Deep Learning for Ultrasonic Crack Characterization in NDE

Richard J. Pyle, Rhodri L.T. Bevan, Robert R. Hughes, Rosen K. Rachev, Amine Ait Si Ali, Paul D. Wilcox, *Member, IEEE*

***Abstract*— Machine learning for Non-Destructive Evaluation (NDE) has the potential to bring significant improvements in defect characterization accuracy due to its effectiveness in pattern recognition problems. However, the application of modern machine learning methods to NDE has been obstructed by the scarcity of real defect data to train on. This paper demonstrates how an efficient, hybrid finite element and ray-based simulation can be used to train a Convolutional Neural Network (CNN) to characterize real defects. To demonstrate this methodology, an inline-pipe inspection application is considered. This uses four plane wave images from two arrays, and is applied to the characterization of cracks of length 1-5 mm and inclined at angles of up to 20° from the vertical. A standard image-based sizing technique, the 6 dB drop method, is used as a comparison point. For the 6 dB drop method the average absolute error in length and angle prediction is ±1.1 mm, ±8.6° while the CNN is almost four times more accurate at ±0.29 mm, ±2.9°. To demonstrate the adaptability of the deep-learning approach, an error in sound speed estimation is included in the training and test set. With a maximum error of 10% in shear and longitudinal sound speed the 6 dB drop method has an average error of ±1.5 mm, ±12° while the CNN has ±0.45 mm, ±3.0°. This demonstrates far superior crack characterization accuracy by using deep learning rather than traditional image-based sizing.**

***Index Terms*— Ultrasound, defect characterization, deep learning, neural networks, plane wave imaging, simulation**

I. Introduction

NON-Destructive Evaluation (NDE) techniques are used to extend component life spans, reduce manufacturing costs and increase safety. Inspections typically involve capturing the response of a material to a physical stimulus such as ultrasound, eddy-currents or X-rays. This is often carried out from many angles and locations to build up an image of the internal structure of a component and identify the presence of damage. To detect smaller defects and characterize them more accurately data capture and signal processing techniques are constantly advancing. The quantity and complexity of NDE data relating to a potentially defective region is therefore getting larger every year. Traditionally, defect characterization is achieved by inspection of NDE data by a human operator, which inherently causes problems in consistency, especially when data is very complex. As the data volume gets larger, inspection by a single operator becomes very slow. This can be accelerated by multiple operators working in parallel, but the results then become even more inconsistent and analysis of data by either one or multiple skilled operators is expensive. These issues with consistency, data complexity, speed and cost are motivation for an automated defect characterization method. Since this is a pattern recognition task, machine learning is well suited.

Broadly, machine learning can be split into shallow and deep learning. Shallow learning gives a prediction based on features selected by the operator from the raw data, followed by statistical analysis, whereas deep learning is an end-to-end method that extracts the desired result directly from the raw data. Shallow learning in NDE dates back to at least 1991 with the use of decision trees to detect defects using the loss in amplitude of ultrasonic wall reflections [1] and has continued to be an active area of research [2]. For example, fully-connected neural networks have been used to classify defects [3], [4] and estimate material properties [5], [6] from ultrasonic measurements and support vector machines used to size cracks from eddy current field peaks [7].

Shallow learning is attractive as it reduces the dimensionality of the input, therefore also the complexity of the network needed to be trained, by manually selecting parameters. This in turn reduces the size of training data required. However, the success of the network is heavily reliant on selecting the correct parameters and a lot of information, that may well have been useful, is unused. Deep learning can make use of all available information and produce a more accurate result [8], [9] whilst also increasing automation, lowering the chance of human error causing mistakes [10]. The shift to deep learning is already well underway in medical imaging [11] but application in NDE has been hindered by the cost and difficulty of gathering enough defect data to train an effective network. Attempts to overcome this in the past have mostly relied on data augmentation (cropping, translating, zooming etc. in the context of photographic image analysis) to create a large pool of defect data from a handful of real defects [9], [12]–[14]. However, while for photographic images, zooming and translation produces realistic examples, this is not necessarily true for all

This research is funded by the Engineering and Physical Sciences Research Council (EPSRC, grant number EP/L015587/1) via the Research Centre for Non-Destructive Evaluation (RCNDE), with additional funding provided by Baker Hughes, Cramlington, UK.

Richard Pyle, Rhodri Bevan, Robert Hughes, Rosen Rachev and Paul Wilcox are with the Department of Mechanical Engineering, University of Bristol, UK

Richard Pyle, Rosen Rachev and Amine Ait Si Ali are with Baker Hughes, Cramlington, UK

This is a peer-reviewed, accepted author manuscript of the following article: Pyle, R., Bevan, R. L. T., Hughes, R. R., Rachev, R. K., Ali, A. A. S., & Wilcox, P. (2021). Deep learning for ultrasonic crack characterization in NDE. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control. https://doi.org/10.1109/TUFFC.2020.3045847

NDE modalities. With most NDE techniques there is rarely a linear relationship between defect and indication size, and non-uniform sensitivity maps [15] mean defect responses are not translationally invariant. In some situations deep learning has been applied to NDE where by good fortune or large expense a large enough number of defects are available [16], [17] or by limiting the scope of the problem and accepting a smaller training set [18]. However, none of these methods supply a solid, general case answer to creating a large enough training set for NDE deep learning.

This paper aims to show how accurate and efficient forward modeling of data can be used to train a deep learning network that can generalize to experimental data it has never seen before. While this approach could be applied to any inspection scenario, here the application considered is sizing surface-breaking cracks in ultrasonic inline pipe inspection. Note that to avoid confusion the term 'network' is used in this paper to refer to any machine learning predictor while 'model' is used exclusively to describe physics-based forward models such as Finite Element (FE) and ray tracing.

Ultrasonic inline pipe inspection uses transducers mounted on a 'pig' (pipeline inspection gauge) that travels in the flow of product to detect issues such as corrosion and cracks. The pig travels many kilometers, capturing data every 1-10 mm. When online processing flags a defect, the data from that position is compressed and stored for offline analysis, often involving significant operator input. Increases in computational power and efficient methodologies offer the potential for automated real-time analysis in the future. The pig considered here contains a ring of ultrasonic arrays that capture the data required to produce multiple overlapping images of any part of the cross-section of the pipe. This paper uses four distinct images per defect to predict its orientation and length. While imaging provides a large amount of data size reduction, learning from four images directly is still a very high dimensional problem. Convolutional Neural Networks [19] are a natural answer to this as they connect only nearby pixels at each layer, vastly reducing the complexity of the network. They have also seen widespread success with natural [20], medical [21]–[23] and NDE [13], [14], [16], [17] images in the past. There are many well-known Convolutional Neural Network (CNN) architectures for image characterization such as LeNet, DenseNet, Inception, AlexNet and ResNet [24]. The broad structure for the network used in this paper takes inspiration from networks such as AlexNet and VGG-19 and makes use of advances such as dropout, ReLu activations and max pooling [24] to assist in generalizing to experimental data after training on simulated data.

Typically, medical and NDE deep learning papers use 500-10,000 examples in their training sets. However, in the wider machine learning community sets such as ImageNet are being used that have more than 10,000,000 examples. It is generally accepted that the power of a deep learning network hinges heavily on the size of its training set. As in NDE, samples containing real or manufactured flaws are expensive, there is a shortage of experimental training data. This paper intends to show how by using forward models to create training sets the NDE community can begin to unlock the power of the state-of-the-art deep learning being used elsewhere. To simulate defect responses accurately a local FE model is used to generate scattering matrices [25]. Then, to efficiently create Plane Wave Capture (PWC) data a ray-based model [26] is used. The structural and grain noise is included by gathering experimental PWC data from a defect-free steel plate and adding it to the simulated data [27] which is then filtered and imaged. To cover a parameter space of varying crack length, angle and position this hybrid approach is used to generate 25,625 image sets that train the sizing network while 999 purely experimental sets from samples containing notches, made using Electrical Discharge Machining (EDM), are used to evaluate its accuracy. Section II describes the creation of these data sets. Section III outlines the deep learning method used to characterize them, as well as a more traditional sizing technique, the 6 dB drop method. Section IV presents results for the accuracy of the methods in sizing experimental defects and demonstrates the adaptability of the deep learning approach, by sizing defects imaged with incorrectly estimated sound speed. Overall success is judged by comparison in sizing error to the 6 dB drop method.

