University of North Dakota
UND Scholarly Commons

Civil Engineering Faculty Publications | Department of Civil Engineering

5-2020

Evaluation of Bridge Decks with Overlays Using Impact Echo, a Deep Learning Approach

Sattar Dorafshan
University of North Dakota, sattar.dorafshan@und.edu

Hoda Azari

How does access to this work benefit you? Let us know!

Follow this and additional works at: https://commons.und.edu/cie-fac

 Part of the Civil and Environmental Engineering Commons

Recommended Citation

Sattar Dorafshan and Hoda Azari. "Evaluation of Bridge Decks with Overlays Using Impact Echo, a Deep Learning Approach" (2020). *Civil Engineering Faculty Publications*. 3.
https://commons.und.edu/cie-fac/3

This Article is brought to you for free and open access by the Department of Civil Engineering at UND Scholarly Commons. It has been accepted for inclusion in Civil Engineering Faculty Publications by an authorized administrator of UND Scholarly Commons. For more information, please contact und.commons@library.und.edu.

Evaluation of Bridge Decks with Overlays Using Impact Echo, a Deep Learning Approach

Sattar Dorafshan[1*], Ph.D., Hoda Azari[2**], Ph.D.

[1] Assistant Professor, Email: sattar.dorafshan@und.edu (Corresponding Author)

[2] Nondestructive Evaluation Program and Laboratory Manager, Email: hoda.azari@dot.gov

[*] Department of Civil Engineering, University of North Dakota, Grand Forks, ND 58202

[**] U.S. Department of Transportation, Turner-Fairbank Highway Research Center, Federal Highway Administration, McLean, VA 22101

ABSTRACT

In this paper, the feasibility of using deep learning models (DLMs) for evaluation of bridges with overlay systems is investigated. Several laboratory-made concrete specimens with artificial subsurface defects and overlay systems (bonded and debonded) made of cement and asphalt overlay materials were tested using impact echo (IE). One-dimensional (1D) and two-dimensional (2D) convolutional neural networks (CNNs) were developed, trained, and tested on the IE data. The proposed 1D CNN was the most successful in detecting debonding and subsurface defects; it achieved an average accuracy of 0.68 on the cement overlay specimens and 0.58 for asphalt overlay specimens. Maps of the defects and debonding were generated using the DLMs and were compared to the conventional method for analyzing the IE data. The 1D CNN produced the most accurate defect maps while successfully detected sound, debonded, and defected regions, particularly on the specimens with cement overlay.

Keywords: impact echo; deep learning; concrete bridge deck evaluation; subsurface defects; overlay; debonding; delamination; nondestructive evaluation; convolutional neural network.

INTRODUCTION

For the past 60 years, overlay systems have been applied on concrete bridge decks for rehabilitation and prolonging their service lives. Overlay systems seal the surface cracks and openings of bridge decks and stop penetration of water and corrosive materials in the decks [1]. Over time, the bond between the deck and the overlay becomes weak, which accounts for the majority of defects in overlay decks [1-2]. Once the bond is gone, the overlay deck is prone to formation of a debonded layer between the overlay system and the underlying deck. The debonded layer could inhibit effective evaluation of the deck during inspections [1]. In addition, debonding allows moisture and corrosive materials to penetrate bridge decks. Therefore, it is important to detect the debonded regions. Nondestructive evaluation (NDE) methods have been used in the past to locate subsurface defects or to obtain material and geometry properties of bridge decks. Lin et al. (2018) conducted a comprehensive study to determine the effectiveness of using different NDE methods to detect subsurface defects in bonded and debonded overlay in laboratory reinforced concrete specimens [1]. They showed the possibility of using IE method to detect subsurface defects and debonding in the specimens with cement-based overlay systems. The IE was not effective when used on the specimens with asphalt-based overlays since the acoustic properties of the overlay system and the underlaying concrete were significantly different. The IE data interpretation in Lin et al. (2018) was performed by analyzing the IE waveforms in the frequency domain with the Fourier transform, i.e., the peak frequency method. In this method, the peak value of the IE frequency spectrum is obtained and analyzed to determine the condition of concrete. The peak frequency selection can be automated but often requires manual re-picking of poor quality data which requires an expert interference for field implementation. The reliance on the expert's opinion for data interpretation has been identified

as one of the reasons State DOTs are reluctant to use IE and other NDE methods during bridge inspections [3-4]. There are also pre- and post-processing operations involved in analyzing IE data before the peak frequency method can be effectively used, such as filtering out unwanted frequencies. Finally, prior knowledge of the inspected area and its thickness is also necessary to use the peak frequency method. Recent availability of robotics platforms for bridge evaluation, such as the RABIT [5], has made more NDE data available from bridges. Among other information, the Federal Highway Administration (FHWA) InfoBridge, provides access to NDE data from several bridges in the United States inspected periodically using RABIT [6]. As more NDE data become available, automation emerges as a necessity for bridge evaluation. Automation could reduce the cost and time associated with the manned inspection by minimizing the role of the human inspector or the expert [7-9]. With more data, autonomous data interpretation becomes even more critical in the future of infrastructure inspection and maintenance. Automation in data analysis could be achieved through using artificial intelligence. Artificial neural networks (ANNs) are a branch of artificial intelligence methods which have been effectively used for big data classification and interpretation [10-13].

Inspired by biological neural networks, ANNs consist of one or multiple layers with a set of learnable parameters. In the literature, these learnable parameters are also known as neurons or weights, depending on ANN's type and architecture. The weights are placed between an input layer and an output layer and are determined through training on an annotated dataset, i.e., training dataset (supervised learning). The input layer converts the data into a format compatible with the ANN's architecture. Then, random values are assigned to the weights. The output layer connects the weight to the desired output (if the task of the network was classification, then the desired output would be the target classes). Determination of weights is done by iteration and

back propagation on an annotated dataset to minimize a loss function by gradually updating the weights. The loss function represents deviation of the classes assigned to the input data by the network and the ground truth. Deep learning models (DLMs) are one of the most popular ANNs. DLMs have exceeded human accuracy in object and pattern recognition on visual images [14-15]. For structural condition assessment, the DLMs have been used to detect surface defect [16-21] and have achieved satisfactory accuracies (more than 90%). However, the application of DLMs for NDE classification could be challenging. One reason is the lack of annotated data to train and validate the DLMs. Obtaining an annotated training dataset validated with the ground truth for IE (and other NDE methods) is significantly more difficult than visual images; because the exact locations of the defects cannot be determined without performing destructive tests such as coring, hydro-demolition, and machine mounted demolition. However, researchers have tried semi-automated methods such as statistical classifiers (i.e., machine learning) to locate the rebar levels in bridge decks using ground penetration radar (GPR) [22-24], and several NDE methods including IE [25-26] to locate delamination. DLMs have been used to detect delamination in carbon fiber reinforced polymer composite from X-ray computed tomography images [27], to classify GPR B-scans to find buried objects [28], subsurface defects in metal casting from X-ray images [29], vibration-based damage classifications of buildings [30], fault diagnostics of bearings using acoustic emissions [31], and subsurface defects in plate-like structures using acoustic emission [32-33]. For bridge evaluation purposes, DLMs were used to locate the rebar position in bridges from GPR data [34]. Dorafshan and Azari (2019) used six DLMs to classify IE data acquired from eight reinforced concrete specimens [35]. They found that DLMs were not only capable of classifying IE data, but they also provided more reliable defect maps compared to the conventional method for analyzing IE data (peak frequency). However, the specimens

used in reference [35] were bare and without any overlay system. According to a national survey published by the Federal Highway Administration (FHWA), many state DOTs use overlay systems to extend the service life of bridge [36]. Therefore, developing robust methodologies for the evaluation of bridges with overlay system is important. This paper investigates the feasibility of using DLMs to detect subsurface defects and debonding in reinforced concrete bridges with cement and asphalt overlay systems using IE. The paper has the following sections: description of the specimens, IE introduction, IE data acquisition procedure, data acquisition, investigated DLMs, results, and conclusions.

DESCRIPTION OF EXPERIMENT

Eight reinforced concrete specimens, 3.0 m long, 1.0 m wide, and 0.2 m thick, were constructed at the FHWA Advanced Sensing Technology (FAST) NDE laboratory using normal-weight concrete mix with a water-to-cement ratio of 0.37 and a 28-day minimal compressive strength of 27.5 MPa, with two mats of uncoated steel reinforcement with No. 5 rebar (15.8 mm diameter) spaced 203 mm in both longitudinal and transverse directions. Four types of artificial defects: shallow delamination (above the specimens' top rebar level), deep delamination (above the specimens' bottom rebar level), honeycombing, and voids (Figure 1), were constructed in the specimens. These defects represent the most common subsurface defect types in bridges [37]. Initially, the specimens were constructed to investigate the effectiveness of NDE methods for evaluation of bridges with different types of overlay systems [37].

