Depth Evaluation for Metal Surface Defects by Eddy Current Testing Using Deep Residual Convolutional Neural Networks

Tian Meng, Yang Tao, Ziqi Chen, Jorge R. Salas Avila, Qiaoye Ran, Yuchun Shao, Ruochen Huang, Yuedong Xie, *Member, IEEE*, Qian Zhao, Zhijie Zhang, Hujun Yin, *Senior Member, IEEE*, Anthony J. Peyton, and Wuliang Yin, *Senior Member, IEEE*

***Abstract*— Eddy current testing (ECT) is an effective technique for evaluating depth of metal surface defects. However, in practice, evaluation primarily relies on the experience of an operator and is often carried out by manual inspection. In this article, we address the challenges of automatic depth evaluation of metal surface defects by virtual of state-of-the-art deep learning (DL) techniques. The main contributions are threefold. First, a highly integrated portable ECT device is developed, taking the advantage of an advanced field-programmable gate array (Zynq-7020 system on chip) and provides fast data acquisition and in-phase/quadrature demodulation. Second, a dataset, termed metal defects of different depths by ECT (MDDECT), is constructed using the ECT device by human operators and made openly available. It contains 48 000 scans from 18 defects of different depths and liftoffs. Third, the depth evaluation problem is formulated as a time series classification problem, and various state-of-the-art 1-D residual convolutional neural networks are trained and evaluated on the MDDECT dataset. A 38-layer 1-D ResNeXt achieves an accuracy of 93.58% in discriminating the surface defects in a stainless steel sheet with depths from 0.3 to 2.0 mm in the resolution of 0.1 mm. In addition, the results show that the trained ResNeXt1D-38 model is immune to liftoff signals.**

***Index Terms*— Convolutional neural network, deep learning (DL), eddy current testing (ECT), metal surface defect evaluation, nondestructive testing (NDT).**

I. Introduction

EDDY current testing (ECT) is a nondestructive testing (NDT) method harnessing the principle of electromagnetic induction, which, compared to other NDT methods, has the virtue of high speed, low cost, and no contact [1]. These features make ECT an attractive technique in the detection and evaluation of surface defects for conductive materials [2]. Recovering the profiles of a defect, e.g., location and depth, from sensor impedance signals is a major topic in the research of ECT, where machine learning (ML) plays an important role [3].

Conventional ML algorithms have been adopted in various ECT applications, and many of these studies generally utilized a two-step approach. First, raw eddy current (EC) signals would be subject to a feature transform or extraction process, such as principal component analysis [4]–[9], time–frequency analysis by the Rihaczek distribution [10], wavelet transform [4], [11], [12], Hilbert–Huang transform [13], [14], geometry recognition from the Lissajous figure [15]–[20], and convolutional sparse coding [21]. Next, the resultant feature representations, in order to achieve the ultimate task of detecting and classifying defects, would be fed to a classification or clustering algorithm such as support vector machine (SVM) [7], [11], [14], [15], [20], multilayer perceptron [4], [8], [11], [12], [15]–[19], K-means [6], [10], K-nearest neighbors [11], [20], decision tree [17]–[20], and naive Bayes [17]–[20]. These conventional ML algorithms still remain vibrant today in the research of ECT, and yet, deep learning (DL) methods prevail more recently, which is encouraged by the remarkable success of DL in many other areas.

A deep belief network (DBN), constructed by stacking multiple restricted Boltzmann machines, was exploited in [22] so as to, from the EC scan images of the defects on the surface of a titanium sheet, extract features that were then fed to a least-square SVM algorithm to classify the defects. The dataset was also evaluated in [23] with a plain convolutional neural network (CNN), which, in contrast to [22] where the feature extractor and classifier were separate, was trained end-to-end from EC signals to the classification labels of defects. Classification accuracy increased to 99.79% from the 96% in [22]. In [24], an encoder–decoder CNN, named EddyNet, was proposed aiming at learning an inverse model, which predicted a crack profile given an EC signal. However, the inverse model was more difficult to solve than the associated forward model. As a result, training samples were procured from the forward model, with inputs and outputs exchanged, which

Manuscript received March 26, 2021; revised July 6, 2021; accepted September 22, 2021. Date of publication October 4, 2021; date of current version October 15, 2021. This work was supported in part by the U.K. Engineering and Physical Sciences Research Council through the Project [Real-time Inline Microstructural Engineering (RIME)] under Grant EP/P027237/1 and in part by the U.K. Research Centre in Non-Destructive Evaluation (RCNDE) under Contract EP/L022125/1. The Associate Editor coordinating the review process was Dr. Sasan Bakhtiari. *(Corresponding author: Yang Tao.)*

Tian Meng, Yang Tao, Ziqi Chen, Jorge R. Salas Avila, Qiaoye Ran, Yuchun Shao, Ruochen Huang, Hujun Yin, Anthony J. Peyton, and Wuliang Yin are with the Department of Electrical and Electronic Engineering, School of Engineering, The University of Manchester, Manchester M13 9PL, U.K. (e-mail: mchikyt3@gmail.com).

Yuedong Xie is with the School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing 100191, China.

Qian Zhao is with the College of Engineering, Qufu Normal University, Jining, Shandong 273165, China.

Zhijie Zhang is with the School of Instrument and Electronics, North University of China, Taiyuan, Shanxi 030051, China.

Digital Object Identifier 10.1109/TIM.2021.3117367

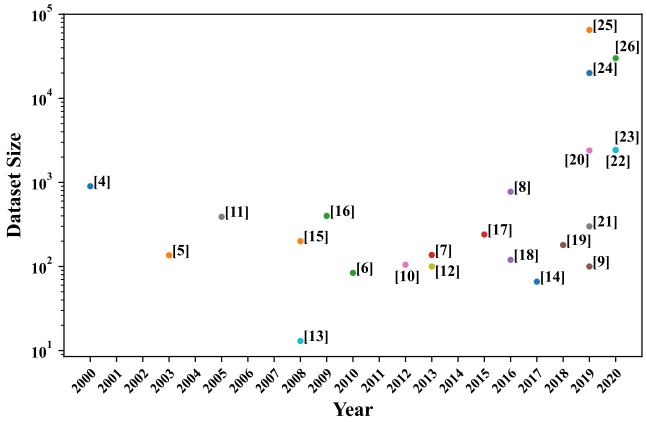

Fig. 1. Summary of the sizes of datasets used in [4]–[26]. In terms of the conventional ML works before 2019, the sizes of the datasets are below 1000. However, in 2019 and 2020, the numbers of training samples are more than 20 000 in the DL-motivated works [24]–[26].

were randomly generated crack profiles and the corresponding EC signals, respectively. The trained CNN achieved a mean absolute error of 0.198 between the predicted profiles and ground truth. In terms of pulsed ECT, a multitask CNN was developed in [25], which installed a softmax layer and a full connection layer as two outputs in order to simultaneously classify the type and predict the depth of flaw, respectively. The total loss was an addition of the classification loss and regression loss. Accuracies of over 90% were obtained for different experiments. In [26], a plain CNN was used to estimate the crack depth for a heat transfer tube of the steam generator of a pressurized water reactor. Compared to conventional numerical models, the trained CNN was less computationally expensive at inference time. In addition, crack depth could be estimated with a high accuracy of about 0.05-mm error under the condition that the liftoff fluctuation was about ± 1 mm. These DL-motivated studies shared some common features. First, they all entailed a larger dataset compared to those using conventional ML algorithms, which can be seen from Fig. 1 that displays the sizes of the datasets in [4]–[26]. Specifically, the numbers of training samples in [24]–[26] were more than 20 000, while most others were a few hundreds. Second, the samples were labeled into multiple classes and the training was in a supervised manner. Finally, an end-to-end strategy was adopted, and domain knowledge- or hand engineering-based feature extraction was absent. Instead, properly designed loss function and network architecture played important roles. In [23]–[26], CNN was used, which was one of the most popular networks in DL research. Nonetheless, the adopted CNNs were wide and shallow, which was at variance with the "deep" feature of modern neural networks.