II. Inspection setup and Data Set Creation

In this section the inspection setup, imaging technique used, and the method for simulating a training set is described. The parameter space covered by the simulated and experimental sets is also given.

A. Inspection Setup

A major objective in inline pipe inspection is to detect and size the cracks that might occur on the outer or inner surfaces of the pipe. These are usually caused by manufacturing flaws such as weld toe cracks or lack of fusion, or in-service mechanisms such as stress corrosion cracking or thermal cycling fatigue and most commonly occur at the outer surface of the pipe. Sizing these surface-breaking defects, at the outer surface, with detection assumed already complete, is the focus of this paper. The experimental inspection set up used here to mimic in-line pipe inspection is described in Fig. 1a.

As pigs are typically used for pipes of least 40 cm diameter the effect of curvature is ignored, and a flat, 10 mm thick stainless-steel plate sample is used to represent the pipe wall material. Most pipelines for this application are made of carbon-steel and contain oil. Stainless-steel has been chosen for these samples to avoid corrosion and water used instead of oil to reduce cost. These replacements are acceptable as sound speeds, attenuation and levels of grain noise are comparable. Plane waves are fired in turn from a commercially available 5 MHz, 0.3 mm pitch, 40 element standard phased array at $\phi = 0°$ and $\pm 19°$ to the vertical in water to create longitudinal waves at $\psi = 0°$ and shear waves at $\pm 45°$ inside the stainless-steel sample. In this paper the two angled waves are used to characterize defects while the 0° wave is only used to calculate standoff (ς) and thickness (Γ). Note that as $\phi = 19°$ is beyond the first critical angle there are no longitudinal waves transmitted into the stainless-steel. The array receives on all elements individually with a sample rate of 16.7 MHz, a point

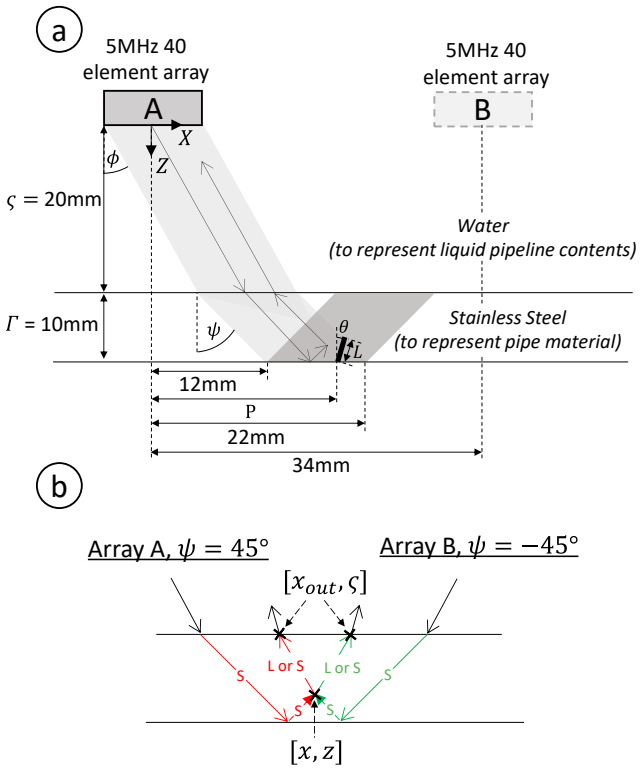

Fig. 1. a) A diagram of the inspection scenario using a plane wave at angle ψ to the vertical transmitted in the sample with a standoff and thickness of ς and Γ where L, θ and P represent the crack length, angle and position and b) all half skip shear (S) and longitudinal (L) mode ray-paths used in this paper where x, z are the co-ordinates of the imaging point and x_{out}, ς the co-ordinates of the returning ray on the front wall.

sufficiently above the center frequency of the array to capture most of its bandwidth without needlessly using excess data storage (an important resource in inline pipe inspection). Sound speeds in the steel are calculated using a calibration sample of known thickness (10 mm) and standoff (20 mm) giving longitudinal speed to be 5759 m/s, a shear speed of 3165 m/s and water speed to be 1480 m/s. All experimental defects are made using EDM to create 0.3 mm wide notches described by their angle from the vertical (θ), length (L) and horizontal distance from the array center (P), as indicated in Fig. 1a. Note that in this paper, crack length is defined as the distance between the defect's root and tip rather than its extent perpendicular to the surface. While EDM notches are simpler in shape and texture than most cracks found in in-service pipelines they allow for very accurate true length measurement. Research into the behaviour of this method when applied to more realistic cracks is left to future work.

B. Imaging

Plane Wave Capture (PWC) data can be focused on reception to create images, this is called Plane Wave Imaging (PWI) [28]. Multiple images, termed views, can be formed of the same physical region in the sample by considering different ray paths. Views are named according to the wave modes (L for longitudinal, S for shear) in the sample on the transmit leg and receive leg, with the two legs separated by a hyphen indicating the imaging point. In this application, half-skip shear ray-paths in transmission and direct shear or longitudinal ray-paths in reception have been found to provide the strongest signal response and clearest images of the defect, hence the views SS-L and SS-S and are used throughout. The intensity of the PWI image I for view SS-γ (where γ is L or S) at position x, z due to the plane wave at angle ψ in the sample is defined by

$$I_{\gamma,\psi}(x,z) = \left| \sum_j h_{j,\psi}(t_\psi^T + t_{j,\psi,\gamma}^R) \right| \quad (1)$$

where $h_{j,\psi}(t)$ is the complex, filtered A-Scan for receiving transducer j, and the ultrasonic transit times between the array and image point in transmission, t_ψ^T, and between the imaging point and receiving transducer, $t_{j,\psi,\gamma}^R$, are calculated using

$$t_\psi^T(x,z) = \frac{\varsigma}{c_c \cos(\phi)} + \frac{\Gamma}{c_S \cos(\psi)} + \frac{\varsigma + \Gamma - z}{c_S \cos(\psi)} \quad (2)$$

$$t_{j,\psi,\gamma}^R(x,z) = \frac{\sqrt{(x - x_{out})^2 + (z - \varsigma)^2}}{c_\gamma} + \frac{\sqrt{(x_{out} - x_j)^2 + (\varsigma)^2}}{c_c} \quad (3)$$

where ς is standoff, Γ is thickness, x_{out} is the position of the exiting ray on the front wall (as described in Fig. 1b), c_c is the speed of sound in the couplant, c_S is the shear speed in the sample and c_γ is the speed of the return ray. Note that x_{out} must be found using an iterative method such as Newton-Raphson [29] to minimize the time of flight between the imaging point and receiving transducer. The array is assumed to be parallel to the X-axis and positioned at $z = 0$.

C. Simulation

To simulate a training set large enough to enable deep learning in reasonable time, an efficient method is needed. To this end, a mixture of FE models in the defect locality and ray-based models for the whole region of interest are used. Structural and grain noise are included by summation of the simulated data with data collected from a defect-free sample. A flow chart describing this process is given in Fig. 2a.

The local FE model functions by exciting a scatterer with a uni-modal plane wave and recording the angle-dependent scattered wave amplitude to calculate its scattering matrix [25]. In this paper, the scatterers are surface-breaking cracks represented as 0.3 mm wide perfect reflectors with flat tips. The local FE model can be conducted independently of the ray-based model, so each defect length and angle combination need only to be simulated once, no matter where in the image the defect lies. This is significantly more efficient than using a fully FE approach, which would demand a model encompassing the whole region of interest to be run for all positions, as well as all lengths and angles. It should be noted that this method assumes that the receiving element is in the far field of the defect, where the scattered field decays monotonically with distance. In the immersion set-up considered here, the approach has been found to be applicable for defects of up to 5mm in length. For longer defects, the far-field assumption is not satisfied, and their simulated image responses are noticeably distorted.