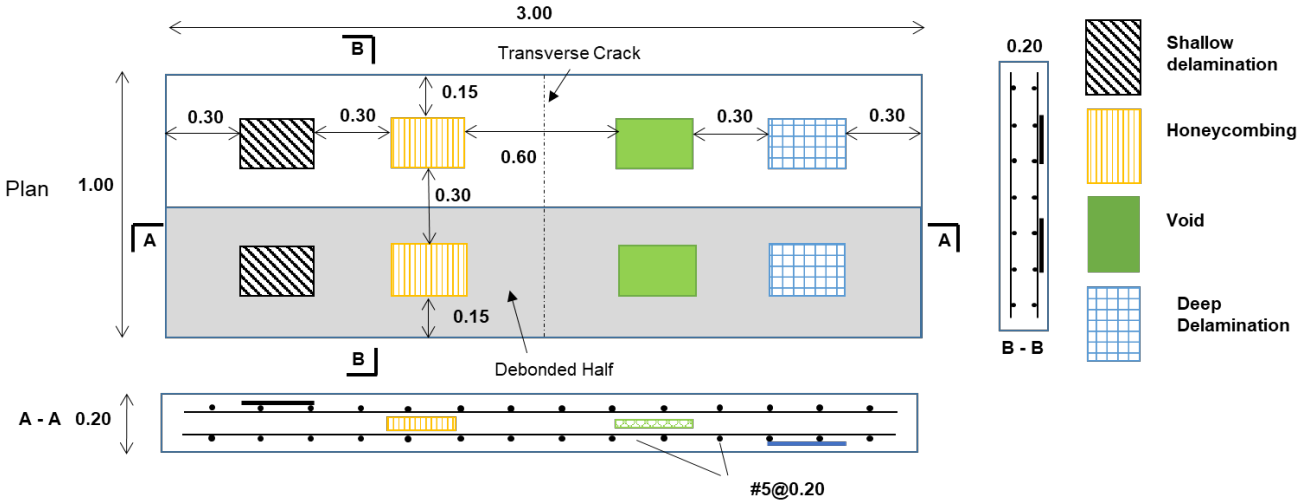


Figure 1 Layout of the laboratory bridge decks

Figure 2 shows the placements of the defects in the specimens before pouring the concrete. All defects were 0.30 m long and 0.20 m wide. A transverse crack was also artificially made in the middle of all specimens. More details about the defects, their materials, and their construction procedure can be found in [1,38].

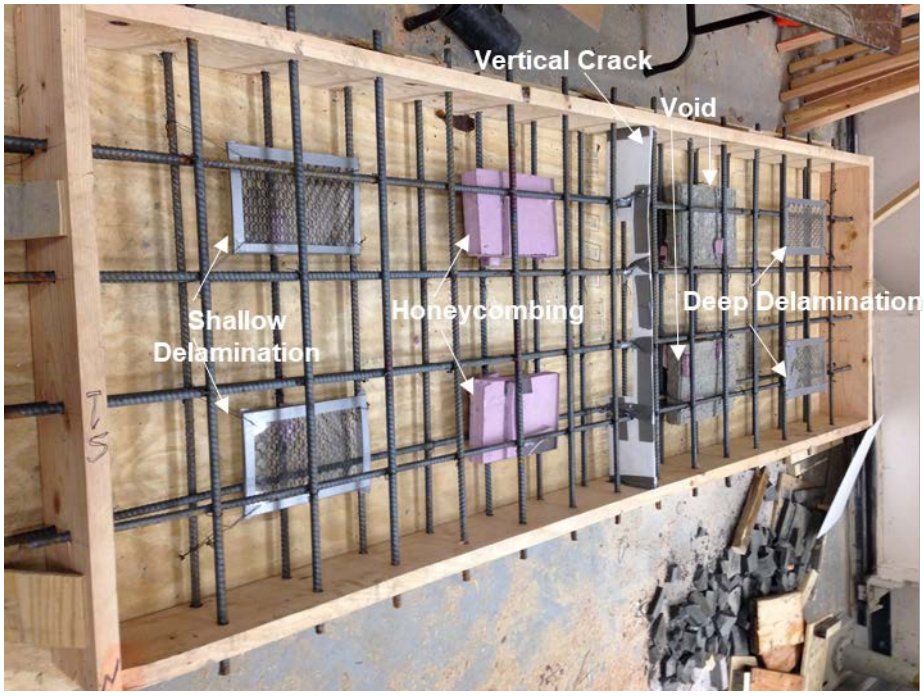


Figure 2 Artificial defects in the specimens

Cement-based overlay systems were placed on four of the specimens: a 2-layer epoxy polymer (S1E), a 38-mm thick latex modified concrete (S3L), a 38-mm thick silica fume modified concrete (S6S), and a 19-mm thick polyester polymer concrete (S8P). Three asphalt-based overlay systems were also constructed and placed on the specimens: asphalt with a sheet membrane (S4AS), asphalt with a liquid membrane (S5AL), and asphalt without a membrane (S7A). One of the specimens (S2) was kept bare for control. The investigated overlay systems represent the most common materials used as the overlay layer in the United States [36]. Before the placing of the overlay systems, half of the area of each specimen was covered with a plastic sheet in the longitudinal direction (shaded area Figure 1). The plastic sheets mimicked the debonding while the other half was shot blasted for full bonding. Details about overlay systems placement and construction considerations can be found in [1].

IE METHOD

ASTM adopted IE technique for thickness measurements of plate-shaped concrete structures and published the IE standard in 2005 [39]. The IE device in this study consisted of steel sphere impactors, an accelerometer, and a data acquisition system (Figure 3). The impactors generate seismic waves through the inspected region. The waves are reflected when reaching the bottom of the medium (backwall reflection) or subsurface irregularities. These reflected waves, i.e., echoes, then travel back to the surface where their amplitudes are recorded by the receiver, i.e., accelerometer.

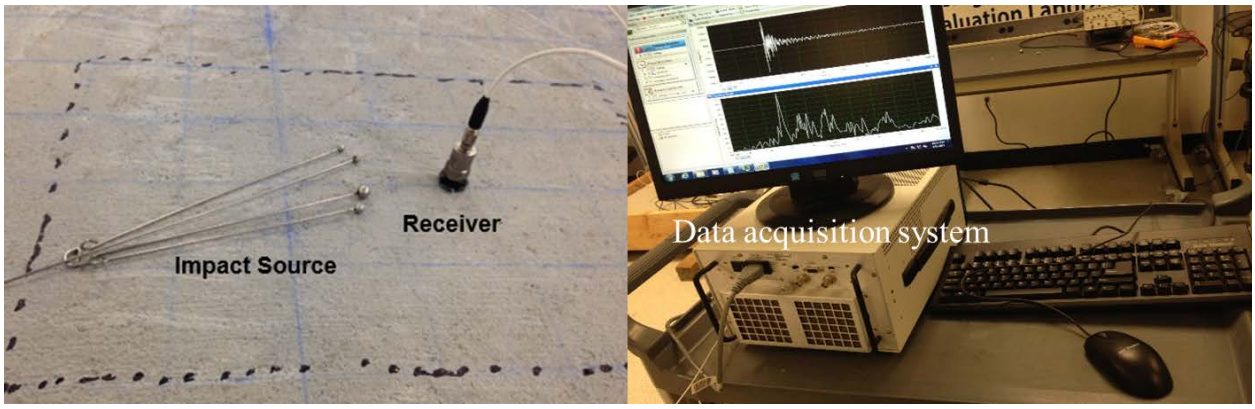


Figure 3 IE device and data acquisition system

Frequency response of the echo is commonly used to interpret IE data. In theory, a distinctive maximum in the frequency response spectrum indicates a sound region if the peak matches a frequency corresponding to the backwall reflection (also known as the thickness frequency) [40]. On the other hand, presence of defects could result in two distinct peaks a single peak at higher frequencies than the thickness frequency, or a low-frequency response of the flexural mode [41]. In practice, the described shapes of the frequency responses could be unlikely to appear in the IE spectrum without processing the data. The processing could be done by applying user-defined filters to remove the noise from the IE signals. Appropriate frequency ranges and threshold values should also be determined before the peak frequency method can be effectively used. The processing operation is not generic and is done by an inspector with expertise in IE analysis which could limit the application of IE for bridge evaluation.

IE DATA ACQUISITION

The IE device was used to inspect the specimens before and after overlay placement. The specimens were marked with a marker to create a grid system on their surface with 100 mm spacing (longitudinally and transversely) which produced a total number of 261 test points. This grid system, partially shown in Figure 3, consisted of a set of markings along each dimension of

the specimens with the mentioned spacing. The impactor and receiver were placed on two adjacent grid points in the longitudinal direction to cover the inspected specimens. The sampling frequency of the IE tests was 200 kHz, and the IE data were collected for 10 milliseconds on each test point. Several steel spheres were initially tested as the impactor as seen in Figure 3; but the IE was performed using a sphere with the diameter of 11 mm. No IE tests were carried out within 100 mm of each side of the transverse crack in the middle of the bare specimens. Since the crack was sealed by the overlay system; all test points could be used to perform the IE test. This was the case for specimens with cement overlay, as seen in Table 1; however, the asphalt-based overlay specimens (S4AS, S5AL, and S7A) had 100 less IE tests since the operator decreased the spatial resolution in the debonded regions. Note that asphalt overlay properties depend on the temperature and the mix design. These specimens were tested at an ambient temperature of 0°C. This temperature was selected after investigating IE signals on asphalt overlay specimens at -19, 0, and 29°C temperatures which showed 0°C was the operating temperature to used IE [38]. The authors classified the IE data into three groups: sound (S), defected (D), and debonded (DB). The D class included the IE data associated with the subsurface defects (in the specimens with and without overlay system), the DB class included the IE data of the debonded half of the overlay specimens, and the S class was assigned to the remainder of the IE data. Table 1 shows the classification of the IE data. Therefore, the IE dataset in this study consisted of 736, 715, and 2092, samples for the D class, the DB class, and the S class, respectively.

