0.84; F1-Score: 0.66 A: 0.83; F1-Score: 0.61 A: 0.80; F1-Score: 0.61	Limited number of training samples; no data augmentation; not deployed on field data; model performance needs to be validated on different training and testing ratios.
		Lab (3543 instances) 3282 for training 261 for testing	1D-CNN 2D-CNN AlexNet- full training AlexNet-transfer learning	A:0.82 (Concrete); 0.71(Asphalt) A:0.80 (Concrete); 0.45 (Asphalt) A:0.80(Concrete; 0.56 (Asphalt)	
IRT	Zhang	Dataset from McLaughlin	Xception and DeepLabV3+	IoU: 0.98, 0.95 and 0.96 for training, test and validation. Accuracy: 99.4%, 97.83% and 97.96% for training, test and validation.	Limited number of training samples; no data augmentation
	Cheng	Lab (1369 images) 70% training and 30 % testing	DenseNet	Non-augmented data IoU: 0.66-0.84 (validation); 0.55-0.7 (test). Augmented data IoU: 0.84-0.92 (validation); 0.68-0.8 (test)	Can't handle environmental noises (such as uneven dark colors, artificial heat/cold sources, and shadows etc.), and cause inaccurate predictions. Performance of other architectures for segmentation need further evaluations.
	Mc Laughlin	Field (500 images) 261 delamination; 239 sound	MobileNet (for feature extraction)	IoU: 0.76 (training); 0.73 (Validation); 0.75 (Testing)	Limited number of training samples; no data augmentation

			DeepLabV3 (for semantic segmentation)		
	Sandra	Field (517 images) 477 training and 40 testing	VGG16 ResNet 18 ResNet 50 MobileNet V2	A: 72.9% IoU:0.58 F1-score:0.84 A: 70.4% IoU:0.55 F1-score:0.83 A: 70.9% IoU:0.58 F1-score:0.83 A: 74.2% IoU:0.54 F1-score:0.85	Subjectivity in manual labeling; Limited number of training samples; no data augmentation
	Liu et.al.	Field 2370 for training 1622 for testing	SSD	Precision: 90.9% AUC: 0.94	Second layer rebars are not detected. Only horizontal flip and scaling are considered in data augmentation.
	Ahmed	Field 28,091 for training 5168 for validation	Deep Residual Network ResNet 152 DenseNet 161	A: 99.42 (training); 97.20 (validation) A:99.30 (training); 97.19 (validation)	The performance of model with varying ratio of training and test data needs to be validated. The block-based sliding window size can differ in real-world scenarios to accurately localize rebar signatures within the rebar detection and localization systems.
GPR	Xiang	Field 48 images are collected and split into small portions for training and testing	AlexNet	A: 94.51% (WS: 200x80) :87.78 % (WS: 150X50) :82.31% (WS: 250x100)	Detection accuracy varies with changing window sizes. Small dataset for training Feasibility of other DL models needs to be verified.

			:72.68% (WS: 120X30)	
Hou	Field 95 GPR scans 85% for training and 15% for testing	Mask R-CNN	A: 97% (training) Average precision: 0.42 Average Recall: 0.53	Requires a large dataset for training the model. The efficacy of the results depended on the quality of data related to soil condition (high water content)
Ahmadvand	Lab 17388 signals for training 2898 signals for validation 2898 signals for testing	1D-CNN	A: 84% Coefficient of variation: 0.04	Imbalanced dataset used for training. Not verified on field data. Model was not trained to detect rebar corrosion and rebar location

Note: Accuracy (A)= $\frac{TP+TN}{TP+FP+TN+FN}$; IoU=$\frac{TP}{TP+FP+FN}$;Precision = $\frac{TP}{TP+FP}$; Recall = $\frac{TP}{TP+FN}$;F1-Score=$\frac{2(Precision \times Recall)}{Precision+Recall}$; WS – window size.