A ray-based model is used to simulate the Full Matrix Capture (FMC) data, tracing all relevant paths (direct and half

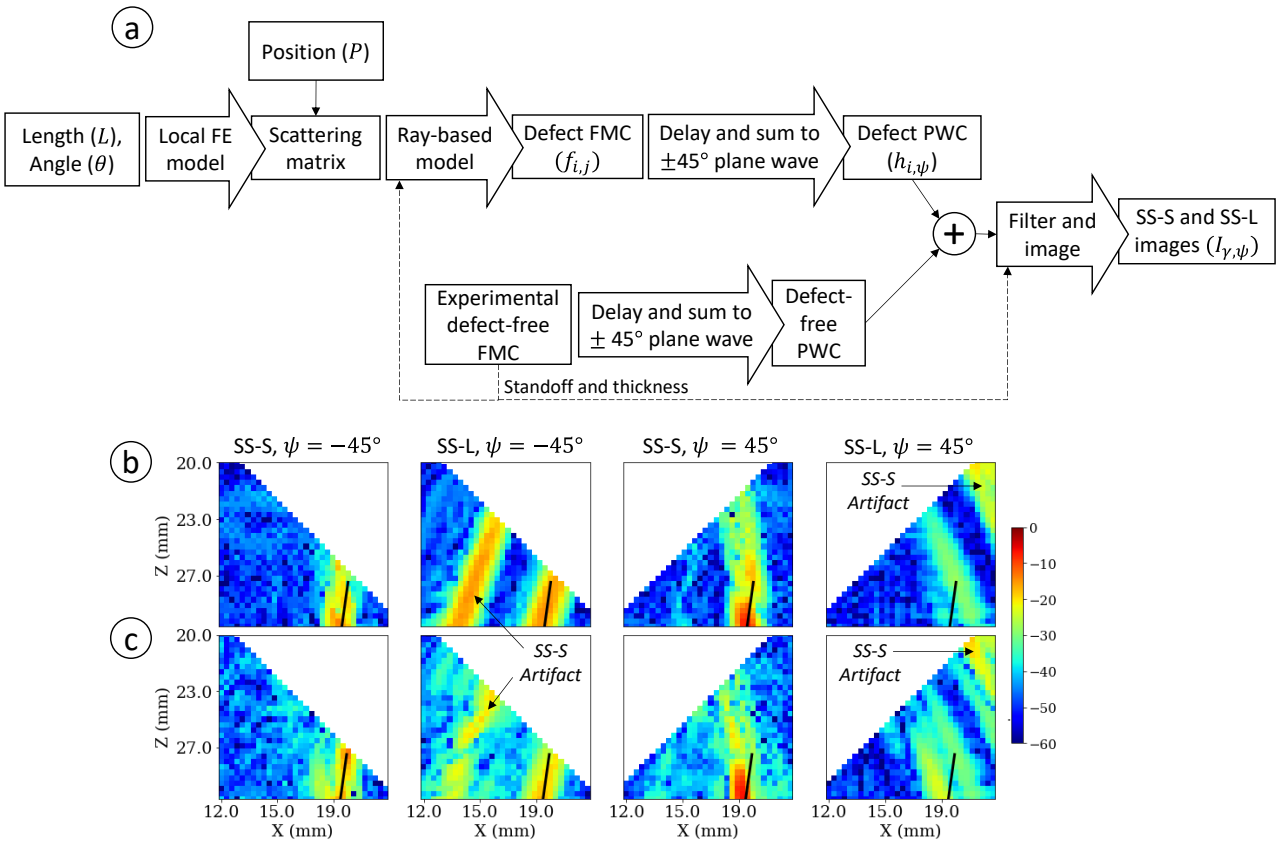

Fig. 2. a) A flow chart describing the method used to create the simulated data set given a defect's length (L), angle (θ) and position (P), b) an example set of simulated images for a defect with $L = 3$ mm, $\theta = 8°$ and $P = 19.2$ mm and c) a fully experimental set of images for a defect of the same parameters. Note that the black lines show the true extent of the defects and all images are on the same dB color scale, normalized to the maximum intensity in the experimental set.

skip) from the array to the defect and using scattering matrices to calculate the phase and amplitude of its reflections [26], [30]. The FMC dataset, $f_{i,j}$, is used to generate PWC data, $h_{j,\psi}$, with

$$h_{j,\psi}(t) = \sum_i f_{i,j}(t - \tau_{j,\psi}) \tag{4}$$

by summation over receiving transducers, i, where the appropriate delay, τ, is given by

$$\tau_{j,\psi} = (x_j - x_r)\frac{sin(\psi)}{c_S} \tag{5}$$

where x_r is the x position of an arbitrary reference element in the array, chosen to be the central element in this paper. Note that scattering induced attenuation has not been included in the simulation. This is because through the 10mm thick stainless-steel sample its effect is minimal.

Accurately representing structural and grain noise in the training data is achieved efficiently by collecting FMC datasets from a defect-free sample and combining them with the simulated defect data [27]. This is implemented here by choosing, at random, one of 24 FMC data sets obtained from a defect-free stainless-steel plate. To ensure the arrival times in these data sets match accurately with the simulated data, the standoff (ς) and sample thickness (Γ) used both in imaging and the ray-based model are calculated using the experimental 0° PWC data ($h_{j,0}$). These are calculated with

$$\varsigma = \frac{\sum_j t_{j,F}}{2Nc_w} \quad , \quad \Gamma = \frac{\sum_j (t_{j,B} - t_{j,BFB})}{2Nc_L} \tag{6}$$

where $t_{j,\mathrm{F}}$ and $t_{j,\mathrm{B}}$ are the arrival times of the front and backwall reflections, N is the number of transducers in the array and $t_{j,\mathrm{BFB}}$ is the arrival time of the first reverberation inside the sample. Amplitudes are ensured to be on the same scale by normalizing both defect and defect-free sets to a backwall reflection in the 0° PWC data set.

The resulting PWC data now contains signals due to grain noise, front and back wall reflections and all direct and half skip ray-paths from the defect. A Gaussian filter centered at 5 MHz with a -40 dB half width of 4.5 MHz is applied to the PWC data to remove data outside the frequency range of the transducer. Finally, PWI using Eq. (1) is used to create SS-S and SS-L images from the arrays on each side of the defect. An example of the resulting synthetic images and an equivalent, fully experimental set, are given in Fig. 2b,c. In these images, indications at the expected location of the defect are caused by corner reflections and tip diffractions while the 'artefacts' at other locations are due to these same effects but from a ray-path other than the one being imaged. For example, in Fig. 2c an artefact from the SS-S ray-path from the defect is very clear in the SS-L image.

TABLE I
SIMULATED DATA SET SUMMARY

Parameter	Range	Step	Number
Crack Length, L (mm)	0.2 to 5	0.2	25
Crack Position, P (mm)	13 to 21	0.2	41
Crack Angle, θ (°)	-24 to 24	2	25
Non-Defect Scan			24
Total = 25×41×25=**25,625 data sets**			

TABLE 2
EXPERIMENTAL DATA SET SUMMARY

		Crack Angle, θ (°)						
		0	±2	±5	±8	±12	±15	±20
Crack Length, L (mm)	1	✓						
	2	✓	✓	✓			✓	✓
	3	✓	✓	✓	✓	✓	✓	✓
	4	✓	✓	✓	✓	✓	✓	✓
	5	✓						
37 length/angle combinations								

	Range	Step	Number
Crack Position, P (mm)	13 to 21	0.2	27
Total = 37× 27=**999 data sets**			

D. Data Set Summary

The defects of interest are surface-breaking cracks between 1 and 5 mm in length, angled at most 20° from the vertical. The root of the defect can be positioned anywhere between $P = 13$ mm and $P = 21$ mm which corresponds to all bar 2 mm of the insonified backwall region. 25,625 image sets are simulated and are summarized in Table. I. The simulated set covers lengths and angles beyond that of the experimental set to ensure the resulting network learns across the full parameter space. Lengths larger than 5 mm are not included as they extend beyond the imaging domain, hence will be sized at 5 mm, a level which is already well above the critical crack length. Experimentally, FMCs are gathered with an array at 27 different positions (P) relative to each of the 21 available defects. The samples are rotated 180° to obtain data from defects with both positive and negative angles. The resulting 999 image sets are summarized in Table. II. The maximum intensity of defect indications in the simulated set is found to have a Mean Absolute Error (MAE) relative to the experimental set of 0.97%. Along with the visual similarity of the images such as in Fig. 2b,c this low level of MAE is considered to validate the simulation. The methodology for creating the standard data sets has now been described, the following subsection (Section II.E) outlines how this is altered to create sets with varying sound speeds.