The recent advancement of CNN was largely driven by the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) [27]. AlexNet [28], the winner in 2012, was regarded as the breakthrough and drew the attention on CNNs. In 2014, two very deep CNNs emerged. The first one was GoogLeNet (with inception modules) [29], which adopted a sparsely connected architecture of stacking inception modules composed of filters of various sizes. In contrast, the second, VGG [30], exploited smaller filters of the same size for all the convolutional layers and increased the depth. Both very deep CNNs were able to achieve compelling performances, and however, they usually suffered the degradation problem [31] that training accuracy would saturate and then degrade as the depth increased. In 2015, ResNet [32] was proposed to address the degradation problem and won the ILSVRC-2015 with an ultradeep network of 152 layers. The fundamental idea was to let the network fit a residual mapping, instead of the original propagation, by adding skipping connections between some layers, so that in principle, a deep network would not produce higher training error than its shallower counterpart. The additional shortcut connections enabled gradients to propagate backward to earlier layers more easily and hence resulted in easier training than VGG. Later, ResNet evolved to the second version [33], where in each unit, the activation layer preceded the convolutional layer. In 2016, ResNeXt [34] was proposed, which, in the residual module, harnessed the split-transform-merge pattern akin to the inception module. These revolutionary CNN architectures have influenced many deep networks in applications beyond image classification. In particular, the state-of-the-art residual CNNs, such as ResNet and ResNeXt, can be applied to the research of defect depth estimation with ECT; however, it has not been seen in the literature.

In this article, the problem of estimating the depth of a surface defect of a metallic sheet is addressed using a new ECT device and the state-of-the-art DL techniques. The main contributions are threefold. First, a portable multifunctional ECT device is introduced, which integrates a field-programmable gate array (FPGA), an ARM processor, and the Windows 10 operating system. Second, the defect depth estimation problem is formulated as a time series classification problem, and a dataset using the ECT device is constructed and made openly available. We name the dataset as metal defects of different depths by ECT (MDDECT) and aim to initiate a data-sharing campaign. It can serve as a testbed and would encourage advancing the research of ECT in light of modern DL techniques. Finally, the state-of-the-art residual CNNs are applied for the first time, to our knowledge, in the research of ECT. An accuracy rate of 93.58% is achieved using a 38-layer 1-D ResNeXt for defects with a depth resolution of 0.1 mm for a stainless steel sheet.

The remainder of this article is organized as follows. Section II presents the hardware design of the integrated ECT device. The architectures of the 1-D residual CNNs are explained in detail in Section III. Section IV introduces the procedures of data collection and the details of the MDDECT dataset. The configurations of hyperparameters and training process are demonstrated, along with the results of different CNNs and discussions in Section V. Section VI concludes the work and suggests further research directions.

II. Hardware Design of ECT Device

The architecture of the ECT device is shown in Fig. 2. The system mainly consists of four components, which are a replaceable coil probe sensor, an SoC composed of an FPGA and ARM processer, front-end circuits, and a host PC.

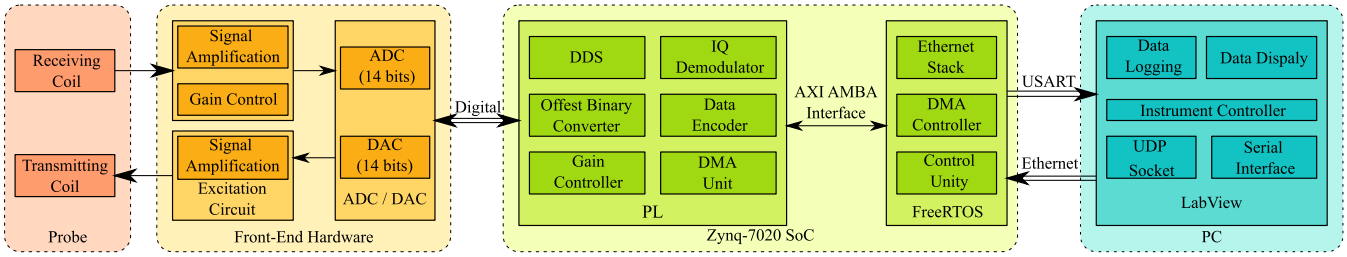

Fig. 2. Block diagram of the architecture of the ECT device. The system is mainly composed of four modules, including a replaceable coil probe sensor, a system on chip (SoC) of an FPGA and ARM processor, front-end circuits, and a host PC running Windows 10 system.

Zynq-7020 SoC is the cornerstone of the system, which integrates an ARM dual Cortex-A9 processor and a Xilinx 7-series FPGA. This module is responsible for generating excitation signals, implementing in-phase and quadrature (I/Q) demodulation and transferring data between the module and the host PC. The system is capable of providing a multi-frequency excitation signal, and the received signal can be demodulated at each frequency simultaneously. The front-end circuits consist of analog-to-digital (ADC)/digital-to-analog (DAC), signal amplification, and gain control modules. During the measurement process, the host PC displays the received signals from the coil probe sensor and allows users to operate the instrument and save data through a LabVIEW-based GUI.

A parallel digital interface is exploited to connect the front-end hardware and the SoC via an FPGA Mezzanine Card (FMC) connector. The Zynq-7020 SoC is connected to the host PC via Ethernet, and user datagram protocol (UDP) is used as the transport protocol. Universal asynchronous receiver–transmitter (UART) serial communication is available for debugging purpose. The system takes advantage of the Ethernet communication, which provides a fast and robust data transmission. The data transmission speed can reach up to 10 000 samples/s. With respect to the coil probe sensor, a receiving coil is connected to the signal amplification circuit, and a transmitting coil is connected to the excitation circuit. A LabVIEW-based data acquisition and signal processing application runs on the host PC. Such configurations as amplification gain, frequency, and sampling rate are easily adjustable from the PC interface, which provides good compatibility with a variety of sensor designs and applications.

Fig. 3 shows a scene of a human operator holding a probe sensor and scanning a stainless steel sheet. The received signals and statistic information are displayed on the screen of the ECT device in real time. The geometry dimensions of the steel sheet, which is used to collect data from, are shown in Fig. 4. The schematic of the probe sensor and the geometry dimensions are shown in Fig. 5

$$\delta = \frac{1}{\sqrt{\sigma \pi f \mu}}. \tag{1}$$

The penetration depth δ of the electromagnetic field is limited by the skin effect, which is shown in (1), where f is the frequency of the excitation current and μ and σ are the permeability and conductivity of the specimen, respectively. Therefore, the operating frequency of the device depends on the electromagnetic properties of the specimen and the depth of the defect. Generally, low- and high-frequency excitations would be applied for detecting subsurface and on-surface defects, respectively. The typical excitation frequency of our device ranges from 10 to 100 kHz. In the experiment, the relative permeability of the stainless steel plate is 1 and the conductivity is 1.46×10^6 S/m. Therefore, the electromagnetic field generated by the device can penetrate up to 4.17 mm beneath the surface of the specimen. During the experiments, we set the excitation frequency to 20 kHz, which gives 2.94 mm of penetration depth that covers the depth of defects on the specimen. The excitation current amplitude is 200 mA.

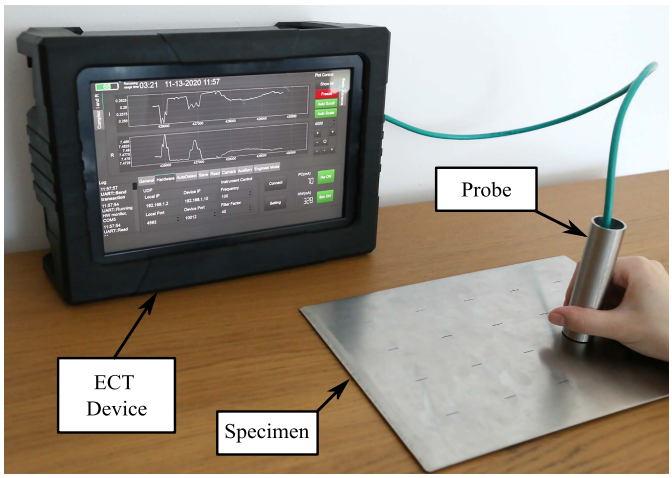

Fig. 3. Scene of a human operator scanning a steel sheet using a hand-hold probe sensor with the ECT device.