Table 1 The IE dataset description

Specimen No	Condition	Class Size			Specimen No	Condition	Class Size		
		D	DB	S			D	DB	S
S1	*Bare*	64	0	188	S1E	*Overlay*	32	145	84
S2	*Bare*	64	0	188	S3L	*Overlay*	32	145	84

S3	*Bare*	64	0	188	S4AS	*Overlay*	32	45	84
S4	*Bare*	64	0	188	S5AL	*Overlay*	32	45	84
S5	*Bare*	64	0	188	S6S	*Overlay*	32	145	84
S6	*Bare*	64	0	188	S7A	*Overlay*	32	45	84
S7	*Bare*	64	0	188	S8P	*Overlay*	32	145	84
S8	*Bare*	64	0	188					

A training dataset was formed that included all IE data acquired on bare specimens. In addition, the IE data of six of the overlay specimens were included in the training while the data of the remaining overlay specimen was used to as a testing dataset. The testing dataset was set to include the IE data from each overlay specimen while the rest of specimens were added to the training dataset. Therefore, seven datasets were created for training the DLMs and seven datasets were created for testing the DLMs. The size of the testing and training datasets are shown in Table 2.

Table 2 The IE dataset description

ID	**Size of Training Dataset**			**Size of Testing dataset**		
	D	**DB**	**S**	**D**	**DB**	**S**
S1E	704	570	2008	32	145	84
S3L	704	570	2008	32	145	84
S4AS	704	670	2008	32	45	84
S5AL	704	670	2008	32	45	84
S6S	704	570	2008	32	145	84
S7A	704	670	2008	32	45	84
S8P	704	570	2008	32	145	84

Figure 4a-c show examples of the IE waveform for the S, the D, and the DB classes, respectively. Figure 4d-f are the spectrograms associated with the waveforms shown in Figure 4a-c. The spectrograms were generated by applying short-time Fourier transforms (STFT) on the IE waveforms with a sliding window length of 30 time steps and an overlap of 50 percent between the windows. In the spectrograms, the horizontal axis represents the time, and the vertical axis represents the signal frequency. To follow a generic approach, all frequencies in the

IE spectrums were used to construct the spectrograms. Finally, the spectrograms were converted into color images with red, green, and blue channels.

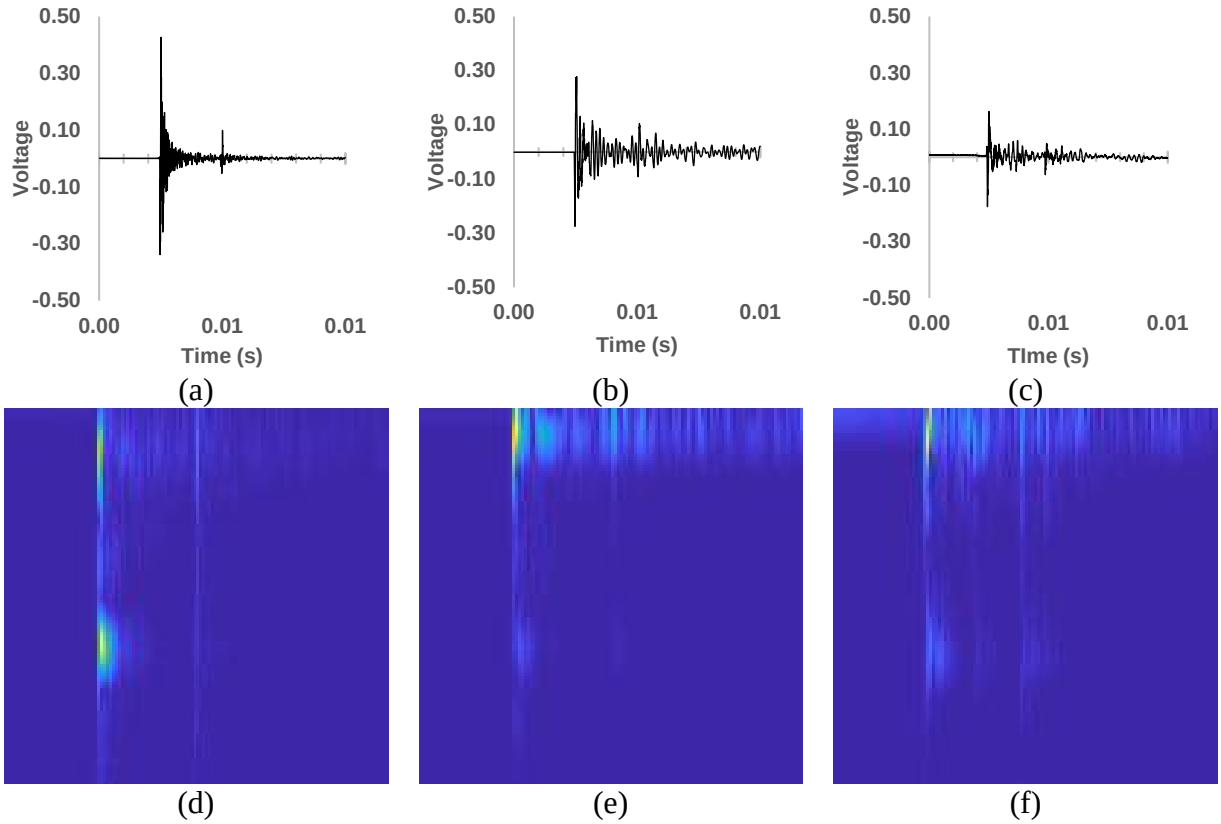


Figure 4 IE waveform representing the (a) S class, (b) D class, (c) DB class, IE spectrograms with time on the horizontal and frequency on the vertical axis, repressing the (d) S class, (e) D class, (f) DB class.

INVESTIGATED DLMs

Convolutional neural networks (CNNs) are one the most successful DLMs for image classification and segmentation in the computer science community. The weights in a CNN are stored in filters that are applied to the input data using convolution operation. It is also possible to design CNNs for 1D data such as IE waveforms [42].

Both 1D and 2D deep learning models were investigated in this paper. For IE waveforms, a 1D CNN was designed inspired by the one that has been successfully applied on the bare concrete specimens [35]. The architecture of the 1D model is shown in Figure 5a. The network

comprised of a combination of basic deep learning layers that included three layers of convolution (Conv.), batch normalization (Norm), and rectified linear unit (ReLU). These layers are called CBR for brevity, and they were connected using two max pooling (MP) layers. The model also had three layers of fully connected (FC), ReLU, and drop out. The combination of these layers is referred to as FRD. The network would assign a probability to each class of the IE data (the D, DB, and S classes). After the last FRD layer, a softmax layer was placed to ensure the sum of all of probabilities for each class did not exceed 1.0. Finally, a classification layer would assign the class with the highest probability to the IE waveforms at the end of the network.