5.2 Impact Echo

In this subsection, the performance of various deep learning model architectures employed in conjunction with impact echo for delamination detection in RC bridge decks is provided. Based on the published literature, impact echo technique is used in conjunction with CNN to detect delamination. Dorafshan et. al. [71] evaluated bridge decks with overlays using 1D and 2D CNN. AlexNet was implemented with full training and transfer learning. Experimental deck specimens were generated with different overlay systems (bonded and debonded) and the data was acquired using impact echo. In total, 700, 570 and 2000 instances of defect, debond and sound class labels were used for training purpose. Note that impact echo data is a one-dimensional signal. To perform 2D CNN, the one-dimensional signal was converted to 2D CNN (also referred to as spectrogram) using short-time Fourier Transform with sliding window of 50% overlap. While cement overlay system yielded an accuracy of 0.68, asphalt overlay yielded an accuracy of 0.58. 1D CNN exhibited more promising results compared to 2D CNN. In another study [72], the authors have reported the feasibility of deep learning models for defect detection using impact echo on decks without overlay, bare decks. Specifically, 1D CNN, 2D CNN (AlexNet, GoogleNet and ResNet) and recurrent neural network using bidirectional long-short term memory units were investigated. In 2D CNN transfer learning was implemented wherein only the weights in the fully connected layers are trained. However, for 1D CNN and biLSTM full training was carried out. 1D CNN exhibited highest overall accuracy of 0.88 followed by 2D CNN (83-86%) and biLSTM (80%).

5.3 GPR

In this subsection, the performance of various deep learning model architectures employed in conjunction with GPR for rebar detection and localization in RC bridge decks is provided. Interpretation of hyperbolic signature is the most crucial step for detecting and locating the rebar in GPR data. Based on the published literature, different deep learning models explored in the past includes AlexNet, Multi-Layer Perceptron (MLP), SSD, Deep Residual Networks and Mask R-CNN.

Liu et. al. [73] employed Single Shot Multibox Detector (SSD) model to detect and localize the rebar in GPR B-scan. Firstly, SSD was trained and hyperbolic signatures in the scan was detected. Then the migration and binarization operations were carried out on the GPR data to identify the location or depth of the bar. While the idea of migration is to collapse the hyperbolic signature to concentrated blobs at their apex, binarization aids in highlighting the pixels which are utilized to locate the rebar. Details about migration and binarization concetps could be found elsewhere and is avoided here for the sake of brevity. In total 2370 GPR images were used for training purpose and 1622 images were used for testing purpose. An average precision of 90.9% was achieved for detection of hyperbolic signature. The area under the Receiver Operating Characteristic (ROC) curves (AUCs) of SSD was calculated to be 0.94. For locating the depth of rebar, the error in estimation was found to be less than 1.5mm.

Ahmed et. al. [74] devised an approach that automatically detects and localizes rebar in concrete. Specifically, deep residual network and k-means clustering technique (effective separation between the foreground and background regions within the GPR images) was employed. Data acquisition was carried out from 9 bridges using GPR. Architectures ResNet-18, ResNet-34,

ResNet-50, ResNet-101 and ResNet-152 frameworks were investigated and influence of hyper tuning parameters (e.g. e.g. number of epochs, batch size, and number of layers) on the performance of proposed rebar detection and localization system was assessed. The proposed method outperformed state-of-the art methods with an overall accuracy of 91.91%, IoU of 0.9 and F1-Score of 95.58%. Furthermore, the performance metrics were found to be positively correlated with the number of layers.

Xiang et. al. [75] employed AlexNet architecture for automatic detection of rebar in GPR scan. The configuration of the architecture consisted of 5 convolutional layers and 3 fully connected layers. ReLU activation function was used and 'max' pooling was implemented. For the purpose of comparison another deep learning model TraNet was constructed. TraNet configuration consisted of 5 layers in which every alternate layer is a convolutional layer and pooling layer. Upon investigating the influence of different window sizes (120×30, 150×50, 200×80 and 250×100), it was observed that AlexNet outperformed TraNet. An overall accuracy of 94.5% was achieved for window size of 200× 80.

Hou et. al. [76] implemented mask R-CNN architecture to detect and segment anomaly/ hyperbolic signatures from GPR scans. A novel anchoring scheme is proposed and integrated with Mask R-CNN to improve performance of anomaly or object detection. Anchors extract the regions of hyperbolic signatures and locate the bounding boxes. Authors have obtained anchors by screening feature map using sliding window. The performance of the proposed model is found to be superior to the HOG-based feature extracted SVM classifier model. Transfer learning is used in the training process.