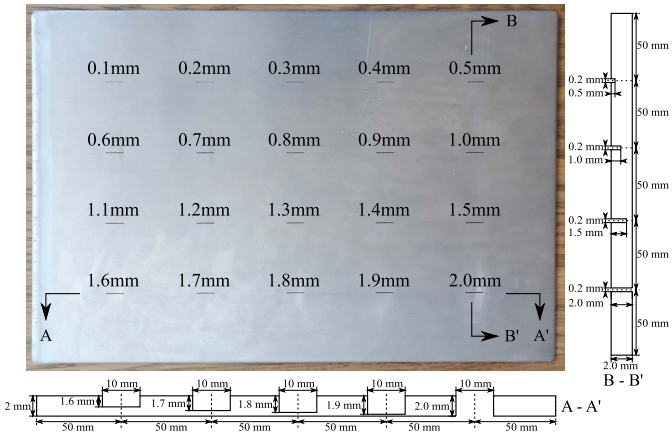

Fig. 4. Illustration of the geometry dimensions of a stainless steel sheet that is used as a specimen to collect data from.

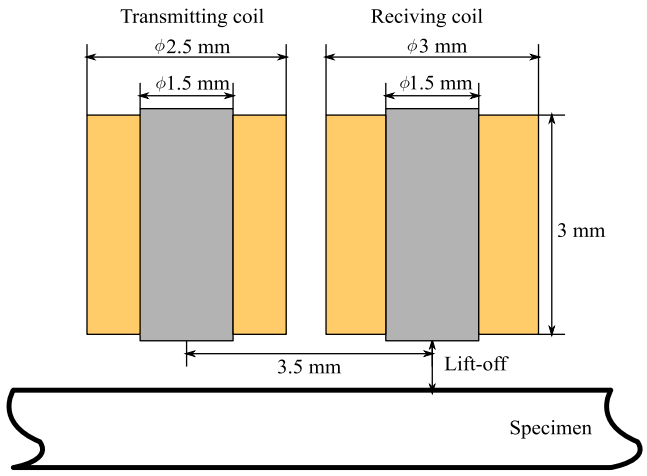

Fig. 5. Schematic of a probe sensor, which consists of two cylindrical ferrite-cored coils.

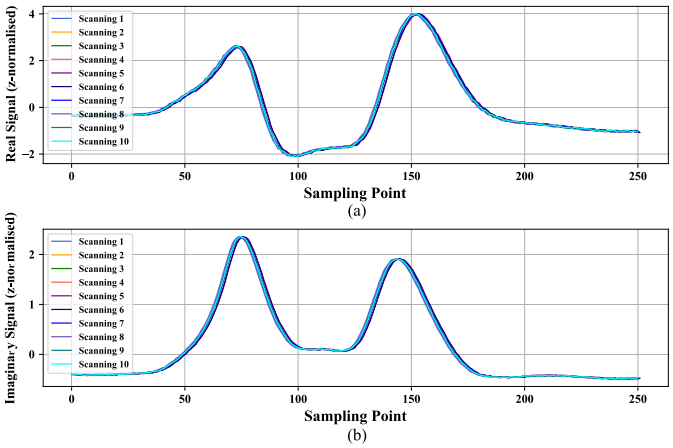

Fig. 6. Signals of ten scans using a stepper motor for the 2.0-mm-deep defect under the same scanning conditions. (a) and (b) I/Q signals, respectively.

The performance of the device is evaluated by its repeatability of signal and signal-to-noise ratio (SNR). In terms of repeatability, the 2.0-mm-deep defect was scanned ten times using a stepper motor under the same scanning conditions (angle, speed, liftoff, starting point and endpoint, and so on). The resultant signals are plotted in Fig. 6. It can be seen that the signals almost coincide with each other. Furthermore, the correlation value is larger than 0.99 between the signals of any two of the ten scans, which indicates a high repeatability of the device. In addition, the SNR of the received signal is above 80 dB on average.

III. Architecture of 1-D Residual CNN

A residual CNN, e.g., ResNet, is constructed by stacking "residual units," which learns a residual function $\mathcal{R}(x) := \mathcal{F}(x) - x$ where $\mathcal{F}(x)$ is the original underlying mapping [32]. Formally, a residual unit conducts calculations as expressed in (2), in which x_l and y_l are the input and output of the lth residual unit, respectively, and ϱ is an activation function. The block $\mathcal{R}$ takes x_l as input and performs transformations with weights W_l. In the original version of ResNet, the activation function ϱ is a rectified linear unit (ReLU) function. The block $\mathcal{R}$ is chosen from either a stack of two convolution units or a "bottleneck" unit. A convolution unit is a convolution layer followed by a batch normalization (BN) layer and a ReLU layer. A bottleneck unit comprises three convolution units, with the first and last convolution layers being 1×1 convolutions, which are used to reduce dimensionality hence computational complexity

$$y_l = \varrho(x_l + \mathcal{R}(x_l, W_l)) \tag{2}$$

$$y_l = x_l + \mathcal{R}(x_l, W_l) \tag{3}$$

$$y_l = x_l + \sum_{i=1}^{C} \mathcal{T}_i(x_l, W_{\text{li}}). \tag{4}$$

In the second version of ResNet, the residual unit performs calculations as shown in (3), where the activation is an identity function [33]. However, the convolution unit in the block $\mathcal{R}$ is preactivated, i.e., the BN and ReLU layers precede the convolution layer. It was verified in [33] that (3) enabled gradients to propagate to any layers more easily than (2). In ResNeXt [34], the residual unit performs calculations as shown in (4), where the block $\mathcal{R}$ is an aggregation of a number of transformations $\mathcal{T}_i$, and the number C is the cardinality. In practice, the split-transform-merge block $\mathcal{R}$ is usually implemented using an equivalent grouped convolution, where the number of groups equals the cardinality C. It is noted that here, we assume that ResNeXt inherits from the second version of ResNet. In addition, the input x_l and output y_l share the same dimension in (2)–(4) so as to illustrate the idea. When the dimensions are different, a convolution layer will replace the identity connection to match the dimensions.

The convention of naming a network is followed in this article by appending the version and depth to the type of network. However, because the convolution layers used are 1-D instead of 2-D, we append "1-D" to the name in order to differentiate the networks from the original 2-D ones. Usually, residual CNNs comprise multiple stages, each one of which has one or a stack of multiple residual units. In this article, we unify the number of stages to four. The first residual unit of each stage doubles the channel dimension while halving the temporal dimension. In order to clarify details, Fig. 7 shows the architectures of ResNet1Dv1-14, ResNet1Dv2-14, and ResNeXt1D-14, which serve as the baseline networks for the defect depth classification task. They all have the same depth of 14 essential layers.

IV. MDDECT Dataset

The success of DL in an application relies heavily on the effort of constructing large and well-labeled datasets. For instance, ImageNet contains 14 million high-quality images in 22 000 visual categories [27]. However, it is difficult in principle to develop a universal dataset like ImageNet for defect depth estimation by ECT because the impedance signal captured by an ECT sensor is determined by a plurality of factors.