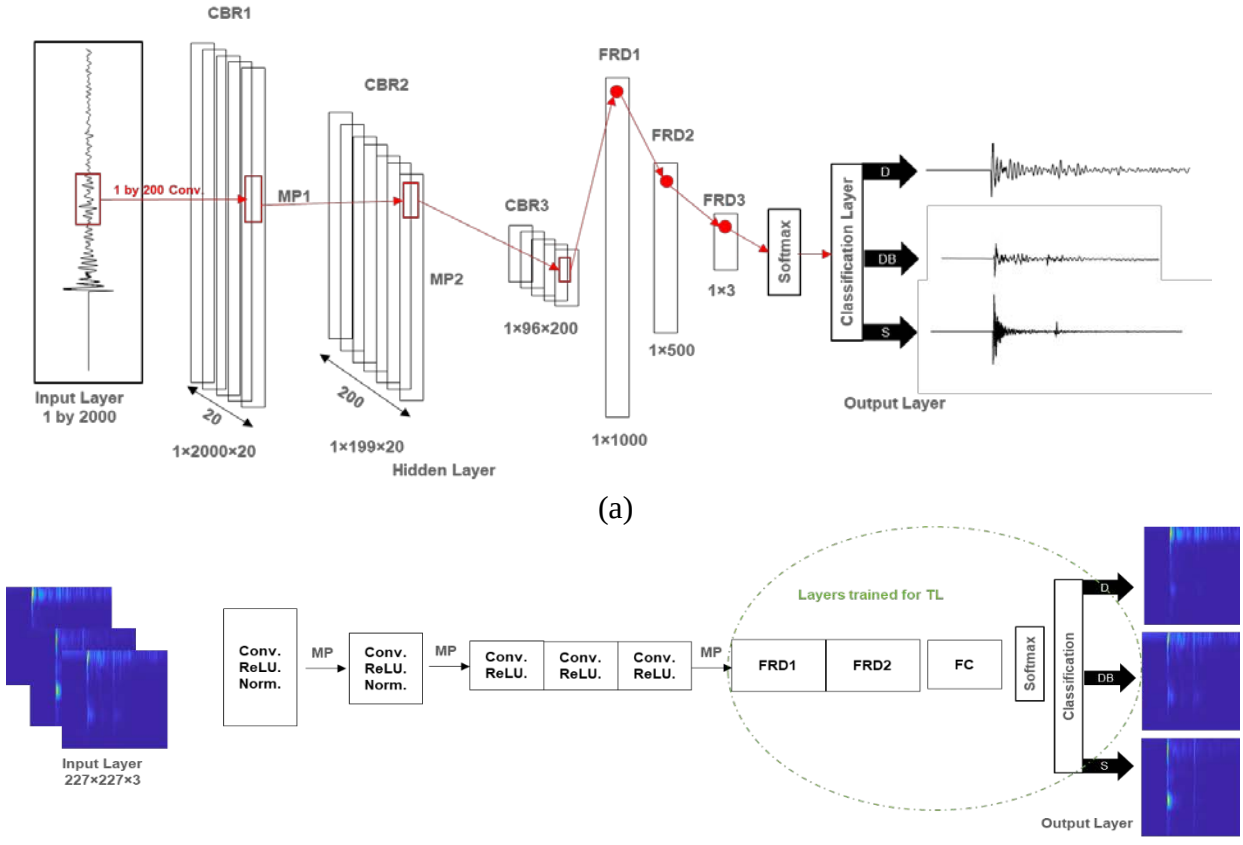


Figure 5 The architecture of CNNs (a) proposed 1D CNN (b) AlexNet

Conventionally applied on visual images, CNNs became famous after the introduction of AlexNet [43], which won the ImageNet classification contest [44] in 2012. Compared to other renowned architectures, AlexNet can be trained easier when limited amount of the annotated data is available [12,35]. AlexNet includes two Conv, ReLU, and Norm layers connected with MP layers, followed by Conv and ReLU layers. The rest of the network consisted of two FRD layers, one FC layer, one softmax layer, and one classification layer, as seen in Figure 5b. For AlexNet, and some other famous CNN architectures, pretrained models on ImageNet are already available. It is possible to use these models for classification of IE spectrograms; however, the end layers of the pretrained network should be retrained to match a new dataset and new number of classes. This mode of training is called transfer learning (TL); as opposed to training a model from scratch (full training (FT)). In the TL mode, the weights towards the end of the network are obtained through training on the IE dataset. The layers holding these weights in shown in Figure 5b (inside of the oval). Transfer learning is commonly used when the number of images in the training dataset is limited. An AlexNet FT model could also be more accurate in detecting defects, while an AlexNet TL model was shown to be more immune to false positive reports [12, 17]. Therefore, both modes of training were investigated in this study.

As seen, the proposed 1D CNN was formed using basic deep learning components which are not inherently different than the AlexNet. However, the 1D CNN was designed to be compatible on 1D data, i.e. IE signals, which would result in a lighter model during the training.

All computations were performed on a desktop workstation with a 64-bit operating system, 32 GB memory, Intel® Core™ i7 CPU, and two GeForce GTX 1080 Ti graphics processing units (GPU). MATLAB 2019a was used to program, train, and test the DLMs. Before training began, several training parameters, i.e., hyperparameters, were determined for each

model in an empirical manner. Table 3 shows the selected values for the hyperparameters. In this table, the optimizer refers to the type of algorithm used to minimize the loss of the model. For the proposed 1D CNN, adaptive moment estimation (ADAM), Root Mean Square Propagation (RMSprops), and stochastic gradient descent with momentum (SGDM) were tested which showed the model achieved higher accuracy when used with SGDM (5% higher than ADAM and 9% higher than RMSprops). Mini batch size is the size of the segments of the training data used to calculate the loss and error of the models. Iteration is when the models go through a forward and backward processing for a batch of data while one epoch is when all training data is processed once over the training dataset. Learning rate is the tolerance of parameters to optimize the gradient descent and minimize the loss. The number of trained layers in each network is also shown in the table (note that only the last layers were trained in the TL mode).

Table 3 Training parameters for each investigated network

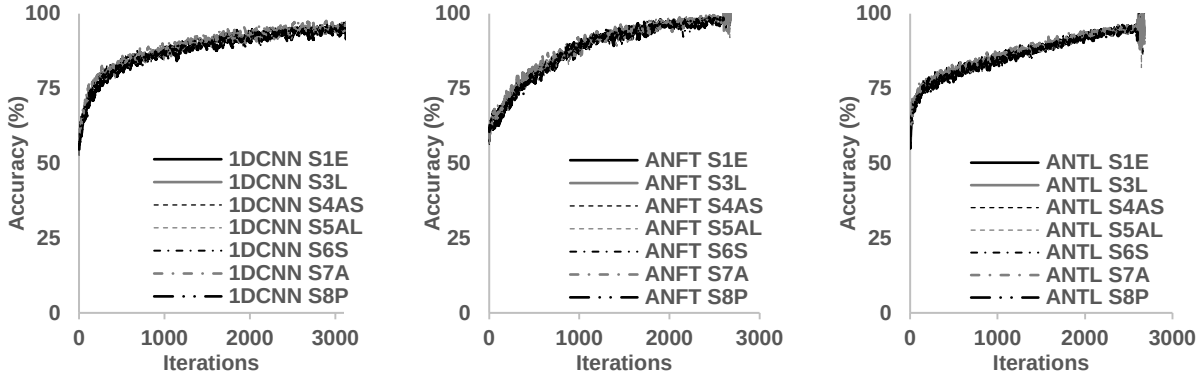
Network	Mode	Optimizer	Mini Batch Size	Max Epochs	Learning Rate	No Trained Layers
1D CNN	FT	SGDM	50	40	1e-3	18
AlexNet CNN	FT	SGDM	50	40	1e-3	25
AlexNet CNN	TL	SGDM	50	40	1e-4	3

RESULTS

Training Results

The results of the training of each DLM are shown in Figure 6. The legends of these plots represent the specimen id that was used for testing. For instance, a training process was labeled as S1E if the DLMs were trained on all the specimens, expect for S1E. After the training was complete, the DLMs were used to label the IE data from the remaining specimen, i.e., S1E, as the testing dataset. All models reached an accuracy of 95 percent or higher towards the end of the training process. Each model required a different amount of time to finish one round of iteration.

On average, the AlexNet FT model required 0.36 s per iteration while the AlexNet TL model finished one iteration in 0.28 s. The training time for the 1D CNN was only 0.05 s per iteration, which made this model considerably faster than the 2D CNNs. This was mainly due to the difference between the size of the 1D and 2D data. For the 1D CNN, the input had 2,000 elements, representing the voltage amplitude for each time step, whereas each input for the 2D CNNs was a color image with 154,587 elements, total number of pixels in the spectrograms including the three channels 3×227×227. Therefore, the number of calculations required to finish the training increased significantly for the 2D CNNs. In addition, the proposed 1D CNN also had 50 percent less weights in its architecture compared to the AlexNet in both the FT and TL modes. The fully connected layers included about 90 percent of the parameters in the AlexNet architecture. The training in the TL mode was not substantially faster than in the FT mode since the fully connected layer had to be trained in both modes.



(a) (b) (c)

Figure 6 Training results (a) 1D CNN, (b) AlexNet FT, (c) AlexNet TL

Testing Results

Table 4 shows the testing results of the investigated DLMs. In the table, the true classification rate (TCR) for each class is reported. The TCR values for each class was calculated

by dividing the total number of correct classification of IE signal in each class, to the total number of IE signals in that class. The DLMs assigned three probabilities to each input data corresponding to the three classes. The classification layer assigned the input data to the class with the highest probability. A true detection was when the DLM assigned the same label to the input data as the ground truth. In addition to TCR values, the weighted average of the TCR is also reported. The weighted average is the total number of correct classifications to the total number of IE data, which represented the accuracy of each DLM (ACC).

Overall, the DLMs were more successful on the specimens with cement-based overlay systems. The asphalt overlays had significantly different material properties (lower acoustic impedance) than the base reinforced concrete [38], therefore they were more difficult to classify than the cement-based overlays. The same observation was made when the peak frequency method was used in [1]. The proposed 1D CNN achieved the highest accuracy on the specimens with cement-based overlays (0.65. 0.57, and 0.82 for S3L, S6S, and S8P, respectively) except for S1E, where the AlexNet models reached the highest accuracy of 0.80. With 0.68, the average accuracy of the 1D CNN was higher than AlexNet FT with 0.61, and AlexNet TL with 0.64. This model also detected debonding more accurately than did the AlexNet models, with the average TCR of 0.66 compared to the TCR of 0.51 achieved by the AlexNet FT model and 0.57 achieved by the AlexNet TL model. Even though the 2D spectrograms preserved both the time and frequency information of the IE data, they were inherently different than the color images that AlexNet was invented to classify. Unlike the color images, the vertical and horizontal axis in spectrograms presented two different physical concepts, i.e., time and frequency. However, AlexNet FT was the most successful model to classify the D class, with a mean TCR of 0.60 on the specimens with cement-based overlays. The AlexNet TL reached the highest average TCR of

0.83 in the S class; and both 1D CNN and AlexNet FT had an average TCR value of 0.79. Since the AlexNet TL model benefited from being trained on a larger dataset, i.e., ImageNet, it was less susceptible to false positives. Note that the spectrograms could be obtained using different methods, such as wavelet, which could potentially lead to different classification results.