The common use of GPR has been to localize reinforcement bars and their level of corrosion in RC bridge decks. Ahmadvand et al. [77] took a new approach for using GPR in bridge deck evaluation. They hypothesized that presence of any irregularities, such as delamination not induced by corrosion, will interrupt the GPR signals due to change in dielectric constant of delaminated versus sound concrete. GPR data collected from several RC bridge deck specimens with artificial defects were used to train a one-dimensional CNN and other Machine Learning algorithms. The model was trained on 17,388 GPR signals, validated on 2898 GPR signals, and tested on 2898 GPR signals. The GPR dataset was imbalanced with 80% of signals collected from a sound region and 20% from delaminated region. A 5-fold training and validation procedure was adapted where the whole dataset was spilt in 8 equal segments. In each round, 6/8, 1/8, and 1/8 of GPR data was for training, validation, and testing, respectively. This was repeated eight times until all split was used for testing exactly one time. In each round, the performance of the models was described with an average value and a coefficient of variation value. The 1DCNN was superior to SVM and KNN models by achieving an average accuracy of 84%. The proposed 1DCNN also produced the most consistent result by achieving a coefficient of variation of 0.04.

6. Synthesis

This section is dedicated to summarize the achievements, draw attention to drawbacks, and to outline the future research and practice need to further use deep learning for bridge deck evolution.

6.1 Potentials and Achievements

Deep learning models have shown remarkable performance when properly applied on NDE data. CNNs were used on timeseries, image representations of timeseries of IE data and showed promising result by being superior to conventional IE data interpretation method. The same trend was observed when deep learning was developed, trained, and tested for GPR classification. Within the IRT image segmentation, the DL was found to outperform the conventional thresholding methods. The review of published records of using deep learning models showed that their consistency and high accuracy could potentially address the two bottlenecks of widespread use of NDE for bridge deck evaluation: dependency on inspector's skills, low detection rate (especially with IRT). Based on the review, deep learning models can be considered a powerful tool for evaluation of reinforced concrete bridge decks using NDE since:

- Deep learning can reduce or eliminate the need for skilled operators interpreting the NDE data to identify subsurface deterioration. Automation in data interpretation is crucial since conventional methods derived from a specific NDE's physic-based features are not always successful in subsurface defect detections. For instance, conventional image processing can only be 70% accurate if white spots in thermal images shown in Figure 2 are segmented as delamination [78].
- Consistency of data interpretation among different bridge decks (spatial), and the same bridge deck at different times of inspection is substantially more achievable when deep learning models are used compared to conventional data analysis techniques.
- Implementation of deep learning models in long-term bridge deck management has the potential to generate more informed bridge deterioration models;
- Data fusion for concurrent analysis of heterogenous can be achieved through deep learning which is a clear future need since bridge decks are composite and complicated systems;
- The proper adaptation of deep learning in bridge deck inspection is aligned resiliency and sustainability goals of transportation infrastructure in the U.S.

6.2 Challenges, Limitations and Future Need

In this survey paper a brief overview of the popular NDE methods for subsurface defect detection in bridge decks is provided and the various DL models developed to automate the identification of defects is explored. Specifically, the working principles of NDE methods is described along with their limitations and the efficacy of DL models is highlighted. In total, 15 publications were reviewed to obtain the information about application of DL models in bridge deck subsurface inspection. This information is categorized based on their integration with the specific sensor type i.e., IRT, IE and GPR. While it is evident that there is still lot of scope for implementation and development of DL models in the field of bridge inspection, some general challenges and future prospects are highlighted below.

- Deep learning is known to generate the generalized model that predicts the target variables reasonably well on unseen data. However, it needs to be ensured that: (1) dataset is as large as possible and is of high quality, and (2) the data set covers the wide range of real-world scenarios. Currently, the efficacy of proposed DL models is either validated on the lab data or the dataset size that is small. It may be practically difficult to capture wide

spectrum of uncertainties in the lab data which are known to arise in real-field. For instance, the temperature effects, wind effects, environmental effects etc. Therefore, the generalizability of proposed deep learning models needs to be verified for the unseen data in the real-world scenario.