E. Sound Speed Variation

With machine learning, creating a network that can cope with expected variations in inspection conditions is achieved by including these variations in the training set. Here the case of inaccurate knowledge of sound speeds is considered as an example. In practice, the variation would be in the physical measurements and the image reconstruction sound speeds would be fixed. However, because it is not readily possible to obtain a large amount of experimental data from physical systems with different sound speeds, the sound speeds used for image reconstruction are varied instead. Varying the reconstruction sound speed is not directly equivalent to varying the specimen sound speed, as the latter causes changes in physical quantities such as the crack length to wavelength ratio. However, in terms final image distortion, these are second order effects compared to a mismatch between the specimen and reconstruction sound speeds.

It is assumed that a sensor is available to get an accurate reading of temperature in the couplant from which its speed of sound can be estimated from previously acquired speed vs. temperature calibration data. Because in practice the pipeline product acts as the couplant, there will be some uncertainty in its sound speed due to uncontrolled variables, such as the exact composition of the product. Shear and longitudinal speeds in the steel pipe have larger potential uncertainty caused by effects such as variation in material composition, corrosion and temperature change due to the external environment. To include these variations in sound speed, random uniform multipliers are used at the imaging stage. These are

$$\begin{aligned} 0.99 < \beta_W < 1.01 \\ 0.9 < \beta_S < 1.1 \\ 0.9 < \beta_L < 1.1 \end{aligned} \tag{7}$$

where β_W, β_S and β_L are multipliers for the water speed, c_w, shear speed, c_S, and longitudinal speed, c_L, used in Eqs. (2-6). These values are larger than the true variation in material sound speed is likely to be; for example, carbon steel experiences less than a 10% variation in sound speed [31] across the full temperature range an inline pipe inspection tool is able to operate in (−10 to 50°C [32]). These large values of β are chosen to demonstrate the effectiveness of this method even under extreme conditions. As is evident from Eq. (6), the calculated thickness and standoff will change proportionally to longitudinal and water speeds, respectively. The coordinates of the imaging mesh are moved to consistently sit at the predicted position of the plane wave aperture on the backwall. An

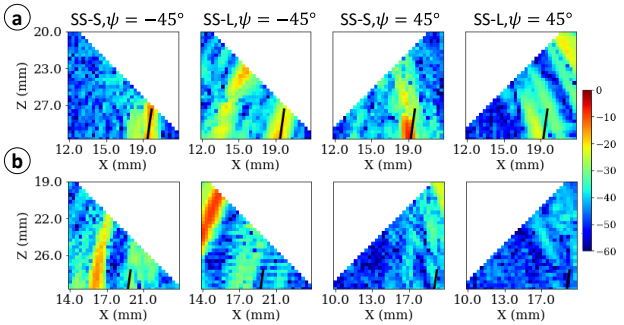

Fig. 3. a) An experimental image set for a defect with $L = 3$ mm, $\theta = 8°$, $P = 19.2$ mm and no sound speed variation and b) the same PWC data imaged with $\beta_S = 1.1, \beta_L = 0.9$ *and* $\beta_W = 0.99$. All images are on the same color scale in dB, normalized to the maximum intensity in the experimental set. The black lines show the true extent of the defect.

example of images produced with the most severe set of errors is given in Fig. 3b where it can be seen that the sound speed errors have caused significant spatial movement of defect responses, total loss of co-registration and a change in indication amplitude and size for some cases.

III. Defect-Characterization Algorithms

In this section, the process for implementing the 6 dB drop sizing method will be explained, the CNN architecture used will be described and the training method outlined.

A. 6 dB Drop Method

The 6 dB drop method is a common way to size defects in ultrasonic images and is presented here as a comparison for the deep learning approach. The 6 dB drop method is based upon the idea that if a defect is the strongest indicator in an image the region of the image that is within 6 dB of the peak value can be used as a good approximation of the size of the defect. This is implemented by calculating the minimum area of a rectangular box that encloses all pixels within 6 dB of the peak value and taking the crack length and angle as those of the major axis of the enclosing box [33]. Pixels above -6 dB must be within a certain distance of each other to be considered part of the same defect. In this paper the maximum distance is set at 4 pixels (1.27 mm). 6dB drop is deemed to be the most appropriate traditional sizing technique as amplitude based methods for large surface breaking defects suffer from constant amplitude corner reflections [34], [35], tip diffraction signals aren't consistently strong enough to enable temporal based techniques and the restricted range of incident and reflected angles means that scattering matrices [36] cannot be calculated. The reader is directed to [35] for a comprehensive review of traditional NDE sizing techniques.

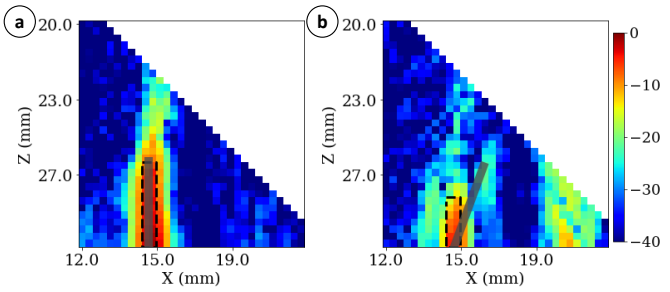

Fig. 4. Experimental SS-S images for a) a defect with $L = 4$ mm, $\theta = 0°$ and $P = 15$ mm, predicted by the 6 dB drop method to have $L = 3.5$ mm and $\theta = 0°$ and b) a defect with $L = 4$ mm, $\theta = 20°$ and $P = 15$ mm, predicted to have $L = 1.9$ mm and $\theta = 0°$. Black dashed lines indicate the box fit by the 6 dB drop method and grey solid lines the true extent of the defect. Note that the images are in dB, each image is normalized to its own maximum value.

This method has a few advantages over a machine learnt one. For example, it requires no tuning other than setting the range at which indications are considered to be from different defects, it is simple to execute, and is not a 'black box' method (a common criticism of deep learning). It can be argued to be a physics-based approach in that a single transducer above a large, planar defect will return half the amplitude when 'half on, half off' the defect compared to a measurement from directly above it [35], [37]. As this occurs at the edges of the defect the indication in a simple B-Scan should be described by a 6 dB drop. Figure 4a shows an experimental example where this works well, with the 6 dB box describing the extent of the defect quite accurately, undersizing by only 0.5 mm. However, this method performs poorly in more complex scenarios. For example, Fig. 4b shows a more angled defect from which the specular and tip reflections are well below -6 dB so only the corner indication is picked up, resulting in undersizing by 2.1 mm. Importantly, it is also difficult to make use of information from more than one image using 6 dB drop. In this application, defects of $|\theta| > 12°$ are much more accurately visualized in the SS-L view than SS-S, however, effectively deciding which one to use without prior knowledge of the defect is challenging. The SS-S view has been used for 6 dB sizing throughout this work as on average it gives a more accurate result.