Table 4 Summary of results

Specimen No	TCR (1DCNN)			TCR (AN FT)			TCR (AN TL)			ACC (%)		
	D	DB	S	D	DB	S	D	DB	S	1DCNN	AN FT	AN TL
Cement-based overlay systems												
S1E	0.59	0.59	0.88	0.59	0.82	0.86	0.44	0.79	0.95	0.68	**0.80**	**0.80**
S3L	0.44	0.49	1.00	0.56	0.19	0.93	0.25	0.32	0.95	**0.65**	0.47	0.51
S6S	0.69	0.68	0.33	0.47	0.13	0.68	0.41	0.29	0.79	**0.57**	0.35	0.46
S8P	0.41	0.88	0.88	0.75	0.88	0.70	0.78	0.89	0.64	**0.82**	0.81	0.80
Asphalt-based overlay systems												
S4AS	0.34	0.47	0.73	0.00	0.51	0.65	0.16	0.47	0.76	**0.58**	0.48	0.56
S5AL	0.06	0.82	0.89	0.22	0.76	0.71	0.31	0.71	0.48	**0.71**	0.63	0.51
S7A	0.53	0.60	0.33	0.34	0.44	0.49	0.72	0.33	0.21	**0.45**	**0.45**	0.35

The proposed 1D CNN achieved the highest accuracy in all specimens with asphalt overlay (0.58. 0.71, and 0.45 for S4AS, S5AL, and S7A, respectively). The DLMs achieved considerably less TCR rates for the D class in the specimens with asphalt overlays (average 0.30) compared to the specimens with cement overlays (average 0.53). The performance of the DLMs in classifying the DB class, however, was not compromised as significantly as it was for the D class. The average TCR was 0.58 in the specimens with asphalt overlays and 0.58 in the specimens with cement overlays. The DLMs were the most successful in the specimen with asphalt liquid (S5AL) among asphalt-based overlays, with accuracies of 0.71 for 1D CNN, 0.63 for AlexNet FT, and 0.51for AlexNet TL.

Defect Maps

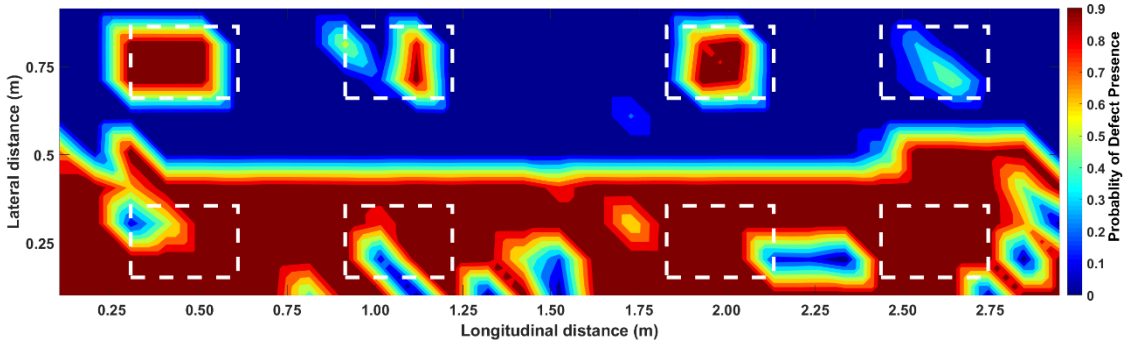
Presenting IE results in color-coded contour maps is more common in the NDE literature, rather than reporting discrete performance metrics (TCR and ACC). Since it is important to

detect both debonding and subsurface defects during bridge deck condition assessment [40], the sum of the assigned probabilities to the D and the DB classes were used to indicate the "defected" regions, and the assigned probability to the S class by the DLMs was used to represent the sound concrete. These probabilities were stored in the softmax layer of each DLM. By extracting these probabilities and mapping them on the spatial coordinates of the specimens, a contour map can be generated. In addition, a set of defect maps were generated using the peak frequency method. For brevity, the defect maps of one specimen with cement (S3L) and one specimen with asphalt (S5AL) overlay are presented and discussed here; however, the content of the following discussions could be expanded to the rest the specimens. Note that these specimens were selected to no not represent DLMs' best performance (refer to Table 4).

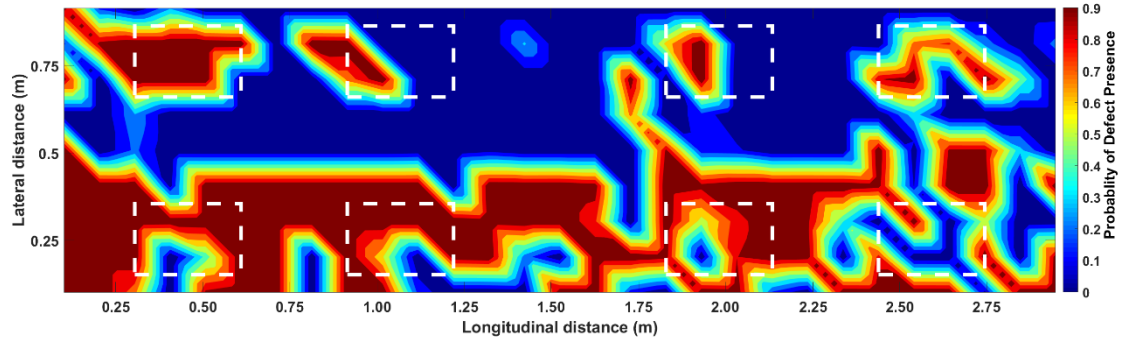
Figure 7a-c show the contour maps of one of the specimens with cement overlay (S3L) generated by the investigated DLMs. The location of each subsurface defect (as previously shown in Figure 1) is also mapped as a dashed-line box. As seen, the DLMs could show indication of defects on the top half of the specimens with the most accurate localization for the shallow delamination and the void (the first and third defects from the left). The overlay was debonded from the deck on the lower halves of the specimens, which was also detected by the DLMs for the most part. Figure 7d shows the defect map of specimen S3L using the peak frequency method which mapped the peak frequency values on the specimen spatial grid. Peak frequencies lower than 8.5 kHz can be detected in the debonded regions on the lower half of the specimen. The proposed 1D CNN provided an equally reliable defect map. In addition, the 1D CNN outperformed the 2D CNNs in defect map generation. This was expected due to the 1D CNNs higher TCR rate and accuracy. The 1D CNN was able to localize shallow delamination and honeycomb, with probabilities more than 80 percent. The deep delamination was also

detected with a lower probability (between 30 and 50 percent) in the bonded zone. In addition, the majority of the lower half of the deck had a defect probability of 80 percent and higher, which matched the ground truth.

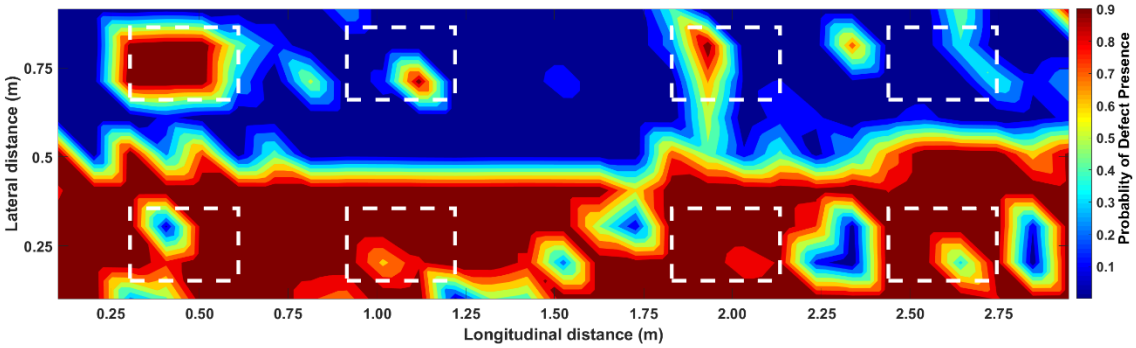
(a)



(b)



(c)



(d)

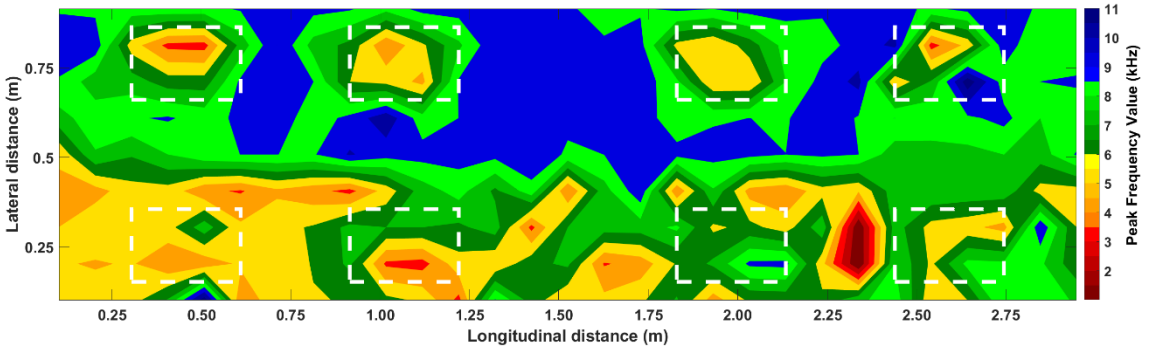


Figure 7 Defect maps of a specimen with cement overlay (S3L), (a) 1D CNN, (b) AN FT, (c) AN TL, (d) peak frequency