- The size of the available NDE datasets for the bridge subsurface defects is very limited. Unlike the crack detection problem wherein annotated open source datasets are available to train deep learning models, these datasets are rare for subsurface defects. This is mainly due to the fact that annotation of subsurface defect datasets requires some level of destructive verifications. Although data augmentation techniques could enhance the learning of DL architectures, the scope of capturing various uncertainties that could affect the prediction may not be possible. While creating a real-world scenario in the lab is difficult, it is also challenging to acquire data in real field because it is not easy to know when the uncertainties would arise. Therefore, an open source dataset needs to be created by archiving the sensor data obtained through various resources around the globe which incorporates factors like age of bridge, weather conditions, material degradation properties, sensor specifications etc.
- DL involves determination of millions of parameters referred to as weights in the network. Often it may be essential to increase the depth of the network to capture the complex nature of the problem. Furthermore, tuning of hyperparameters such as activation functions, regularization parameters, learning rate etc. may also be required. In such case the computational burden may increase and the current computational resources may not be enough. This hinders the development of accurate model. In other words, a compromised DL model at the cost of unexplored combinations of hyperparameters will be obtained.
- It may be time consuming to train the DL model network from the scratch. Transfer learning and other domain adaptation methods can facilitate faster learning by determining the weight in only the last few layers of a model fully trained on large image datasets (such as ImageNet). Using transfer learning the model can be retrained when new data is acquired.
- DL should be implemented to find hidden features in NDE data that are commonly neglected. However, these features are only effective if they are extracted based on realistic and validated NDE data. NDE data annotation can be implemented for bridges scheduled for deck repairs. The process normally involves removal of overlay, chain-dragging or milling the deck to identify susceptible regions to delamination, removal of concrete at the susceptible regions, chain-dragging or milling of remainder concrete in the susceptible regions to remove possible deeper delamination. The process continues until no region is marked as susceptible. If NDE data were collected from the same hypothetical deck before the repair process, one can annotate them almost automatically as it has been done for IRT images [79]
- DL architectures are considered as black box models because of inherent complexity. It is difficult to interpret why and how a certain prediction is made. Prior to the deployment of trained DL model, it is also important to know that the outcome of model is interpretable. This will ensure that the physics of the deterioration is captured by the model. Therefore,

the explainable algorithms [80,81] in conjunction with DL model needs to be explored to carry out the inspections in real time.

- It is important to deploy the learned DL models in the real time in conjunction with UAS/ robots to detect the defects on field. This invites for the interdisciplinary collaboration to develop embedded systems for on field applications.
- No single sensor will suffice capturing the defects accurately. Every NDE technique is associated with certain limitations. An integrated system of sensors may improve the condition assessment of bridges. However, fusion of data acquired from different sensors may be a challenge because of their heterogeneous forms and inconsistent scale. For instance, impact echo provides a time-series data which is obtained at discrete locations of bridge, while IRT yields a temperature contour data which is continuous along the bridge. Therefore, there is a need for development of efficient fusion techniques that accommodates heterogeneous data types from multiple sensors and enhance the reliability of prediction.

7. Conclusions

In this survey paper a brief overview of the popular NDE methods for subsurface defect detection in bridge decks is provided and the various DL models developed to automate the identification of defects is explored. Specifically, the working principles of NDE methods is described along with their limitations and the efficacy of DL models is highlighted. In total, 15 publications were reviewed to obtain the information about application of DL models in bridge deck subsurface inspection. This information is categorized based on their integration with the specific sensor type i.e., IRT, IE and GPR. The DL models were found to exhibit superior performance in terms of prediction accuracy for IRT, IE and GPR. However still there is lot of scope for improvement in the context of subsurface defect assessment prior to their deployment in the real-world scenario. The associated challenges and future needs are highlighted.

Data Availability Statement

Data available on request from the authors.

Conflict of Interest

Authors have no conflict of interest relevant to this article**.**

Acknowledgements

The authors would like to acknowledge the financial support from College of Engineering and Mines, and Office of Vice-President for Research at University of North Dakota that supported this research. Furthermore, the help of two PhD students Eberechi Ichi and Faezeh Jafari for processing the data is highly appreciated.

References

[1] FHWA. Bridges & Structures - Federal Highway Administration n.d. https://www.fhwa.dot.gov/bridge/ (accessed July 28, 2022).

[2] Federal Highway Administration, Technology USD of T. National Transportation Atlas Database (NTAD): National Bridge Inventory (NBI) 2008-Present Datasets. United States. Department of Transportation. Bureau of Transportation Statistics; 2020. https://doi.org/10.21949/1520766.