B. Deep Learning

1) Network Architecture

The deep learning architecture used here is convolutional and loosely based on image recognition architectures such as AlexNet and VGG-19 due to their widespread success in image classification and regression [38]. Similar to these architectures,

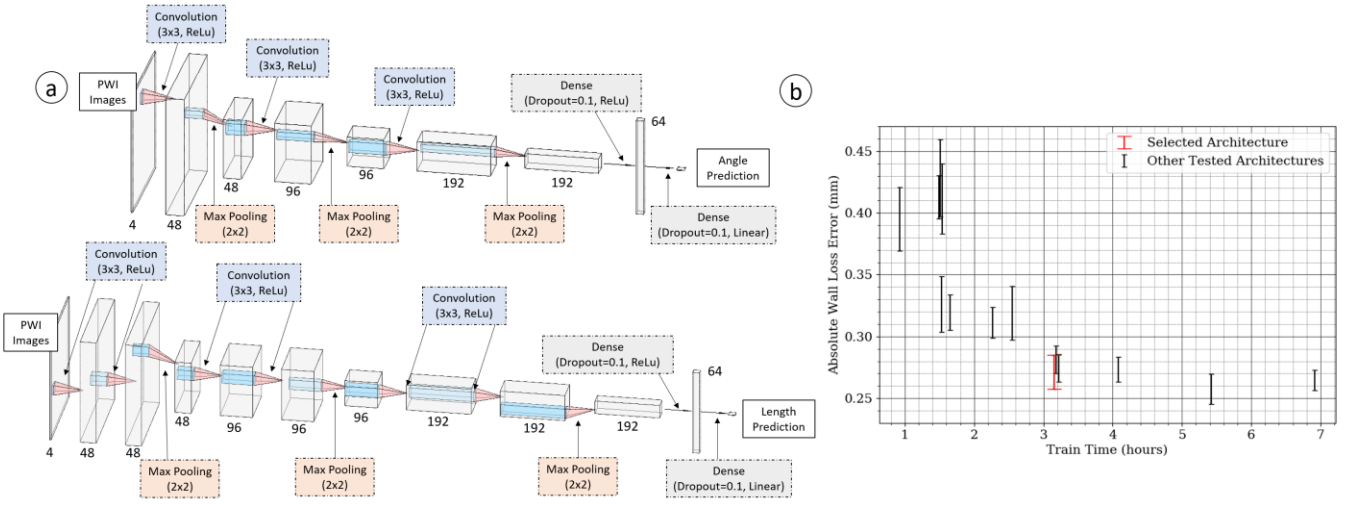

Fig. 5. a) An illustration of the chosen architecture and b) absolute experimental validation set wall loss error for all tested CNN architectures where the error bars represent ± standard deviation over 10 independent initializations. Hyperparameters for all tested architectures can be found in the accompanying data store (https://doi.org/10.5523/bris.1u376lnlam8ac2lf0uato86zrr).

sets of convolutional and max pooling layers with ReLu activation functions are used to achieve feature extraction and are followed by fully connected layers to predict the output. However, all hyperparameters have been tuned for this application. Directly using a well-known architecture 'off-the-shelf' is not possible as the images they are designed for are much larger in size than those used in this paper. It can also not be assumed that the most successful architecture for natural images will be the best choice for NDE images as their content is significantly different in structure. In this paper, as shown in Fig. 5a, dropout before each fully connected layer is used to minimize overfitting to the training set, 10% is chosen as this was found to be a good tradeoff between train time needed to converge and decrease in validation set loss (indicating reduced overfitting). To make use of all four images they are stacked at the input (akin to how natural image CNNs treat red, green and blue channels) producing a 32x32x4 input. This input is then fed into one network that predicts crack length and another network that predicts angle. These networks are decoupled to allow them to learn the image features that are most useful in predicting each property. Note that the outputs can take on any real value, making this a regression, rather than classification network.

The route to arriving at the final networks shown in Fig. 5a is by trialing different numbers of layers, filters and filter sizes to increase complexity (and therefore train time) until the improvement in accuracy is minimal. To simplify this analysis, length and angle predictions are combined into 'wall loss,' defined as $Lcos(\theta)$, which is usually the metric of interest when deciding upon the safety of a pipeline. The result of this study, using the experimental validation set to calculate wall loss, is shown in Fig. 5b with error bars representing the standard deviation in results over 10 independent initializations of each architecture using different starting weights and train/test/validate shuffles. Figure 5b shows that there is a diminishing return in adding complexity to the network. The architecture selected is chosen as it is accurate enough for this stage of research with more complex networks (with longer train times) giving minimal decrease in wall loss error. For real world implementation, as training need only occur once, more filters could be used at each stage to slightly increase prediction accuracy.

2) Training

For use in machine learning the simulated and experimental sets are each further split into two more. The simulated images are split 85% (21781 examples) for training and 15% (3843 examples) for validation while the experimental images are split 75% (749 examples) for testing and 25% (250 examples) for validation. The purpose of these sets are as follows:

Simulated, train: Used to iteratively update the weights and biases of the network.

Simulated, validation: Automatically analysed to implement the training stop condition, minimizing overfitting to the simulation.

Experimental, validation: Used during research and design stages to ensure the network is not overfitting to the simulated images and to implement the training stop condition.

Experimental, test: To evaluate the performance of the trained network on previously unseen data.

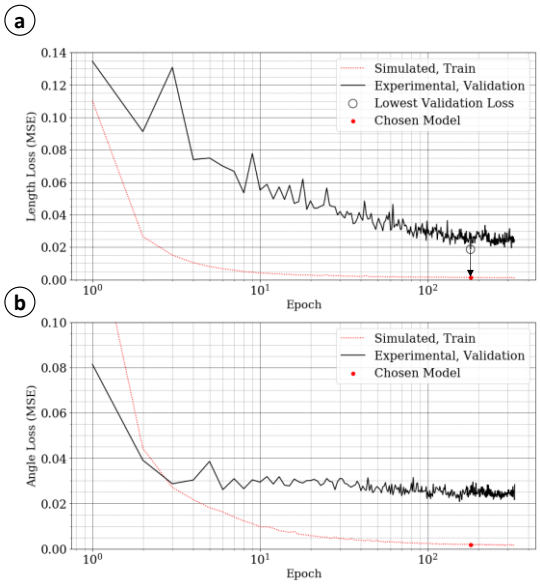

Fig. 6. An example pair of training progress graphs for a) length and b) angle predictions.

The CNN is implemented in Python using TensorFlow, trained using a Mean Square Error (MSE) loss function and the state-of-the-art 'Adam' optimizer [39] with a learning rate of 0.001, in mini-batches of 128 for a maximum of 400 epochs with a patience of 150. For machine learning terminology and definitions see [40]. This learning rate was selected by raising it from a low value until any significant instability in simulated training set loss appeared. The mini-batch size of 128 is selected to reduce overfitting without causing train time to increase

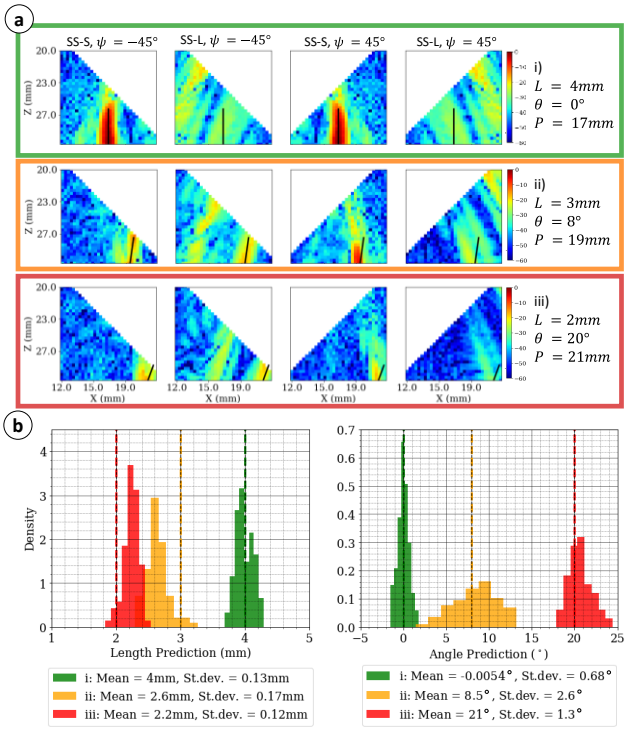

Fig. 7. a) Three experimental image sets with black lines indicating the true defect extent (all images are on the same color scale in dB, normalized to the maximum intensity in the experimental set) and b) a histogram showing the length and angle CNN predictions for these defects from 80 different training attempts. Dashed vertical lines represent true length and angle.

dramatically as while a small batch size gives a regularization effect it increases train time [41]. After training, the weights and biases chosen are from the point of minimum experimental validation loss. An example training progress graph is shown in Fig. 6 where it can be seen that a minima in experimental validation loss is reached at 180 epochs, past which point validation loss begins to increase due to overfitting to the simulation. On a workstation GPU (NVIDIA Quadro K620) training 400 epochs takes ~3 hours.