Figure 8a-c show the defect maps of one of the specimens with asphalt overlay (S5AL) generated by the DLMs. On the top half, the 1D CNN was able to localize shallow delamination and honeycomb with probabilities higher than 80 percent, while the deep delamination was detected with a lower probability (between 30 to 50 percent). Shallow delamination was detected accurately by all DLMs. Similar to the specimens with cement overlays, the DLMs were also successful in detection of the void on the top half. As seen in Table 1, only 45 IE test were performed on the debonded regions; therefore, the defect maps on the lower half of these specimens had less spatial resolutions than did the top half maps. Figure 8d shows the defect map of specimen S5AL generated by the peak frequency method. Only the shallow delamination was detected on the top half, and the distinction between the debonded region (lower half) and the bonded region (top half) is impossible. The proposed 1D CNN was able to localize shallow delamination and honeycomb, with probabilities more than 80 percent; but it missed the deep delamination. The AlexNet TL detected all four defects on the top half but misclassified sound regions as defect in this area. Note that the contour maps were generated by interpolating values assigned to each test point on the specimens. Since the lower halves of the asphalt specimens were presented with less values, the distinction between the bonded and debonded regions were not as clear as it was for the specimens with cement overlay in Figure 7.

(a)

(b)

(c)

(d)

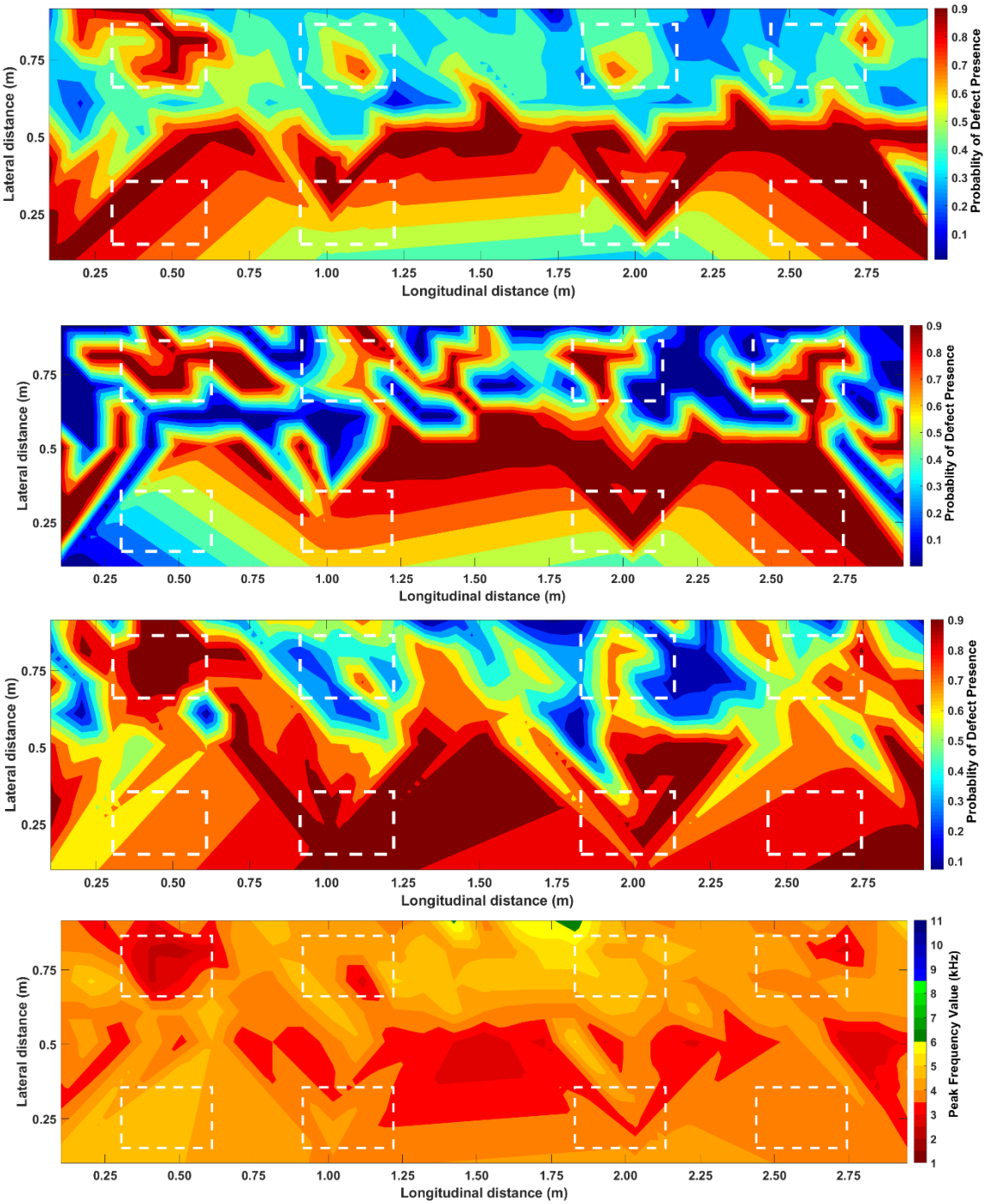


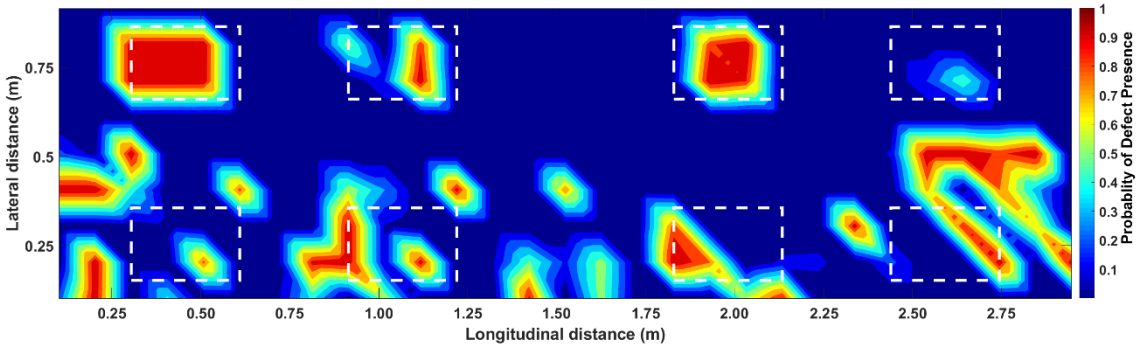
Figure 8 Defect maps of a specimen with asphalt overlay (S5AL), (a) 1D CNN, (b) AN FT, (c) AN TL, (d) peak frequency

The DLMs, the 1D CNN in particular, outperformed the peak frequency method in generating defect maps for the specimens with cement overlay. Their performance exceeded the peak frequency method by detecting most of the defects on the top half. DLMs detected the

debonded region more accurately in the specimens with cement overlay as seen in Figure 7a-c as opposed to the peak frequency method (Figure 7d). For specimens with asphalt overlay, the defect maps were less compatible with the ground truth the specimens with cement overlay, however, still more consistent with the ground truth compared to the peak frequency method.

Reviewing the classification results of specimens with asphalt overlay revealed that the assigned probabilities to the DB and the D classes agreed less to the ground truth classes compared to the cement overlay specimens. The classification results corresponding to the D and DB classes using the 1D CNN are shown for specimens S3L and S5AL in Figure 9. The network detected the D class on the top of Figure 9a, except for the deep delamination in the specimens with cement overlay (S3L). In addition, the debonded and bonded regions were distinguished clearly on the same specimen as seen in Figure 9b. However, the D class regions were misclassified in the bonded region on the top of the specimen S5AL that had asphalt overlay on (Figure 9c). The indications of defect can be seen on the border of the shallow delamination region, i.e., false positive, in this specimen. Even though the 1D CNN detected the D class for the S5AL in Figure 9d; but, it also classified the D classes as the DB class as seen in the top of the figure.