[3] Ahmed H, La HM, Gucunski N. Review of Non-Destructive Civil Infrastructure Evaluation for Bridges: State-of-the-Art Robotic Platforms, Sensors and Algorithms. Sensors 2020, Vol 20, Page 3954 2020;20:3954. https://doi.org/10.3390/S20143954.

[4] FHWA. National Bridge Inventory - Federal Highway Administration 2021. https://www.fhwa.dot.gov/bridge/nbi/no10/condition21.cfm (accessed July 29, 2022).

[5] Kashif Ur Rehman S, Ibrahim Z, Memon SA, Jameel M. Nondestructive test methods for concrete bridges: A review. Constr Build Mater 2016;107:58–86. https://doi.org/10.1016/J.CONBUILDMAT.2015.12.011.

[6] Akgul F. Inspection and evaluation of a network of concrete bridges based on multiple NDT techniques. Struct Infrastruct Eng 2020;17:1–20. https://doi.org/10.1080/15732479.2020.1790016/FORMAT/EPUB.

[7] Gucunski N, Pailes B, Kim J, Azari H, Dinh K. Capture and Quantification of Deterioration Progression in Concrete Bridge Decks through Periodical NDE Surveys. J Infrastruct Syst 2017;23. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000321.

[8] Tsiatas G, Robinson J. Durability Evaluation of Concrete Crack Repair Systems: Https://DoiOrg/103141/1795-11 2002:82–7. https://doi.org/10.3141/1795-11.

[9] Olin BD, Meeker WQ. Applications of statistical methods to nondestructive evaluation. Technometrics 1996;38:95–112. https://doi.org/10.1080/00401706.1996.10484451.

[10] Chakraborty D, McGovern ME. NDE 4.0: Smart NDE. 2019 IEEE Int Conf Progn Heal Manag ICPHM 2019 2019. https://doi.org/10.1109/ICPHM.2019.8819429.

[11] Omar T, Nehdi ML. Condition Assessment of Reinforced Concrete Bridges: Current Practice and Research Challenges. Infrastructures 2018, Vol 3, Page 36 2018;3:36. https://doi.org/10.3390/INFRASTRUCTURES3030036.

[12] Tenžera D, Puž G, Radiæ J. Visual inspection in evaluation of bridge condition. Gradjevinar 2012;64:717–26. https://doi.org/10.14256/JCE.718.2012.

[13] Omar T, Asce SM, Nehdi ; M L, Zayed T, Asce M. Integrated Condition Rating Model for Reinforced Concrete Bridge Decks. J Perform Constr Facil 2017;31:04017090. https://doi.org/10.1061/(ASCE)CF.1943-5509.0001084.

[14] Abdelkhalek S, Asce SM, Zayed T, Asce F. Comprehensive Inspection System for Concrete Bridge Deck Application: Current Situation and Future Needs. J Perform Constr Facil 2020;34:03120001. https://doi.org/10.1061/(ASCE)CF.1943-5509.0001484.

[15] Billah UH, La HM, Gucunski N, Tavakkoli A. Classification of Concrete Crack using Deep Residual Network. 9 th Int. Conf. Struct. Heal. Monit. Intell. Infrastruct., St. Louis: 2019.

[16] Li G, Zhao X, Du K, Ru F, Zhang Y. Recognition and evaluation of bridge cracks with modified active contour model and greedy search-based support vector machine. Autom Constr 2017;78:51–61. https://doi.org/10.1016/J.AUTCON.2017.01.019.

[17] Zhou S, Song W. Deep learning-based roadway crack classification using laser-scanned range images: A comparative study on hyperparameter selection. Autom Constr 2020;114:103171. https://doi.org/10.1016/J.AUTCON.2020.103171.

[18] Maeda H, Sekimoto Y, Seto T, Kashiyama T, Omata H. Road Damage Detection and Classification Using Deep Neural Networks with Smartphone Images. Comput Civ Infrastruct Eng 2018;33:1127–41. https://doi.org/10.1111/MICE.12387.

[19] Sakagami T. Remote nondestructive evaluation technique using infrared thermography for fatigue cracks in steel bridges. Fatigue Fract Eng Mater Struct 2015;38:755–79. https://doi.org/10.1111/FFE.12302.

[20] Zhang Q, Barri K, Babanajad SK, Alavi AH. Real-Time Detection of Cracks on Concrete Bridge Decks Using Deep Learning in the Frequency Domain. Engineering 2021;7:1786–96. https://doi.org/10.1016/J.ENG.2020.07.026.