First of all, a variety of ECT device designs exist, where different excitation strategies are applied and different types of sensors are exploited. In addition, as per applications, the geometry and electromagnetic properties of specimens are

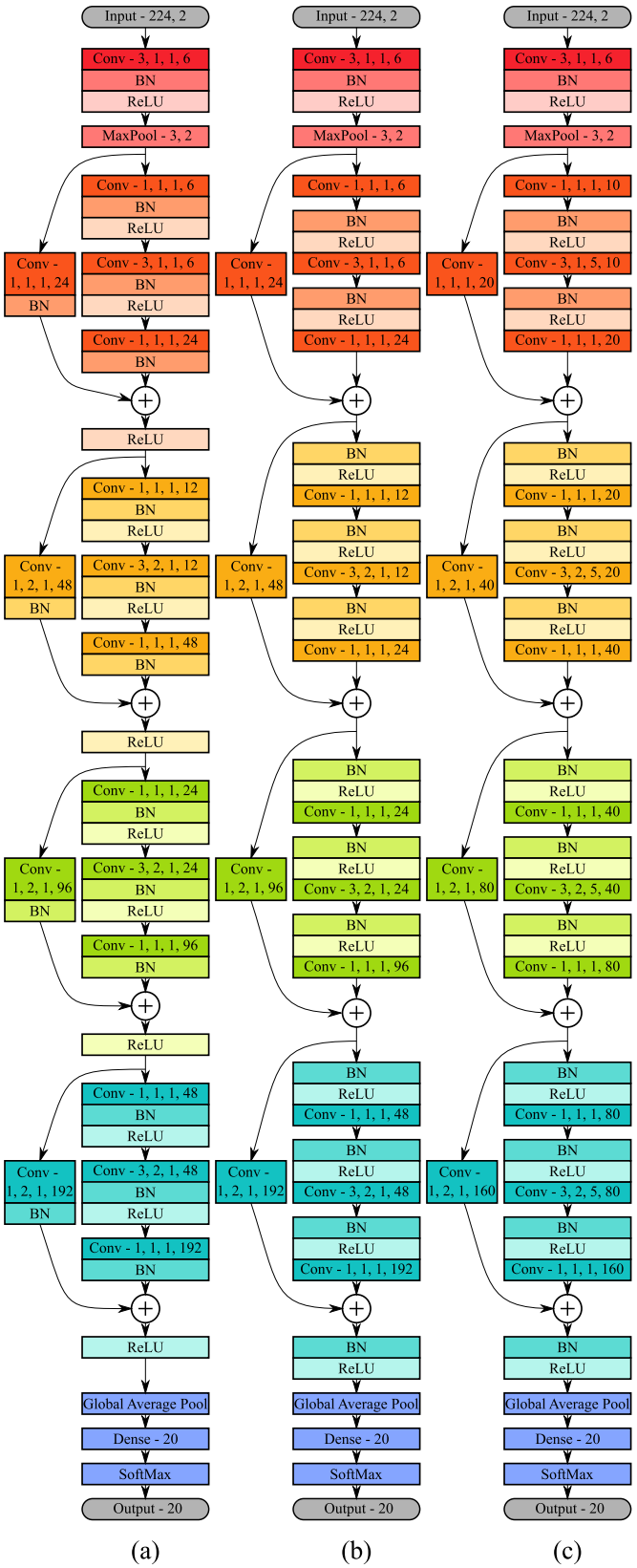

Fig. 7. Detailed architectures of three 1-D residual networks. Four stages in each network are marked in different colors. The input tensor has 224 sampling points and two channels. In terms of the convolution layer, the numbers in the block represent the kernel size, strides, number of groups, and number of filters. In terms of the max-pooling layer, the numbers represent the kernel size and strides. The output tensor has 20 labels. (a), (b) and (c) ResNet1Dv1-14, ResNet1Dv2-14, and ResNeXt1D-14, respectively.

usually distinctive. A widely accepted standard specimen with standard defects is lacking in the ECT research community. An ECT defect depth estimation dataset often corresponds to a specific device and specimen. A large open dataset would help facilitate advancing the ECT techniques.

In this article, the MDDECT dataset was constructed using the integrated ECT device described in Section II, which installed a probe sensor as shown in Fig. 5. The filtering factor and signal data rate were configured as 4000 and 2500 samples/s, respectively. A stainless steel sheet with 20 machine-fabricated slots on the surface was used as the specimen to scan, whose detailed geometry dimensions are shown in Fig. 4. The defects were surface opening cracks and shared the same length and width of 10 and 0.2 mm, respectively. The depth of the defects started from 0.1 mm and incremented by 0.1 mm to the largest 2.0 mm. The defect of 2.0 mm in depth was a through crack, as the thickness of the sheet was 2.0 mm.

Although the MDDECT dataset was specialized to the ECT device and the specimen, many other practical variances, which would nontrivially affect the performance of an ECT system, were considered. First, 30 volunteers, who had no experience operating an ECT device, were invited to scan the defects, hence introducing a great variety of uncertainties, as composed to some research, e.g., [20]–[23], where an automatically controlled movement was harnessed to scan defects. Second, liftoff signals were deliberately collected and labeled so that the classifier should be able to differentiate liftoffs and defects. As shown in Fig. 8(c), liftoff signals were generated by randomly tapping the probe to a defect-free area on the surface of the specimen. The liftoff distances were controlled to be under 3 mm above the surface. In addition, normal signals of defect-free areas were also captured and labeled, and the capturing process is shown in Fig. 8(d). Third, eight different scanning angles between the long-edge line of a defect and the line crossing the axial centers of the two coils were determined, as shown in Fig. 8(a) and (b). Also, the volunteers scanned across a defect along an angle in two directions, back and forth. Finally, when scanning a defect, the volunteers were asked to try to maintain a constant liftoff distance of 0.5 mm, a constant moving speed of 60 mm/s, and vertical to the surface. However, variances were inevitable and represented more practical testing scenarios, hence making the MDDECT challenging.

After a preliminary test, it was found that the signals of the defects of 0.1 and 0.2 mm were under the noise floor, and hence, the data from these two defects were excluded from the dataset. As a result, the total number of classes was 20, including 18 classes of defects, liftoff class, and normal signal class. Each volunteer repeated the same scanning five times. All scans were divided into scan segments, each of time window 0.5 s, which gave rise to the temporal dimension of 1250. Ultimately, the dataset tensor had dimensions of (30, 8, 2, 5, 20, 1250, and 2), each one of which represented the number of volunteers, scanning angles, directions, repeats of each scanning, classes, temporal points, and channels, respectively. The last dimension corresponded to the I/Q channels. In total, the MDDECT comprehended 48 000 scan segments

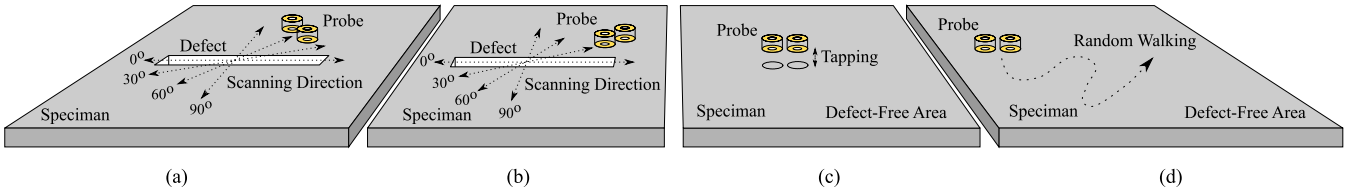

Fig. 8. Illustrations of data collecting processes. (a) and (b) Probe scanning across a defect in eight different angles and two directions (back and forth). Notice that the line passing the axial centers of the two coils is either tangential or normal to the scanning direction. (c) Probe tapping on the surface of a defect-free area in order to generate liftoff signals. (d) Probe moving in a random trajectory on the surface of a defect-free area in order to generate normal signals.