IV. Results and Discussion

As outlined in Section III.B, the weights within the CNN are initialized with a random seed. In addition, the assignment of a particular dataset to the train, test and validation sets is also random to avoid potential bias. The first consideration is therefore repeatability of the trained CNN. This section will also present and discuss length and angle prediction accuracy of the 6 dB drop method in comparison to that of the CNN both with and without errors in sound speed estimation.

A. Deep Learning Repeatability

It is important to know the amount of variation in accuracy over different network initializations as large scatter could suggest poor generalization. This is because in a wide distribution of test results the lower errors may be caused by fortuitous train/test splits rather than better networks. With low scatter, a higher level of confidence can be placed in the model's success not being due to overfitting. To test this, 80 networks are trained from different starting seeds and the spread of their results for three example defects are shown in Fig. 7. The low standard deviations in these results suggests that there is a big enough training set and enough network complexity for the network to generalize and the training to be satisfactorily independent of initial weights and train/test/validation data partitioning. This means that the final network can be picked at random from these 80 realizations.

The larger standard deviation in error for defect ii) compared to iii) is unintuitive as defect ii) has a higher Signal to Noise Ratio (SNR) and its indications better match its true size. Investigation into this found that experimental defects of $8 \leq |\theta| \leq 15$ cause weak reverberations from ray paths not considered in the simulation. These are very low in amplitude relative to the SS-S and SS-L views but cause an average SNR drop of 2 dB across these angles. While this is a small value for high SNR data like this (~30 dB) it is hypothesized to be the cause of the larger spread in error for defect ii). This finding highlights the importance of an accurate simulation. Further research into the cause of these reverberations will allow them to be modelled in the future.

B. Deep Learning vs 6 dB Drop Accuracy

Figure 8a shows the error in characterizing the experimental test set using the 6 dB drop method, a CNN trained without any variation in training set sound speeds and a CNN trained with the sound speed variation described in Section II.E. Table. III gives the mean and standard deviation of these prediction

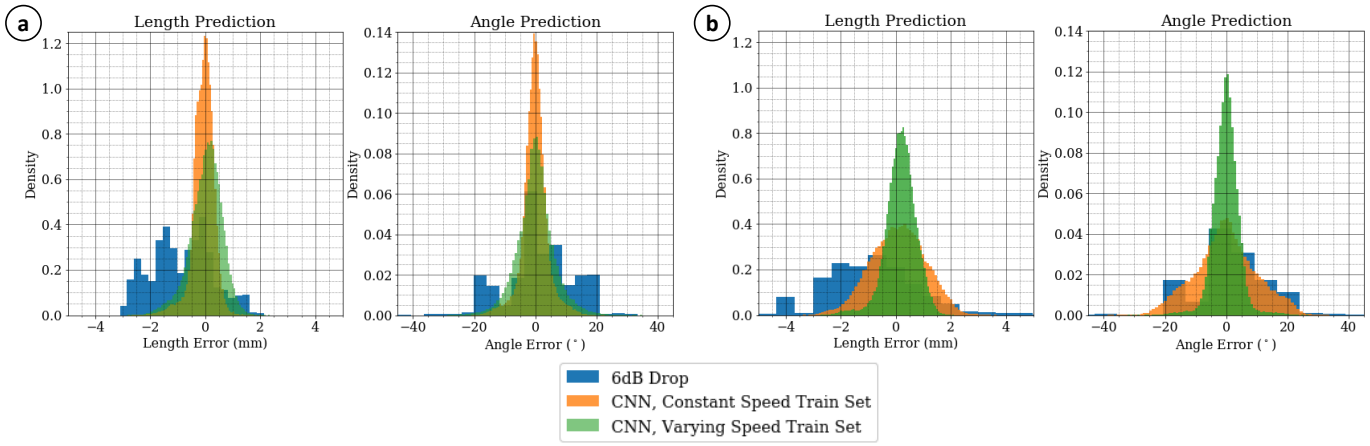

Fig. 8. Histograms of length and angle prediction error of methods applied to a) the standard experimental test set and b) the experimental test set with sound speed variation applied. Note that for CNN results all predictions from 80 independent initializations are shown.

TABLE III
Mean and Standard Deviation of Prediction Errors

		Characterization Algorithm					
		6 dB Drop		CNN (Standard Train Set)		CNN (Speed Varying Train Set)	
		Mean	*St.Dev.*	*Mean*	*St.Dev.*	*Mean*	*St.Dev.*
Test Set, Quantity	Speed Constant, Length (mm)	-0.86	1.1	-0.063	0.39	0.03	0.59
	Speed Varying, Length (mm)	-0.78	1.8	0.088	0.98	0.18	0.56
	Standard, Angle (°)	1.4	12	-0.13	4.1	0.062	4.1
	Speed Varying, Angle (°)	-2	20	-0.15	10	-0.048	4.2

errors. In terms of length prediction, the 6 dB drop method shows a non-negligible mean prediction error of -0.86 mm so is on average under-sizing the cracks. It also has a significant standard deviation of 1.1 mm. Both the standard CNN and the CNN trained with speed variation outperform this with near-zero mean error and standard deviations of 0.39 mm and 0.59 mm, respectively. The results for angle follow a similar pattern. The most successful method for this test set is the standard CNN that has 95% confidence intervals of ± 0.77 mm and $\pm 8.0°$.

Figure 8b shows the performance of the same methods on the experimental test set with sound speed variation included. As shown in Table. III, the standard deviations in length and angle prediction for the 6 dB drop method rise by 64% and 67%, respectively, compared to results on the standard test set. The standard CNN more than doubles in standard deviation. However, while adding sound speed to the training set increases errors for the standard test set the increased generality it creates means that adding speed variation to the test set decreases standard deviation by 5% for length and 31% for angle. This results in a network with 95% confidence intervals of ± 1.1 mm and $\pm 8.2°$ even with uncertainties in material sound speed up to 10%.

Whilst the results presented are for a relatively coarse imaging grid (pixel size = $\frac{\lambda_S}{2} = 0.317$ mm), a finer grid (pixel size = $\frac{\lambda_S}{6}$) provided negligible improvement for either the CNN or dB drop methodologies. For the 6 dB drop, this is because the limitation is accuracy rather than precision, evidenced by 81% of absolute length errors in Fig. 8a being larger than the coarse image pixel size. The standard CNN sizes much fewer defects with errors larger than a pixel (34% in Fig. 8a), but its prediction is not intrinsically based on distances in the image so is harder to relate to the pixel size. Furthermore, as the chosen resolution is already at the diffraction limit ($\frac{\lambda_S}{2}$) reducing it does not provide any further information about the defect to the network.

C. Discussion

This paper shows, once again, that avoiding overfitting is key to the success of deep learning. While this is common knowledge within the machine learning community, its importance cannot be overstated. This is of even more importance when training on simulated data, as for the network to be useful it must be able to operate on real data, despite any simplifications or assumptions the simulation may make. Use of dropout, analysis of validation data and careful training set creation is essential. It must also be ensured that the training set contains all significant variation that is expected to occur in the real inspection. This is demonstrated here with sound speed variation in the training set, but the principle extends to many other properties such as variable attenuation, standoff, surface roughness and array alignment. It is worth noting that finding which simulation inaccuracies cause significant errors in experimental sizing is difficult and not always intuitive. This is exampled in Section IV.A where reverberations not included in the simulation, despite being weak relative to the defect's half skip response, cause non-negligble decreases in angle prediction accuracy. Ultimately, the main limitation of this method is the breadth and accuracy of the training set. Including the correct variation, or somehow accounting for the deficiencies of your simulation, is key to creating a network that is applicable to real data.