[21] Prasanna P, Dana KJ, Gucunski N, Basily BB, La HM, Lim RS, et al. Automated Crack Detection on Concrete Bridges. IEEE Trans Autom Sci Eng 2016;13:591–9. https://doi.org/10.1109/TASE.2014.2354314.

[22] Lim RS, La HM, Sheng W. A robotic crack inspection and mapping system for bridge deck maintenance. IEEE Trans Autom Sci Eng 2014;11:367–78. https://doi.org/10.1109/TASE.2013.2294687.

[23] Gucunski N, Romero F, Imani A, Fetrat F. Nondestructive Testing to Identify Concrete Bridge Deck Deterioration. 2013.

[24] Rashidi M, Azari H, Nehme J. Assessment of the overall condition of bridge decks using the Jensen-Shannon divergence of NDE data. NDT E Int 2020;110:102204. https://doi.org/10.1016/J.NDTEINT.2019.102204.

[25] Gucunski N, Kee SH, La H, Basily B, Maher A, Ghasemi H. Implementation of a Fully Autonomous Platform for Assessment of Concrete Bridge Decks RABIT. Struct Congr 2015 - Proc 2015 Struct Congr 2015:367–78. https://doi.org/10.1061/9780784479117.032.

[26] Oh T, Kee S-H, Arndt RW, Popovics JS, Asce M, Zhu J. Comparison of NDT Methods for Assessment of a Concrete Bridge Deck. J Eng Mech 2013;139:305–14. https://doi.org/10.1061/(ASCE)EM.1943-7889.0000441.

[27] Omar T, Nehdi ML, Zayed T. Infrared thermography model for automated detection of delamination in RC bridge decks. Constr Build Mater 2018;168:313–27. https://doi.org/10.1016/J.CONBUILDMAT.2018.02.126.

[28] ASTM D 4788. Standard Test Method for Detecting Delaminations in Bridge Decks Using Infrared Thermography. ASTM Int 2007.

[29] Ichi E, Dorafshan S. SDNET2021: Annotated NDE dataset for Structural Defects. Datasets 2021. https://doi.org/10.31356/data019.

[30] ASTM C 1383. Standard Test Method for Measuring the P-Wave Speed and the Thickness

of Concrete Plates Using the Impact-Echo Method. ASTM Int 2015.

[31] Carino N. Impact Echo: The Fundamentals . Int. Symp. Non-Destructive Test. Civ. Eng., Berlin, Germany: 2015.

[32] Carino by N, Carino NJ. THE IMPACT-ECHO METHOD: AN OVERVIEW The Impact-Echo Method: An Overview 1. Proc. 2001 Struct. Congr. Expo., 2001.

[33] Scherr JF, Grosse CU. Delamination detection on a concrete bridge deck using impact echo scanning. Struct Concr 2021;22:806–12. https://doi.org/10.1002/SUCO.202000415.

[34] ASTM D 6087. Standard Test Method for Evaluating Asphalt-Covered Concrete Bridge Decks Using Ground Penetrating Radar. ASTM Int 2015.

[35] Travassos XL, Pantoja F M, Ida N. Ground Penetrating Radar: Improving sensing and imaging through numerical modeling. The Institution of Engineering and Technology; 2021. https://doi.org/10.1049/PBCE115E.

[36] Merkle D, Frey C, Reiterer A. Fusion of ground penetrating radar and laser scanning for infrastructure mapping. J Appl Geod 2021;15:31–45. https://doi.org/10.1515/JAG-2020-0004/MACHINEREADABLECITATION/RIS.

[37] ASTM C 876. Standard Test Method for Corrosion Potentials of Uncoated Reinforcing Steel in Concrete. ASTM Int 2015.

[38] Satish V.L. Half Cell Potential Testing for Durability Studies of Concrete Structures in Coastal Environment. NDE 2020 - Virtual Conf. Exhib., 2020.

[39] Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: A review. Neurocomputing 2016;187:27–48. https://doi.org/10.1016/J.NEUCOM.2015.09.116.

[40] Sengupta S, Basak S, Saikia P, Paul S, Tsalavoutis V, Atiah F, et al. A review of deep learning with special emphasis on architectures, applications and recent trends. Knowledge-Based Syst 2020;194:105596. https://doi.org/10.1016/J.KNOSYS.2020.105596.