(or scans for simplicity) in 20 classes. In terms of the split of training and test sets, volunteers and the data scanned by them were randomly selected. As a result, the dimensions for training and test sets are (43 200, 1250, and 2) and (4800, 1250, and 2), respectively,[1] constituting 90% and 10% of the total samples. Hence, the test set consists of three randomly selected volunteers. The MDDECT dataset is available on Kaggle.[2]

V. Experiment Details and Results

Before training a network, we needed to extract and set aside a validation set from the training set to determine hyperparameters and select DL models. The test set must not be used in the training phase and should only serve to produce the final claim on accuracy. As the training set contains data from 27 volunteers, we randomly selected 3 from the 27 volunteers and used their corresponding data as a validation set. In addition, the temporal dimension was decimated from 1250 to 250 by a factor of 5, in order to enable faster training. As a result, the dimensions of the training, validation, and test sets were (38 400, 250, and 2), (4800, 250, 2), and (4800, 250, 2), respectively. The ratio of the number of samples among them was 8:1:1.

A. Normalize and Augment Data

The final training data were applied with a channelwise normalization according to (5), where x_i is the flattened tensor when the channel dimension equals i and μ_i and σ_i are the mean and standard deviation of x_i, respectively. This normalization is termed as z-normalization [35], which nullifies the mean and standardizes the variance of the data in terms of each channel. The resultant training data were then used to train a network. In addition, the calculated μ_i and σ_i from the training data were used to normalize the validation and test data too. In order to appreciate the data in general before training, the z-normalized validation data are plotted all together in a complex plane in Fig. 9, from which it can be seen that the liftoff signals are larger in magnitude and also different in phase compared to the defect signals. However, most signals overlap severely with each other, indicating the difficulties to classify these signals. To be more clear, some typical signals for the defects are plotted in Fig. 10

$$\frac{x_i - \mu_i}{\sigma_i}, i = 1, 2. \tag{5}$$

[1]The numbers 43 200 and 4 800 are calculated from $27 \times 8 \times 2 \times 5 \times 20$ and $3 \times 8 \times 2 \times 5 \times 20$, respectively.

[2]https://www.kaggle.com/mchikyt3/mddect

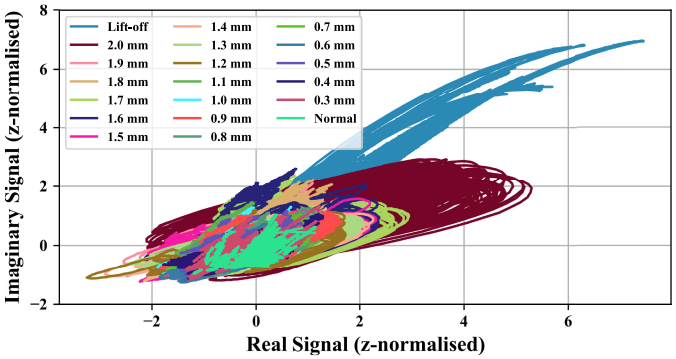

Fig. 9. Plots of the z-normalized validation data all together in a complex plane. The mean and standard deviation used in the normalization were calculated from the training data. The x- and y-axes represent the I/Q channels, respectively.

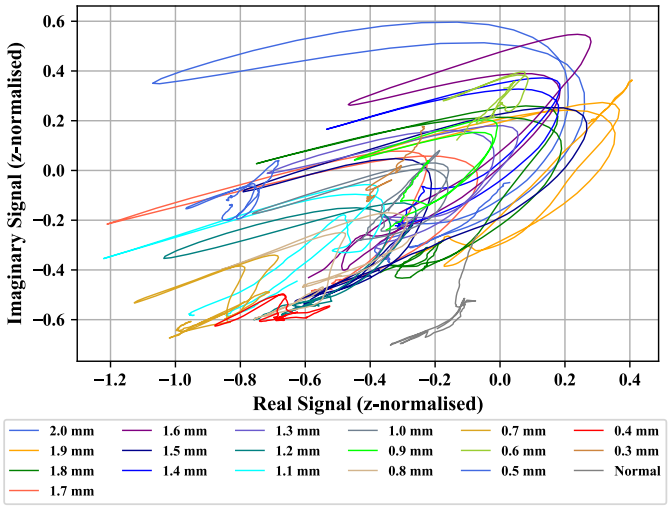

Fig. 10. Plots of typical signals for the defects in a complex plane. The signals correspond to the 90° scanning angle in Fig. 8(a).

After normalization, the data were then configured for train- and test-time augmentations. In terms of train-time augmentation, every training sample was cropped randomly for every epoch in order to introduce certain variances to the training data on the fly. Concretely, a segment of dimensions (224, 2) was cropped randomly under a uniform distribution from the original training sample of dimensions (250, 2). As a result, the input dimensions to a network were (224, 2), where the batch dimension was not shown. In terms of test-time augmentation, ten-crop test was conducted to the validation and test samples, i.e., the final classification result of a sample

TABLE I
ARCHITECTURE DETAILS OF RESIDUAL CNNS

Stage	Output Size	Network Architecture							
		ResNet1Dv1-26	ResNet1Dv2-26	ResNeXt1D-26	ResNet1Dv1-14 (Wider)	ResNet1Dv2-14 (Wider)	ResNeXt1D-14 (Wider 1)	ResNeXt1D-14 (Wider 2)	ResNeXt1D-38
	224	conv - 3, 1, 1, 6	conv - 3, 1, 1, 6	conv - 3, 1, 1, 6	conv - 3, 1, 1, 8	conv - 3, 1, 1, 8	conv - 3, 1, 1, 8	conv - 3, 1, 1, 8	conv - 3, 1, 1, 6
	112	max pool - 3, 2							
1	112	$\begin{bmatrix} 1,6 \\ 3,6 \\ 1,24 \end{bmatrix} \times 2$	$\begin{bmatrix} 1,6 \\ 3,6 \\ 1,24 \end{bmatrix} \times 2$	$\begin{bmatrix} 1,10 \\ 3,10,C=5 \\ 1,20 \end{bmatrix} \times 2$	$\begin{bmatrix} 1,8 \\ 3,8 \\ 1,32 \end{bmatrix} \times 1$	$\begin{bmatrix} 1,8 \\ 3,8 \\ 1,32 \end{bmatrix} \times 1$	$\begin{bmatrix} 1,14 \\ 3,14,C=7 \\ 1,28 \end{bmatrix} \times 1$	$\begin{bmatrix} 1,15 \\ 3,15,C=5 \\ 1,30 \end{bmatrix} \times 1$	$\begin{bmatrix} 1,10 \\ 3,10,C=5 \\ 1,20 \end{bmatrix} \times 3$
2	56	$\begin{bmatrix} 1,12 \\ 3,12 \\ 1,48 \end{bmatrix} \times 2$	$\begin{bmatrix} 1,12 \\ 3,12 \\ 1,48 \end{bmatrix} \times 2$	$\begin{bmatrix} 1,20 \\ 3,20,C=5 \\ 1,40 \end{bmatrix} \times 2$	$\begin{bmatrix} 1,16 \\ 3,16 \\ 1,64 \end{bmatrix} \times 1$	$\begin{bmatrix} 1,16 \\ 3,16 \\ 1,64 \end{bmatrix} \times 1$	$\begin{bmatrix} 1,28 \\ 3,28,C=7 \\ 1,56 \end{bmatrix} \times 1$	$\begin{bmatrix} 1,30 \\ 3,30,C=5 \\ 1,60 \end{bmatrix} \times 1$	$\begin{bmatrix} 1,20 \\ 3,20,C=5 \\ 1,40 \end{bmatrix} \times 3$
3	28	$\begin{bmatrix} 1,24 \\ 3,24 \\ 1,96 \end{bmatrix} \times 2$	$\begin{bmatrix} 1,24 \\ 3,24 \\ 1,96 \end{bmatrix} \times 2$	$\begin{bmatrix} 1,40 \\ 3,40,C=5 \\ 1,80 \end{bmatrix} \times 2$	$\begin{bmatrix} 1,32 \\ 3,32 \\ 1,128 \end{bmatrix} \times 1$	$\begin{bmatrix} 1,32 \\ 3,32 \\ 1,128 \end{bmatrix} \times 1$	$\begin{bmatrix} 1,80 \\ 3,80,C=7 \\ 1,160 \end{bmatrix} \times 1$	$\begin{bmatrix} 1,60 \\ 3,60,C=5 \\ 1,120 \end{bmatrix} \times 1$	$\begin{bmatrix} 1,40 \\ 3,40,C=5 \\ 1,80 \end{bmatrix} \times 3$
4	14	$\begin{bmatrix} 1,48 \\ 3,48 \\ 1,192 \end{bmatrix} \times 2$	$\begin{bmatrix} 1,48 \\ 3,48 \\ 1,192 \end{bmatrix} \times 2$	$\begin{bmatrix} 1,80 \\ 3,80,C=5 \\ 1,160 \end{bmatrix} \times 2$	$\begin{bmatrix} 1,64 \\ 3,64 \\ 1,256 \end{bmatrix} \times 1$	$\begin{bmatrix} 1,64 \\ 3,64 \\ 1,256 \end{bmatrix} \times 1$	$\begin{bmatrix} 1,160 \\ 3,160,C=7 \\ 1,320 \end{bmatrix} \times 1$	$\begin{bmatrix} 1,120 \\ 3,120,C=5 \\ 1,240 \end{bmatrix} \times 1$	$\begin{bmatrix} 1,80 \\ 3,80,C=5 \\ 1,160 \end{bmatrix} \times 3$
	1	global average pool, 20-d fc, softmax							
# Trainable Parameters		9.37×10^4	9.30×10^4	9.38×10^4	1.01×10^5	1.00×10^5	9.77×10^4	1.14×10^5	1.35×10^6
FLOPs		3.70×10^6	3.69×10^6	3.84×10^6	4.11×10^6	4.09×10^6	4.25×10^6	4.99×10^6	5.42×10^6