Due to its simplicity the 6 dB drop method is computationally inexpensive. However, it is shown to give far less accurate predictions than the CNN. A large factor in this is the quantity of information available to each sizing algorithm. While the 6 dB drop method must size a crack solely from its shape in one image the CNN is able to take information from the amplitude and shape of indications and artefacts in multiple images. This could be further capitalized upon if applied to situations with more views such as multi-mode Total Focusing Method (TFM) [39]. However, deep learning is not without its drawbacks as it is often perceived to be a 'black box' method. This makes it difficult to directly relate its predictions to their cause. For a conservative field like NDE where historically, inspections have been qualified using physics-based reasoning this is a big drawback, but one that could be overcome by ongoing work in techniques such as activation and feature visualization [42]–[44] which provide mechanisms to understand the rationale behind the final sizing choice. How deep learning is integrated into the workplace must also be done carefully. Making use of its predictions without introducing unwanted bias or degrading human skills through overreliance are issues demanding thought and care. However, as current deep learning application to safety-critical problems such as self-driving cars has proved, this is certainly achievable. Therefore, the significant increase in characterization accuracy compared to current methods that this paper has presented are a strong motivation for application and further research of deep learning for NDE.

V. Conclusion

This paper has demonstrated how a simulation approach can generate the large training datasets which enable deep learning for crack characterization. The resulting CNN sizes 97% of the tested experimental defects of length 1 to 5 mm within ± 1 mm while the 6 dB drop method only achieves 48%. Even with a maximum of 10% uncertainty in material sound speed the CNN still achieves 91% sizing in the ± 1 mm range, while the 6 dB method drops to 40%. Future research should be carried out in testing the adaptability and limits of this method by characterizing a wider range of defects such as branching cracks, corrosion and cracks at welds. The network could also be improved by exploring methods to add an output that indicates a level of confidence in its characterization of each defect. The deep learning characterization approach identified in this paper is demonstrated to be successful for in-line pipe inspection and is readily applicable to other ultrasonic NDE inspections.

Appendix

Supporting code and data are available at the University of Bristol data repository, data.bris at:
https://doi.org/10.5523/bris.1u376lnlam8ac2lf0uato86zrr

REFERENCES

[1] P. O'Rorke, S. Morris, M. Amirfathi, W. Bond, and D. St. Clair, "Machine Learning for Nondestructive Evaluation," in *Machine Learning Proceedings 1991*, Elsevier, 1991, pp. 620–624.

[2] J. B. Harley and D. Sparkman, "Machine learning and NDE: Past, present, and future," in *AIP Conference Proceedings*, 2019, vol. 2102, no. 1, p. 90001.

[3] S. Sambath, P. Nagaraj, and N. Selvakumar, "Automatic defect classification in ultrasonic NDT using artificial intelligence," *J. Nondestruct. Eval.*, vol. 30, no. 1, pp. 20–28, 2011.

[4] L. Udpa and S. S. Udpa, "Neural networks for the classification of nondestructive evaluation signals," *IEE Proceedings, Part F Radar Signal Process.*, vol. 138, no. 1, pp. 41–45, 1991, doi: 10.1049/ip-f-2.1991.0007.

[5] N. Amiri, G. H. Farrahi, K. R. Kashyzadeh, and M. Chizari, "Applications of ultrasonic testing and machine learning methods to predict the static & fatigue behavior of spot-welded joints," *J. Manuf. Process.*, vol. 52, pp. 26–34, Apr. 2020, doi: 10.1016/j.jmapro.2020.01.047.

[6] M. Mishra, A. S. Bhatia, and D. Maity, "Predicting the compressive strength of unreinforced brick masonry using machine learning techniques validated on a case study of a museum through nondestructive testing," *J. Civ. Struct. Heal. Monit.*, pp. 1–15, Mar. 2020, doi: 10.1007/s13349-020-00391-7.

[7] A. Bernieri, L. Ferrigno, M. Laracca, and M. Molinara, "Crack shape reconstruction in Eddy current testing using machine learning systems for regression," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 9, pp. 1958–1968, 2008, doi: 10.1109/TIM.2008.919011.

[8] Z. Lin, H. Pan, G. Gui, and C. Yan, "Data-driven structural diagnosis and conditional assessment: from shallow to deep learning," in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2018*, Mar. 2018, vol. 10598, p. 38, doi: 10.1117/12.2296964.

[9] J. Ye, S. Ito, and N. Toyama, "Computerized ultrasonic imaging inspection: From shallow to deep learning," *Sensors (Switzerland)*, vol. 18, no. 11, Nov. 2018, doi: 10.3390/s18113820.

[10] D.-P. Marija Bertović, "Human Factors in Non-Destructive Testing (NDT): Risks and Challenges of Mechanised NDT," 2016. Accessed: Jun. 01, 2020. [Online]. Available: www.bam.de.

[11] K. Suzuki, "Overview of deep learning in medical imaging," *Radiological Physics and Technology*, vol. 10, no. 3. Springer Tokyo, pp. 257–273, Sep. 01, 2017, doi: 10.1007/s12194-017-0406-5.

[12] I. Virkkunen, T. Koskinen, O. Jessen-Juhler, and J. Rinta-Aho, "Augmented Ultrasonic Data for Machine Learning," *arXiv Prepr. arXiv1903.11399*, 2019.

[13] J. C. Aldrin and D. S. Forsyth, "Demonstration of using signal feature extraction and deep learning neural networks with ultrasonic data for detecting challenging discontinuities in composite panels," in *AIP Conference Proceedings*, May 2019, vol. 2102, no. 1, doi: 10.1063/1.5099716.

[14] H. wei Huang, Q. tong Li, and D. ming Zhang, "Deep learning based image recognition for crack and leakage defects of metro shield tunnel," *Tunn. Undergr. Sp. Technol.*, vol. 77, pp. 166–176, Jul. 2018, doi: 10.1016/j.tust.2018.04.002.

[15] N. Budyn, R. L. T. Bevan, J. Zhang, A. J. Croxford, and P. D. Wilcox, "A Model for Multiview Ultrasonic Array Inspection of Small Two-Dimensional Defects," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 66, no. 6, pp. 1129–1139, Jun. 2019, doi: 10.1109/TUFFC.2019.2909988.

[16] J. Feng, F. Li, S. Lu, J. Liu, and D. Ma, "Injurious or noninjurious defect identification from MFL images in pipeline inspection using convolutional neural network," *IEEE Trans. Instrum. Meas.*, vol. 66, no. 7, pp. 1883–1892, 2017.

[17] Y.-J. Cha, W. Choi, and O. Büyüköztürk, "Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks," *Comput. Civ. Infrastruct. Eng.*, vol. 32, no. 5, pp. 361–378, May 2017, doi: 10.1111/mice.12263.

[18] M. Meng, Y. J. Chua, E. Wouterson, and C. P. K. Ong, "Ultrasonic signal classification and imaging system for composite materials via deep convolutional neural networks," *Neurocomputing*, vol. 257, pp. 128–135, 2017.

[19] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Computation*, vol. 29, no. 9. MIT Press Journals, pp. 2352–2449, Sep. 01, 2017, doi: 10.1162/NECO_a_00990.

[20] A. Bhandare, M. Bhide, P. Gokhale, and R. Chandavarkar, "Applications of convolutional neural networks," *Int. J. Comput. Sci. Inf. Technol.*, vol. 7, no. 5, pp. 2206–2215, 2016.

[21] B. Kayalibay, G. Jensen, and P. van der Smagt, "CNN-based segmentation of medical imaging data," *arXiv Prepr. arXiv1701.03056*, 2017.

[22] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *2014 13th international conference on control automation robotics & vision (ICARCV)*, 2014, pp. 844–848.

[23] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.

[24] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *arXiv Prepr. arXiv1901.06032*, 2019.

[25] P. D. Wilcox and A. Velichko, "Efficient frequency-domain finite element modeling of two-dimensional elastodynamic scattering," *J. Acoust. Soc. Am.*, vol. 127, no. 1, pp. 155–165, Jan. 2010, doi: 10.1121/1.3270390.

[26] R. K. Rachev, P. D. Wilcox, A. Velichko, and K. L. McAughey, "Plane Wave Imaging Techniques for Immersion Testing of Components with Non-Planar Surfaces," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, pp. 1–1, Jan. 2020, doi: 10.1109/tuffc.2020.2969083.