(a)



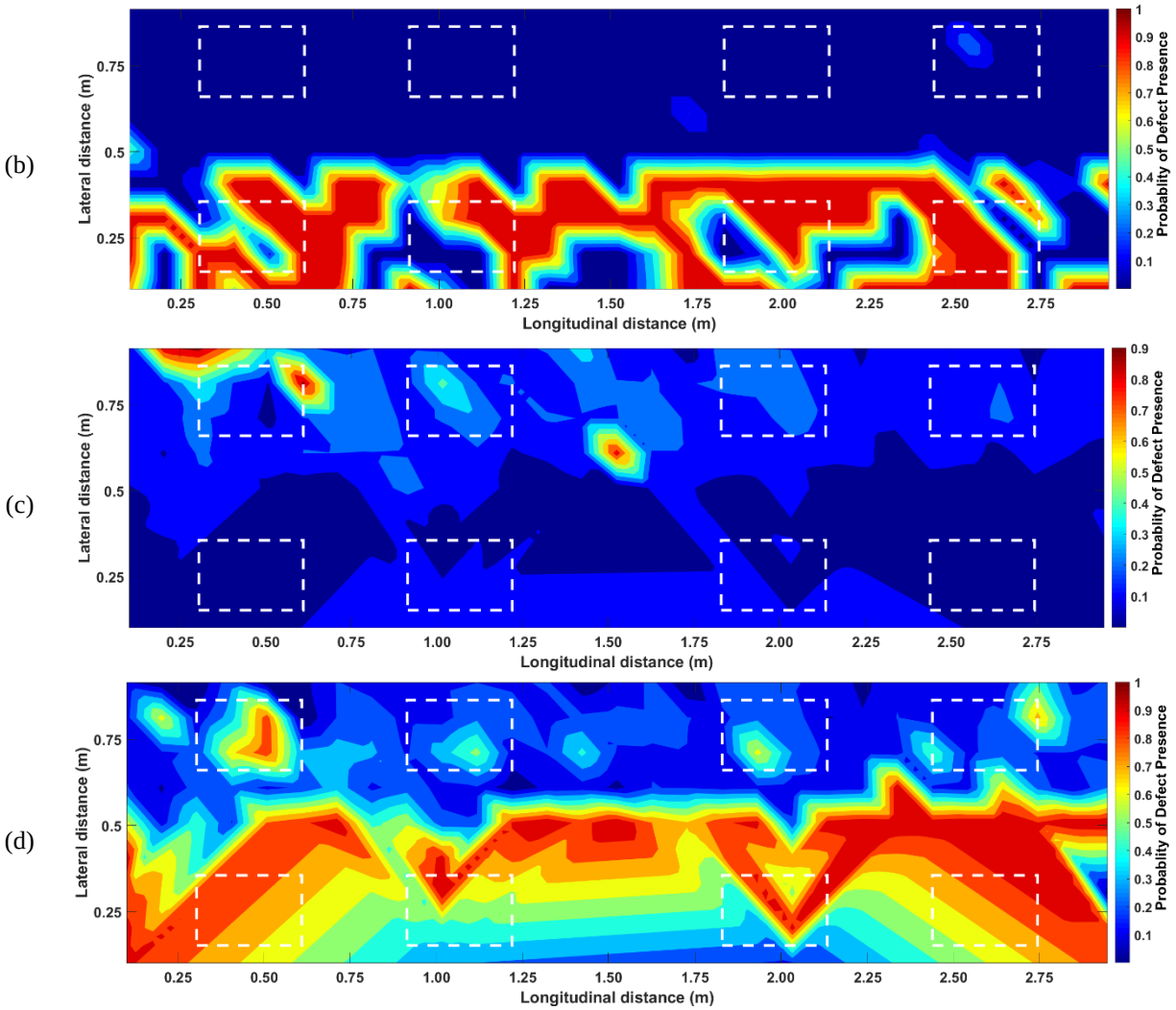


Figure 9 1D CNN Defect map of a specimen with cement overlay (S3L), (a) probability of the D class, (b) probability of the DB class; and a specimen with asphalt overlay (S5AL), (c) probability of the D class, (d) probability of the DB class

The misclassification for the specimens with asphalt overlay could be due the property variation between the base and the overlay materials.

Considering the limited training data that was used for training in this study, the results of the DLMs for IE classification can be considered successful. This study showed that DLMs can be used for structural evolution of bridge decks with overlay systems. Their performance was overall more consistent with the ground truth when classifying specimens with cement overlay system. However, to use trained DLMs in practice, actual IE field data should be included in the

training dataset. In particular, more data is required to improve the detection rate of the debonding in the bridges with asphalt overlay. There are IE data stored in the public repositories such as InfoBridge [6]; however, the IE data for bridges with overlay systems do not exist in InfoBridge since the feasibility of using IE for these bridges was unproven (reference [35] could be regarded as a successful feasibility study where IE were used to evaluate bridges with overlay systems). Even if such data existed, the DLMs could not be validated if used to classify such data due to the absence of the ground truth which is another issue with field implementation of the DLMs.

The 1D CNN was the most accurate model and its architecture was designed for waveforms with a specific frequency and duration (dictated by the IE device). Therefore, the 1D CNN in the current form cannot be used to classify data with a different size. The normalization process in the DLMs were carried out on the amplitude values of the signal amplitude (vertical axis), not the time steps (horizontal axis). However, conventional signal normalization, padding techniques, and domain adaptation can be used to make the field data and the training data compatible [45-46].

CONCLUSIONS

In this study, the feasibility of using deep learning models (DLMs) to locate subsurface defects and overlay debonding from impact echo (IE) data was investigated. Specimens with intentional defects were constructed at FHWA Advanced Sensing Technology (FAST) NDE laboratory. Bonded and debonded overlay systems made of cement and asphalt material were built and placed on the specimens. One dimensional and two dimensional convolutional neural network (CNN) was designed to classify the IE waveforms collected from the specimens. The networks achieved varying accuracies between 0.45 to 0.81 (more accurate on the cement

overlay, average accuracy of 0.68, than on the asphalt overlay, average accuracy of 0.58). Shallow delamination, void, and debonded regions were accurately detected in all specimens using the deep learning models. The proposed 1D CNN showed the most promising results by having the highest accuracies in six out of seven specimens. Compared to the peak frequency, the proposed 1D CNN was a better tool since:

- After the training, it does not require any expert input to classify the IE data;
- It detected debonding with more accuracy in both cement and asphalt overlay systems;
- It was able to differentiate between the debonded regions and the concrete defects.

This study showed the feasibility and application of deep learning models for evaluation of bridge decks with overlay system inspected by IE. With more data, the performance of these models will improve, making them more robust and accurate for bridge inspections.

DATA AVAILABILITY

The datasets used in this paper is published in public dataset [47].

ACKNOWLEDGMENTS

This research was sponsored by the Federal Highway Administration (FHWA). The contents do not necessarily reflect the official views or policies of the FHWA. This research was performed while the first author held a National Research Council (NRC) Research Associateship award administrated by the National Academies of Sciences, Engineering, and Medicine fellowships office and their support is greatly acknowledged.

AUTHOR CONTRIBUTIONS

The authors confirm contribution to the paper as follows: study conception and design:

All authors; data collection: all authors; analysis and interpretation of results: first author; draft manuscript preparation: all authors. All authors reviewed the results and approved the final version of the manuscript.

REFERENCES

[1] Lin, S., Meng, D., Choi, H., Shams, S., & Azari, H. (2018). Laboratory assessment of nine methods for nondestructive evaluation of concrete bridge decks with overlays. *Construction and Building Materials*, *188*, 966-982. https://doi.org/10.1016/j.conbuildmat.2018.08.127

[2] Delatte Jr, N. J., Fowler, D. W., McCullough, B. F., & Gräter, S. F. (1998). Investigating performance of bonded concrete overlays. *Journal of performance of constructed facilities*, *12*(2), 62-70. https://doi.org/10.1061/(ASCE)0887-3828(1998)12:2(62)

[3] Lee, S., & Kalos, N. (2015). Bridge inspection practices using non-destructive testing methods. *Journal of Civil Engineering and Management*, *21*(5), 654-665. https://doi.org/10.1061/9780784413517.132

[4] Dorafshan, S., & Maguire, M. (2018). Bridge inspection: human performance, unmanned aerial systems and automation. *Journal of Civil Structural Health Monitoring*, *8*(3), 443-476. https://doi.org/10.1007/s13349-018-0285-4

[5] Gucunski, N., Kee, S., La, H., Basily, B., & Maher, A. (2015). Delamination and concrete quality assessment of concrete bridge decks using a fully autonomous RABIT platform. *Structural Monitoring and Maintenance*, *2*(1), 19-34. https://doi.org/10.12989/smm.2015.2.1.019

[6] FHWA (2019), LTBP InfoBridge: Data. URL: https://infobridge.fhwa.dot.gov/Data, Accessed October 4, 2019.