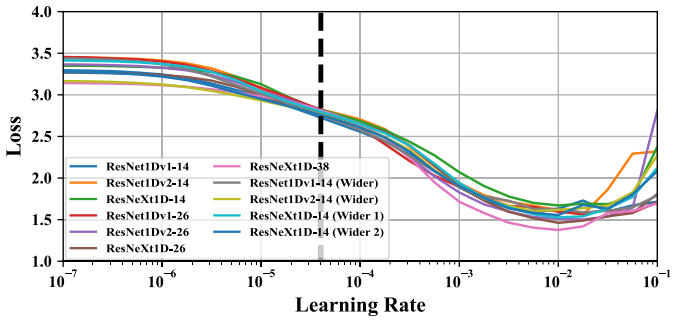

Fig. 11. Plots of loss versus learning rate. The best initial learning rate appears to be the same for all the evaluated networks, which is about 4.0×10^{-5} and marked by a dotted vertical line in the figure.

was determined by averaging the outputs of a network for ten random crops from the sample.

B. Determine Network Architectures and Hyperparameters

As discussed in Section III, the networks exploited in this article were 1-D variants of the residual CNNs. In addition, the CNNs were fixed to have four stages, and hence, the minimal depth was 14 for each network, where only one residual module existed in each stage. Fig. 7 shows the architectures of ResNet1Dv1-14, ResNet1Dv2-14, and ResNeXt1D-14, which served as the baseline networks. The width of them, i.e., the number of filters in the first convolution layer, was set to 6. They all had a similar level of trainable parameters and floating-point operations (FLOPs).

Based on the three baseline networks, we experimented on the depth dimension and doubled the number of residual modules in each stage of the networks, which gave rise to three deeper networks: ResNet1Dv1-26, ResNet1Dv2-26, and ResNeXt1D-26. In addition, we also attempted to expand the width dimension from the baseline networks while maintaining the number of trainable parameters and FLOPs similar to the 26-layer ones, which gave rise to four wider networks. The details of the architectures of these seven new networks are listed in Table I. As they had a similar computation complexity, we would be able to compare their performances and see whether depth or width was more effective under the scope of this article. Finally, we evaluated our deepest network ResNeXt1D-38, where in each stage, there were three residual modules. The details of the architecture of ResNeXt1D-38 are listed in the last column in Table I.

Learning rate is an important hyperparameter affecting the training and performance of a network. In order to find the best initial learning rate, we performed the strategy from [36], where the training started with a small learning rate and increased it epoch by epoch in a geometric progression. After a few epochs, the loss versus learning rate plot could be drawn, and the best initial learning rate located at the point where the loss decreased most rapidly. This strategy was applied to all the networks we evaluated, and the resultant plots of loss versus learning rate are shown in Fig. 11. It can be seen that all the networks appear to share the same best learning rate 4.0×10^{-5}, at which the losses descend rapidly. It is noted that this way of finding the best initial learning rate may not an exact solution. Notice that the x-axis in the figure is in a logarithmic scale.

All the networks were applied with the same following training configurations and hyperparameters. The Adam optimizer was exploited with default values of parameters recommended in [37], and the mini-batch size was set to 128. The weight decay was set to 0.01. The training data were shuffled for every epoch, and each network was trained for 10 000 epochs. The learning rate was initialized to 4.0×10^{-5} and decreased to 4.0×10^{-6} at epoch 5000 and finally to 4.0×10^{-7} at epoch 7500.

C. Train Networks and Analysis Results

The training was coded using the Tensorflow library and conducted on an Nvidia RTX 2080Ti GPU. In total, the training of the networks took about five days to finish. The training and validation accuracies and losses are shown in Fig. 12 for three baseline networks: ResNet1Dv1-14, ResNet1Dv2-14, and ResNeXt1D-14, four deeper networks: ResNet1Dv1-26, ResNet1Dv2-26, ResNeXt1D-26, and