[41] Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. J Big Data 2021 81 2021;8:1–74. https://doi.org/10.1186/S40537-021-00444-8.

[42] Dorafshan S, Thomas RJ, Maguire M. Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. Constr Build Mater 2018;186:1031–45. https://doi.org/10.1016/J.CONBUILDMAT.2018.08.011.

[43] Gopalakrishnan K, Khaitan SK, Choudhary A, Agrawal A. Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. Constr Build Mater 2017;157:322–30. https://doi.org/10.1016/J.CONBUILDMAT.2017.09.110.

[44] Liene S. Automatic image-based road crack detection methods LIENE SOME KTH ROYAL INSTITUTE OF TECHNOLOGY SCHOOL OF ARCHITECTURE AND THE BUILT ENVIRONMENT. KTH Royal Institute of Technology, 2016.

[45] Ali L, Alnajjar F, Khan W, Serhani MA, Al Jassmi H. Bibliometric Analysis and Review of Deep Learning-Based Crack Detection Literature Published between 2010 and 2022.

Build 2022, Vol 12, Page 432 2022;12:432. https://doi.org/10.3390/BUILDINGS12040432.

[46] Shrestha A, Mahmood A. Review of deep learning algorithms and architectures. IEEE Access 2019;7:53040–65. https://doi.org/10.1109/ACCESS.2019.2912200.

[47] Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep Learning for Computer Vision: A Brief Review. Comput Intell Neurosci 2018;2018. https://doi.org/10.1155/2018/7068349.

[48] Aurelio M', Poultney RC, Chopra S, Lecun Y. Efficient Learning of Sparse Representations with an Energy-Based Model n.d.

[49] Vincent P, Larochelle H, Bengio Y, Manzagol PA. Extracting and composing robust features with denoising autoencoders. Proc 25th Int Conf Mach Learn 2008:1096–103. https://doi.org/10.1145/1390156.1390294.

[50] Pu Y, Gan Z, Henao R, Yuan X, Li C, Stevens A, et al. Variational Autoencoder for Deep Learning of Images, Labels and Captions n.d.

[51] Zhang Y. A Better Autoencoder for Image: Convolutional Autoencoder n.d.

[52] Rifai S, Vincent P, Muller X, Glorot X, Bengio Y. Contractive Auto-Encoders: Explicit Invariance During Feature Extraction 2011.

[53] Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. Commun ACM 2017;60. https://doi.org/10.1145/3065386.

[54] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2015;07-12-June-2015:1–9. https://doi.org/10.1109/CVPR.2015.7298594.

[55] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2016;2016-December:770–8. https://doi.org/10.1109/CVPR.2016.90.

[56] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017 2017;2017-January:2261–9. https://doi.org/10.1109/CVPR.2017.243.

[57] Chollet F. Xception: Deep learning with depthwise separable convolutions. Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017 2017;2017-January:1800–7. https://doi.org/10.1109/CVPR.2017.195.

[58] Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications n.d.

[59] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans Pattern Anal Mach Intell 2016;40:834–48. https://doi.org/10.48550/arxiv.1606.00915.

[60] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 3rd Int Conf Learn Represent ICLR 2015 - Conf Track Proc 2014. https://doi.org/10.48550/arxiv.1409.1556.

[61] Shelhamer E, Long J, Darrell T. Fully Convolutional Networks for Semantic Segmentation. IEEE Trans Pattern Anal Mach Intell 2014;39:640–51. https://doi.org/10.48550/arxiv.1411.4038.

[62] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Trans Pattern Anal Mach Intell 2015;39:2481–95. https://doi.org/10.48550/arxiv.1511.00561.

[63] Weng W, Zhu X. U-Net: Convolutional Networks for Biomedical Image Segmentation. IEEE Access 2015;9:16591–603. https://doi.org/10.48550/arxiv.1505.04597.

[64] Qianyun Z, Kaveh Barri, Zhe Wan. A Deep Learning-based Autonomous System for Detection and Quantification of Delamination on Concrete Bridge Decks. Int. Bridg. Conf., Washington DC: 2021.

[65] McLaughlin E, Charron N, Narasimhan S. Combining deep learning and robotics for automated concrete delamination assessment. Proc 36th Int Symp Autom Robot Constr ISARC 2019 2019:485–92. https://doi.org/10.22260/ISARC2019/0065.