[7] Lattanzi, D., & Miller, G. (2017). Review of robotic infrastructure inspection systems. *Journal of Infrastructure Systems*, *23*(3), 04017004. https://doi.org/10.1061/(ASCE)IS.1943-555X.0000353

[8] Dorafshan, S., Maguire, M., Hoffer, N. V., & Coopmans, C. (2017, June). Challenges in bridge inspection using small unmanned aerial systems: Results and lessons learned. In *2017 International Conference on Unmanned Aircraft Systems (ICUAS)* (pp. 1722-1730). IEEE. https://doi.org/10.1109/ICUAS.2017.7991459

[9] Dorafshan, S., Thomas, R. J., & Maguire, M. (2018). Fatigue crack detection using unmanned aerial systems in fracture critical inspection of steel bridges. Journal of bridge engineering, 23(10), 04018078. https://doi.org/10.1061/(ASCE)BE.1943-5592.0001291

[10] Ebrahimkhanlou, A., Dubuc, B., & Salamone, S. (2019). A generalizable deep learning framework for localizing and characterizing acoustic emission sources in riveted metallic panels. *Mechanical Systems and Signal Processing*, *130*, 248-272. https://doi.org/10.1016/j.ymssp.2019.04.050

[11] Dorafshan, S., Thomas, R. J., & Maguire, M. (2019). Benchmarking Image Processing Algorithms for Unmanned Aerial System-Assisted Crack Detection in Concrete Structures. Infrastructures, 4(2), 19. https://doi.org/10.3390/infrastructures4020019

[12] Dorafshan, S., Thomas, R. J., Coopmans, C., & Maguire, M. (2018, June). Deep learning neural networks for sUAS-assisted structural inspections: Feasibility and application. In 2018 International Conference on Unmanned Aircraft Systems (ICUAS) (pp. 874-882). IEEE. https://doi.org/10.1109/ICUAS.2018.8453409

[13] Liu, H. and Zhang, Y. (2019). Image-driven structural steel damage condition assessment method using deep learning algorithm. Measurement,133: 168-181. https://doi.org/10.1016/j.measurement.2018.09.081

[14] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In Proceedings of the IEEE international conference on computer vision (pp. 1026-1034). DOI 10.1109/ICCV.2015.123

[15] Dodge, S., & Karam, L. (2017, July). A study and comparison of human and deep learning recognition performance under visual distortions. In *2017 26th international conference on computer communication and networks (ICCCN)* (pp. 1-7). IEEE. https://doi.org/10.1109/ICCCN.2017.8038465

[16] Dorafshan, S., Thomas, R. J., & Maguire, M. (2018). Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete. *Construction and Building Materials*, *186*, 1031-1045. https://doi.org/10.1016/j.conbuildmat.2018.08.011

[17] Huang, H. W., Li, Q. T., & Zhang, D. M. (2018). Deep learning based image recognition for crack and leakage defects of metro shield tunnel. *Tunnelling and Underground Space Technology*, *77*, 166-176. https://doi.org/10.1016/j.tust.2018.04.002

[18] 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%2F1475921717737051

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

[20] Maeda, K., Takahashi, S., Ogawa, T., & Haseyama, M. (2019). Convolutional sparse coding-based deep random vector functional link network for distress classification of road structures. *Computer-Aided Civil and Infrastructure Engineering*. https://doi.org/10.1111/mice.12451

[21] Huyan, J., Li, W., Tighe, S., Zhai, J., Xu, Z., & Chen, Y. (2019). Detection of sealed and unsealed cracks with complex backgrounds using deep convolutional neural network. *Automation in Construction*, *107*, 102946. https://doi.org/10.1016/j.autcon.2019.102946

[22] Kaur, P., Dana, K. J., Romero, F. A., & Gucunski, N. (2016). Automated GPR rebar analysis for robotic bridge deck evaluation. *IEEE transactions on cybernetics*, *46*(10), 2265-2276. https://doi.org/10.1109/TCYB.2015.2474747

[23] Dinh, K., Zayed, T., Romero, F., & Tarussov, A. (2014). Method for analyzing time-series GPR data of concrete bridge decks. *Journal of Bridge Engineering*, *20*(6), 04014086. https://doi.org/10.1061/(ASCE)BE.1943-5592.0000679

[24] Dinh, K., Gucunski, N., Kim, J., & Duong, T. H. (2017). Method for attenuation assessment of GPR data from concrete bridge decks. *NDT & E International*, *92*, 50-58. https://doi.org/10.1016/j.ndteint.2017.07.016

[25] Völker, C., & Shokouhi, P. (2015). Multi sensor data fusion approach for automatic Honeycomb detection in concrete. NDT & E International, 71, 54-60. https://doi.org/10.1016/j.ndteint.2015.01.003

[26] Epp, T., Svecova, D., & Cha, Y. J. (2018, March). Automated air-coupled impact echo based non-destructive testing using machine learning. In Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018 (Vol. 10598, p. 105981M). International Society for Optics and Photonics. https://doi.org/10.1117/12.2295947

[27] Sammons, D., Winfree, W. P., Burke, E., & Ji, S. (2016, February). Segmenting delaminations in carbon fiber reinforced polymer composite CT using convolutional neural networks. In *AIP conference proceedings* (Vol. 1706, No. 1, p. 110014). AIP Publishing. https://doi.org/10.1063/1.4940585

[28] Besaw, L. E., & Stimac, P. J. (2015, May). Deep convolutional neural networks for classifying GPR B-scans. In *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XX* (Vol. 9454, p. 945413). International Society for Optics and Photonics. https://doi.org/10.1117/12.2176250

[29] Ferguson, M., Ak, R., Lee, Y. T. T., & Law, K. H. (2017, December). Automatic localization of casting defects with convolutional neural networks. In *2017 IEEE International Conference on Big Data (Big Data)* (pp. 1726-1735). IEEE. https://doi.org/10.1109/BigData.2017.8258115

[30] 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. *Neurocomputing*, *275*, 1308-1317. https://doi.org/10.1016/j.neucom.2017.09.069

[31] Appana, D. K., Prosvirin, A., & Kim, J. M. (2018). Reliable fault diagnosis of bearings with varying rotational speeds using envelope spectrum and convolution neural networks. *Soft Computing*, *22*(20), 6719-6729. https://doi.org/10.1007/s00500-018-3256-0

[32] Ebrahimkhanlou, A., & Salamone, S. (2018). Single-sensor acoustic emission source localization in plate-like structures using deep learning. Aerospace, 5(2), 50. https://doi.org/10.1117/12.2296613

[33] Ebrahimkhanlou, A., & Salamone, S. (2018, March). Single-sensor acoustic emission source localization in plate-like structures: a deep learning approach. In Health Monitoring of Structural and Biological Systems XII (Vol. 10600). p. 59. International Society for Optics and Photonics. https://doi.org/10.1117/12.2296613

[34] Dinh, K., Gucunski, N., & Duong, T. H. (2018). An algorithm for automatic localization and detection of rebars from GPR data of concrete bridge decks. *Automation in Construction, 89*, 292-298. https://doi.org/10.1016/j.autcon.2018.02.017

[35] Dorafshan, S., Azari, H. (under review). Using Deep Learning Models for Subsurface Defect Detection in Concrete Bridge Decks. Submitted to Construction and Building Materials.

[36] Lane, S., & Jalinoos, F. (2016). FHWA LTBP Summary—Current Information on the Use of Overlays and Sealers. *FHWA Publication, McLean, VA*. FHWA Publication No.: FHWA-HRT-16-079. Accessed December 12, 2019. https://www.fhwa.dot.gov/publications/research/infrastructure/structures/bridge/16079/16079.pdf

[37] Gucunski, N., Imani, A., Romero, F., Nazarian, S., Yuan, D., Wiggenhauser., H., Shokouhi, P., and Taffe, A. (2013). Nondestructive Testing to identify concrete bridge deck deterioration. Report S2-R06A-RR-1. The second strategic highway research program, Washington, D.C.: Transportation Research Board of the National Academies. ISBN: 978-0-309-12933-6

[38] Azari, H., & Lin, S. (2019). Evaluation of the Impact Echo Method for Concrete Bridge Decks with Asphalt Overlays. *Transportation Research Record*, 0361198119828676. https://doi.org/10.1177/0361198119828676

[39] ASTM C 1383. (2004). Test method for measuring the P-wave speed and the thickness of concrete plates using the Impact-Echo method, ASTM Standards. vol. 04.02. ASTM: West Conshohocken, PA. DOI: 10.1520/C1383-15

[40] FHWA (2015). Nondestructive Evaluation (NDE) Web Manual, Version 1.0. FAST NDE Program, Federal Highway Administration Research and Technology. Available online: https://fhwaapps.fhwa.dot.gov/ndep/. (accessed on 05 May 2019).

[41] Carino, N. J. (2001, May). The impact-echo method: an overview. In *Proceedings of the 2001 Structures Congress & Exposition* (pp. 21-23). American Society of Civil Engineers. https://doi.org/10.1061/40558(2001)15

[42] Rawat, W., & Wang, Z. (2017). Deep convolutional neural networks for image classification: A comprehensive review. *Neural computation*, *29*(9), 2352-2449. https://doi.org/10.1162/neco_a_00990

[43] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097-1105). https://doi.org/10.1145/3065386

[44] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). IEEE. https://doi.org/10.1109/CVPR.2009.5206848

[45] Xie, G. S., Zhang, X. Y., Yan, S., & Liu, C. L. (2015). Hybrid CNN and dictionary-based models for scene recognition and domain adaptation. *IEEE*

Transactions on Circuits and Systems for Video Technology, *27*(6), 1263-1274. https://doi.org/10.1109/TCSVT.2015.2511543

[46] Zhang, W., Li, C., Peng, G., Chen, Y., & Zhang, Z. (2018). A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical Systems and Signal Processing*, *100*, 439-453. https://doi.org/10.1016/j.ymssp.2017.06.022

[47] Dorafshan, S; Azari, H (2020), "Overlay IE2020", Mendeley Data, v1

http://dx.doi.org/10.17632/vdvmj3zwf4.1