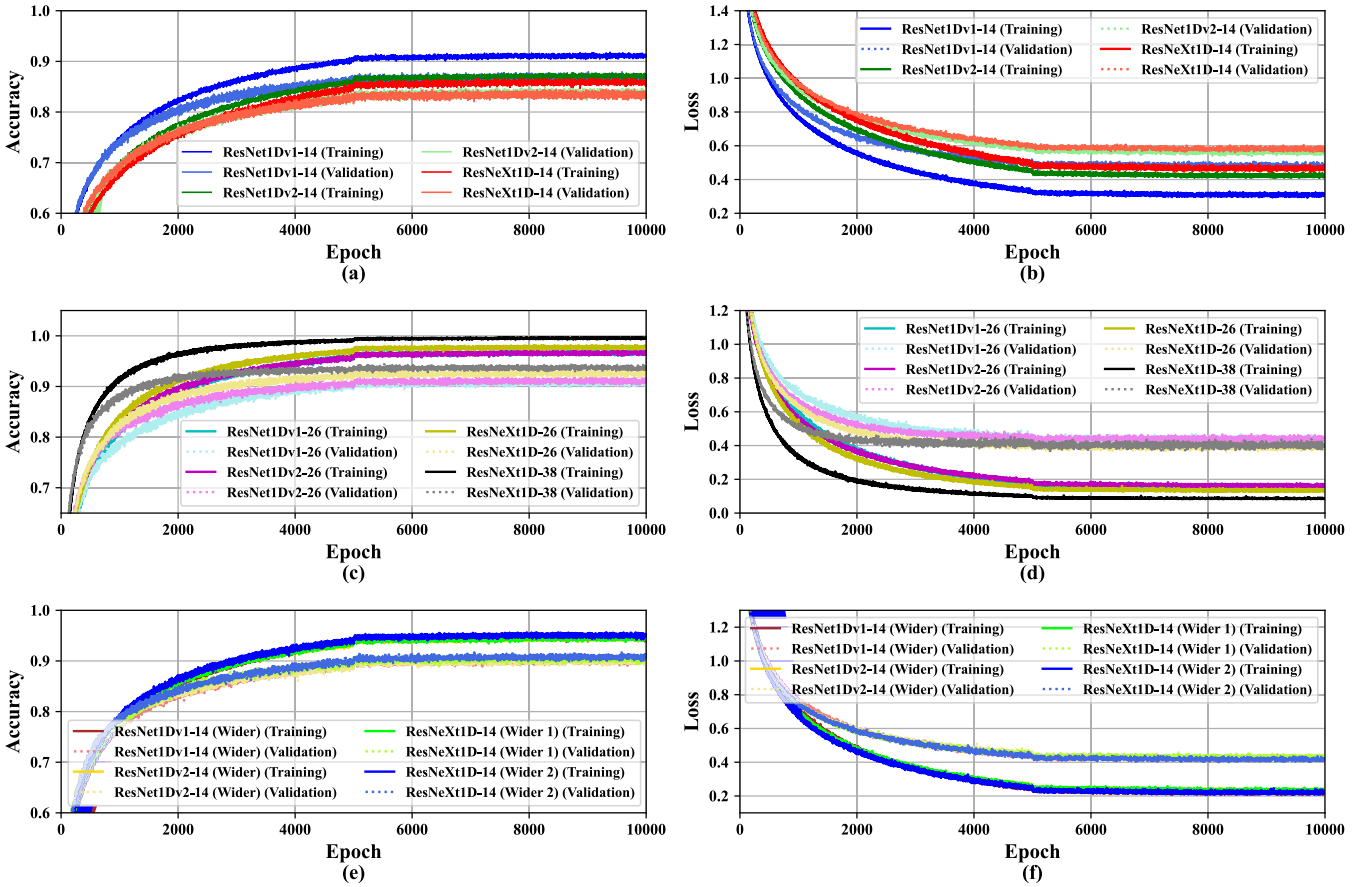

Fig. 12. Training processes of eleven networks. (a), (c) and (e) Training and validation accuracies for three baseline networks, four deeper networks, and four wider networks, respectively. (b), (d) and (f) Corresponding losses. Solid and dashed lines represent the training and validation accuracies/losses, respectively. Results of the same network use similar colors.

TABLE II
CLASSIFICATION RESULTS

Network Architecture	Accuracy	
	Top-1	± 0.1mm
ResNet1Dv1-14	87.88%	95.14%
ResNet1Dv2-14	83.21%	93.01%
ResNeXt1D-14	86.65%	94.51%
ResNet1Dv1-26	92.50%	96.90%
ResNet1Dv2-26	91.85%	96.06%
ResNeXt1D-26	93.15%	96.78%
ResNet1Dv1-14 (Wider)	90.42%	95.88%
ResNet1Dv2-14 (Wider)	89.93%	95.65%
ResNeXt1D-14 (Wider 1)	91.31%	96.23%
ResNeXt1D-14 (Wider 2)	89.38%	95.00%
ResNeXt1D-38	**93.58%**	**97.20%**

ResNeXt1D-38, and three wider networks: ResNet1Dv1-14, ResNet1Dv2-14, and ResNeXt1D-14 with respect to training epoch, which are also available as a TensorBoard experiment.[3]

Three main observations can be taken from the training processes. First, the three baseline networks all suffer the underfitting problem as per Fig. 12(a), that is, they converge to large training errors between 15% and 10%, which implies that their representation powers are

[3]https://tensorboard.dev/experiment/hNURDaEzRr2GQCZeDyqBHw/

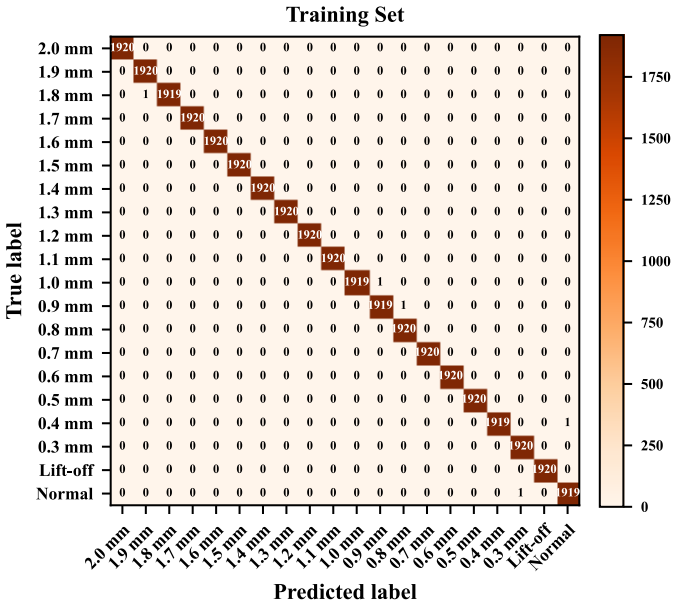

Fig. 13. Confusion matrix of the trained ResNeXt1D-38 for the training set. Every entry in the matrices represents the number of samples that are classified to a specific class.

insufficient for the MDDECT dataset. Such suggestion is verified by the results of larger networks in Fig. 12(c) and (e), where the training errors of ResNet1Dv1-26, ResNet1Dv2-26,

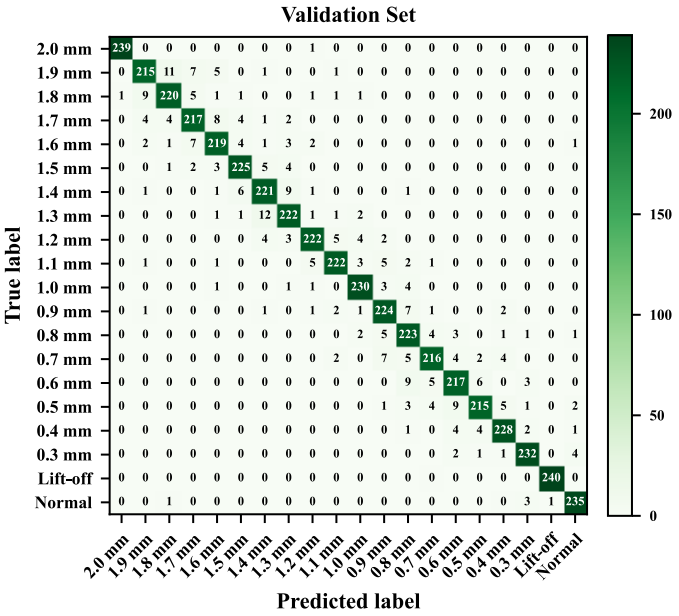

Fig. 14. Confusion matrix of the trained ResNeXt1D-38 for the validation set. Every entry in the matrices represents the number of samples that are classified to a specific class.

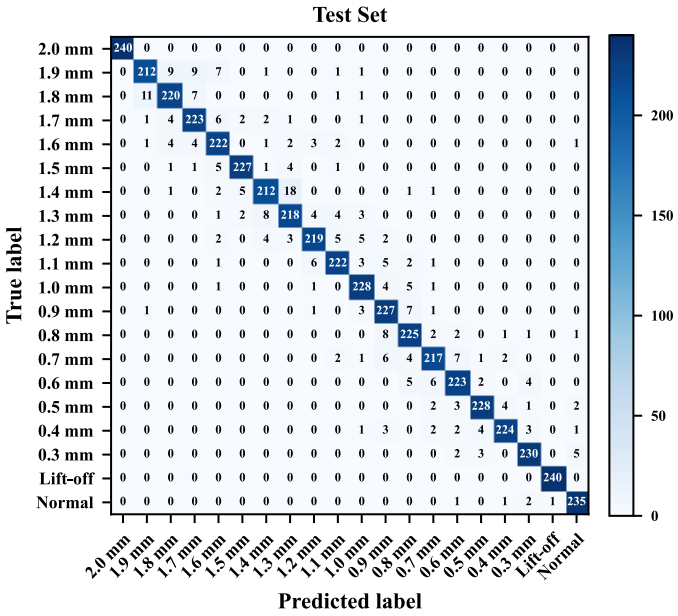

Fig. 15. Confusion matrix of the trained ResNeXt1D-38 for the test set. Every entry in the matrices represents the number of samples that are classified to a specific class.