[27] H. A. Bloxham, A. Velichko, and P. D. Wilcox, "Combining simulated and experimental data to simulate ultrasonic array data from defects in materials with high structural noise," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 63, no. 12, pp. 2198–2206, 2016.

[28] L. Le Jeune, S. Robert, E. L. Villaverde, and C. Prada, "Plane Wave Imaging for ultrasonic non-destructive testing: Generalization to multimodal imaging," *Ultrasonics*, vol. 64, pp. 128–138, 2016.

[29] R. McGill and P. Kenneth, "Solution of variational problems by means of a generalized Newton-Raphson operator," *Aiaa J.*, vol. 2, no. 10, pp. 1761–1766, 1964.

[30] L. W. Schmerr, *Fundamentals of ultrasonic nondestructive evaluation*. Springer, 2016.

[31] Z. Liu, J. Zhao, B. Wu, and C. He, "Temperature dependence of ultrasonic longitudinal guided wave propagation in long range steel strands," *Chinese J. Mech. Eng. (English Ed.*, vol. 24, no. 3, pp. 487–494, May 2011, doi: 10.3901/CJME.2011.03.487.

[32] N. I. Uzelac, K. Reber, M. Belter, and O. A. Barbian, "Ultrasonic In-Line Inspection of Pipelines, New Generation of Tools," in *Rio Pipeline Conference*, 2003.

[33] J. Zhang, B. W. Drinkwater, and P. D. Wilcox, "The use of ultrasonic arrays to characterize crack-like defects," *J. Nondestruct. Eval.*, vol. 29, no. 4, pp. 222–232, 2010.

[34] M. V. Felice, "Ultrasonic array inspections for complex defects." University of Bristol, 2015.

[35] M. V Felice and Z. Fan, "Sizing of flaws using ultrasonic bulk wave testing: A review," *Ultrasonics*, vol. 88, pp. 26–42, 2018.

[36] L. Bai, A. Velichko, and B. W. Drinkwater, "Ultrasonic characterization of crack-like defects using scattering matrix similarity metrics," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 62, no. 3, pp. 545–559, 2015.

[37] J. Krautkrämer and H. Krautkrämer, "Ultrasonic testing by determination of material properties," in *Ultrasonic Testing of Materials*, 4th ed., Springer, 1990, pp. 319–326.

[38] M. Z. Alom *et al.*, "The History Began from AlexNet: A Comprehensive Survey on Deep Learning Approaches," Mar. 2018, Accessed: May 18, 2020. [Online]. Available: http://arxiv.org/abs/1803.01164.

[39] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, Dec. 2015.

[40] Google, "Machine Learning Glossary | Google Developers," *Machine Learning Crash Course*. https://developers.google.com/machine-learning/glossary (accessed Jun. 10, 2020).

[41] D. R. Wilson and T. R. Martinez, "The general inefficiency of batch training for gradient descent learning," *Neural networks*, vol. 16, no. 10, pp. 1429–1451, 2003.
[42] Q. Zhang, Y. Nian Wu, and S.-C. Zhu, "Interpretable convolutional neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8827–8836.
[43] C. Olah, A. Mordvintsev, and L. Schubert, "Feature Visualization," *Distill*, vol. 2, no. 11, p. e7, Nov. 2017, doi: 10.23915/distill.00007.
[44] S. Carter, Z. Armstrong, L. Schubert, I. Johnson, and C. Olah, "Activation Atlas," *Distill*, vol. 4, no. 3, p. e15, Mar. 2019, doi: 10.23915/distill.00015.

Richard J. Pyle was born in Torquay, U.K. in 1996. He received an M.Eng. degree in mechanical engineering from The University of Bristol, U.K. in 2018.

Through the summer of 2017 he worked for Cavendish Nuclear as a graduate design engineer. He is now studying for an Eng.D. degree in ultrasonic phased array signal processing at The University of Bristol, sponsored by Baker Hughes, Cramlington, U.K. His current research interests include phased array imaging, data compression, defect characterization and machine learning.

Rhodri L. T. Bevan received an M.Eng. degree in Civil Engineering in 2005, an M.Res. in Computational Engineering in 2006 and a Ph.D. from Swansea University (Swansea, UK) in 2011.

From 2011 to 2013 he was a Research Assistant in the Civil and Computational Engineering Centre at Swansea University developing high-performance finite element programs. From 2013 to 2016 he was a Research Assistant in Numeric Simulation and Optimisation in the Department of Aerospace Engineering at the University of Bristol. Since 2016, Dr. Bevan has been with the Department of Mechanical Engineering at the University of Bristol and his current research interests include array imaging, data fusion and high-performance algorithm design.

Robert R. Hughes was born in Bristol, U.K., in 1989. He received an M.Phys. degree in physics followed by an Engineering Doctorate (Eng.D.) in non-destructive evaluation from the Department of Physics, University of Warwick, in 2016. His Eng.D. research was sponsored by Rolls-Royce plc., Bristol, where he carried out an industrial placement between 2014 and 2015 focusing on eddy-current array sensor development and data-analysis.

In 2015, Dr. Hughes took up a position as Research Associate with the Department of Mechanical Engineering, University of Bristol, U.K, where he developed eddy-current inspection and data-analysis techniques for characterising surface-breaking defects and carbon-fibre composite structures. From 2019, Dr. Hughes has been a Lecturer in non-destructive testing at the Department of Mechanical Engineering, University of Bristol, U.K where his current research interests include eddy-current inspection, inversion of inhomogeneous materials, defect characterisation and advanced data-analysis techniques, as well as magnetic particle sensing & manipulation in microfluidic environments.

Rosen K. Rachev was born in Burgas, Bulgaria, in 1993. He received the M.Eng. degree in mechanical engineering from the University of Bristol, Bristol, U.K., in 2016, where he is currently pursuing the Eng.D. degree in nondestructive evaluation with Ultrasonics and Non-destructive Testing Research Group.

He is currently working on advanced ultrasonic data processing algorithms for pipeline in-line inspections, and his project is sponsored by Baker Hughes, Cramlington, U.K. His research interests include nondestructive evaluation for the oil and gas industry, ultrasonic phased array surface reconstruction, adaptive imaging, and defect characterization.

Amine Ait Si Ali received the Dipl.-Ing. (M.Eng.) degree in computer science from the University of Science and Technology Houari Boumediene, Algiers, Algeria, in 2009, the M.Sc. degree in embedded intelligent systems from the University of Hertfordshire, Hatfield, U.K, in 2012, and the Ph.D. degree in computer science from the University of the West of Scotland, Paisley, U.K., in 2016.

He took many research positions, including Research Assistant with the School of Engineering, Qatar University, Doha, Qatar and KTP Associate with the Department of Computer and Information Sciences, University of Northumbria, Newcastle upon Tyne, U.K. He is currently a Data Scientist with Process & Pipeline Services, Digital Solutions, Baker Hughes, Cramlington, U.K. His research interests are mainly in machine learning, big data, cloud computing, non-destructive testing, connected health and

custom computing using FPGAs and heterogeneous embedded systems

Paul D. Wilcox was born in Nottingham (England) in 1971. He received an M.Eng. degree in Engineering Science from the University of Oxford (Oxford, England) in 1994 and a Ph.D. from Imperial College (London, England) in 1998. He remained in the Non-Destructive Testing (NDT) research group at Imperial College as a Research Associate until 2002, working on the development of guided wave array transducers for large area inspection.

Since 2002 Prof. Wilcox has been with the Department of Mechanical Engineering at the University of Bristol (Bristol, England) where his current title is Professor of Dynamics. He held an EPSRC Advanced Research Fellowship in Quantitative Structural Health Monitoring from 2007 to 2012, was Head of the Mechanical Engineering Department from 2015 to 2018, and has been a Fellow of the Alan Turing Institute for Data Science since 2018. In 2015 he was a co-founder of Inductosense Ltd., a spin-out company which is commercialising inductively-coupled embedded ultrasonic sensors. His research interests include array transducers, embedded sensors, ultrasonic particle manipulation, long-range guided wave inspection, structural health monitoring, elastodynamic scattering and signal processing.