[66] Cheng C, Shang Z, Shen Z. CNN-Based Deep Architecture for Reinforced Concrete Delamination Segmentation through Thermography. Comput Civ Eng 2019 Smart Cities, Sustain Resil - Sel Pap from ASCE Int Conf Comput Civ Eng 2019 2019:50–7. https://doi.org/10.1061/9780784482445.007.

[67] McLaughlin E, Charron N, Narasimhan S. Automated Defect Quantification in Concrete Bridges Using Robotics and Deep Learning. J Comput Civ Eng 2020;34:04020029. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000915.

[68] Pozzer S, Azar ER, Rosa FD, Pravia ZMC. Semantic Segmentation of Defects in Infrared Thermographic Images of Highly Damaged Concrete Structures. J Perform Constr Facil 2020;35:04020131. https://doi.org/10.1061/(ASCE)CF.1943-5509.0001541.

[69] Rubio JJ, Kashiwa T, Laiteerapong T, Deng W, Nagai K, Escalera S, et al. Multi-class structural damage segmentation using fully convolutional networks. Comput Ind 2019;112:103121. https://doi.org/10.1016/J.COMPIND.2019.08.002.

[70] Montaggioli G, Puliti M, Sabato A. Automated damage detection of bridge's sub-surface defects from infrared images using machine learning. SPIE 11593, Heal. Monit. Struct. Biol. Syst. XV, vol. 11593, SPIE; 2021, p. 427–35. https://doi.org/10.1117/12.2581783.

[71] Dorafshan S, Azari H. Deep learning models for bridge deck evaluation using impact echo. Constr Build Mater 2020;263:120109. https://doi.org/10.1016/J.CONBUILDMAT.2020.120109.

[72] Dorafshan S, Azari H. Evaluation of bridge decks with overlays using impact echo, a deep learning approach. Autom Constr 2020;113:103133. https://doi.org/10.1016/J.AUTCON.2020.103133.

[73] Liu H, Lin C, Cui J, Fan L, Xie X, Spencer BF. Detection and localization of rebar in concrete by deep learning using ground penetrating radar. Autom Constr 2020;118:103279. https://doi.org/10.1016/J.AUTCON.2020.103279.

[74] Ahmed H, La HM, Tran K. Rebar detection and localization for bridge deck inspection and evaluation using deep residual networks. Autom Constr 2020;120:103393.

https://doi.org/10.1016/J.AUTCON.2020.103393.

[75] Xiang Z, Rashidi A, Ou GG. An Improved Convolutional Neural Network System for Automatically Detecting Rebar in GPR Data. Comput Civ Eng 2019 Data, Sensing, Anal - Sel Pap from ASCE Int Conf Comput Civ Eng 2019 2019:422–9. https://doi.org/10.1061/9780784482438.054.

[76] Hou F, Lei W, Li S, Xi J. Deep Learning-Based Subsurface Target Detection from GPR Scans. IEEE Sens J 2021;21:8161–71. https://doi.org/10.1109/JSEN.2021.3050262.

[77] Ahmadvand M, Dorafshan S, Azari H, Shams S. 1D-CNNs for autonomous defect detection in bridge decks using ground penetrating radar. Heal. Monit. Struct. Biol. Syst. XV, vol. 11593, SPIE; 2021, p. 97–113. https://doi.org/10.1117/12.2580575.

[78] Ichi E, Dorafshan S. Effectiveness of Infrared Thermography in Delamination Detection for Concrete Bridge Decks (In Press). Autom Constr 2022.

[79] Besharatian B, Dorafshan S. Non-Contact Bridge Deck Evaluation Using Infrared Thermography, A Pipeline for Data Annotation. 2022 Int Conf Unmanned Aircr Syst 2022:1389–96. https://doi.org/10.1109/ICUAS54217.2022.9836196.

[80] Lavadiya DN. NEURAL NETWORKS AND SENSITIVITY ANALYSIS FOR DETECTION AND INTERPRETATION OF STRUCTURAL DAMAGE. North Dakota State University, 2021.

[81] Barredo Arrieta A, Díaz-Rodríguez N, Del Ser J, Bennetot A, Tabik S, Barbado A, et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 2020;58:82–115. https://doi.org/10.1016/J.INFFUS.2019.12.012.