and ResNeXt1D-26 are well below 5%, and the training errors of the wider ResNet1Dv1-14, ResNet1Dv2-14, and ResNeXt1D-14 are slightly higher than 5%. On the other hand, the validation errors of the larger networks are around 10%. However, it is difficult to conclude that the larger networks are overfitted to the training data, only based on the about 5% difference between the training and validation errors. Second, the deeper networks, in general, have achieved higher accuracies than the wider ones that share a similar level of trainable parameters and FLOPs, as can be seen by comparing Fig. 12(c) and (e). Third, different versions and types of network lead to marginal differences in terms of both the training and validation accuracies. In Fig. 12(c), the discrepancies of the training and validation accuracies of ResNet1Dv1-26, ResNet1Dv2-26, and ResNeXt1D-26 are within 3%. The discrepancies in Fig. 12(e) are even smaller. However, ResNeXt1D-26 has achieved the best performance among the larger networks of similar levels of complexity. As a consequence, we further trained a ResNeXt1D-38, the deepest network we evaluated, so as to test the limit of performance. Its training and validation accuracies are plotted in Fig. 12(c).

In order to claim the final accuracies, the best trained model of a network was selected as the one that achieved the highest validation accuracy during training for that network. Next, the test set data were fed to the chosen model to produce the final claimed accuracy of the network. The final top-1 accuracies of each network are listed in Table II. It can be seen that ResNeXt1D-38 has achieved the highest accuracy of 93.58%, while the second best accuracy is 93.15%, achieved by ResNeXt1D-26. The fact that the additional 12 layers give rise to an improvement of only 0.43% implies that simply increasing the depth may not be able to push the boundary of performance, and overfitting may occur. Moreover, if the accuracy metric is relaxed to tolerate an error of ± 0.1 mm, the "± 0.1-mm accuracies" are also listed in the same table, where ResNeXt1D-38 also wins with 97.20%. If we examine closer on the results of ResNeXt1D-38, its confusion matrices for the training, validation, and test sets are shown in Figs. 13–15, respectively. It can be seen from Fig. 13 that misclassified training samples are sparse, and the training error is nearly zero. Moreover, according to Figs. 14 and 15, mistaken samples of the validation and test set tend to be classified into the adjacent classes of their ground-truth classes, which explains the much improved 97.20% accuracy with ± 0.1-mm tolerances. In addition, it appears that shallower defects are not more difficult to classify than deeper ones. In fact, the defects in the middle range, i.e., from 0.7 to 1.7 mm, have larger errors than those of the remaining defects. Finally, the liftoff samples are all correctly detected. Being immune to liftoff signals is an important and desirable feature for an ECT defect depth estimation method.

In order to further understand the misclassified samples, the test set data labeled by the trained ResNeXt1D-38 are plotted in Fig. 16, from which it can be seen that, in general, the wrongly labeled samples overlap in great deal with the correct samples. We argue that it is almost impossible to estimate by human the depth of a defect in the resolution of 0.1 mm, while ResNeXt1D-38 has achieved a 93.58% accuracy.

Furthermore, the misclassified samples can be divided into two categories. The first type is due to the noisy label in the test set. For instance, a normal sample is predicted as liftoff, whose signal is plotted in Fig. 17 together with those of true normal and liftoff samples. The green trace, a normal signal, clearly exhibits a pattern similar to those of the blue and amber traces, which are liftoff signals. This pattern is very different from those of the red and purple traces, which are normal signals. Therefore, we suspect that the volunteer had a liftoff movement when a normal sample was supposed to be captured. However, this sample has been labeled as normal

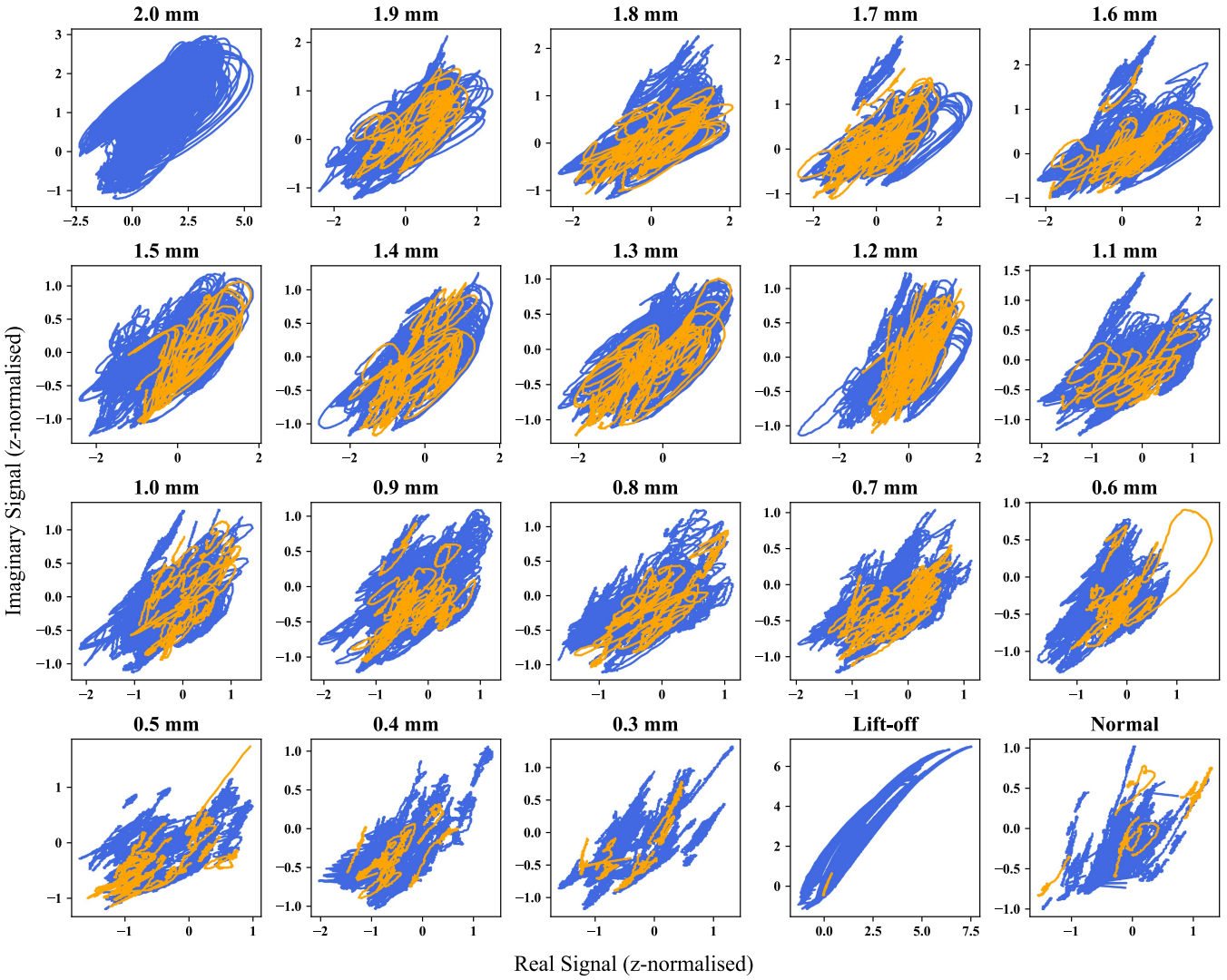

Fig. 16. Plots of the samples in the test set that are grouped into 20 subplots based on the ground-truth labels. The data are z-normalized with the means and standard deviations of the training set data. Blue and yellow traces represent the correctly and wrongly classified samples of the corresponding class, respectively. The predicted labels of samples are from the trained ResNeXt1D-38 model.

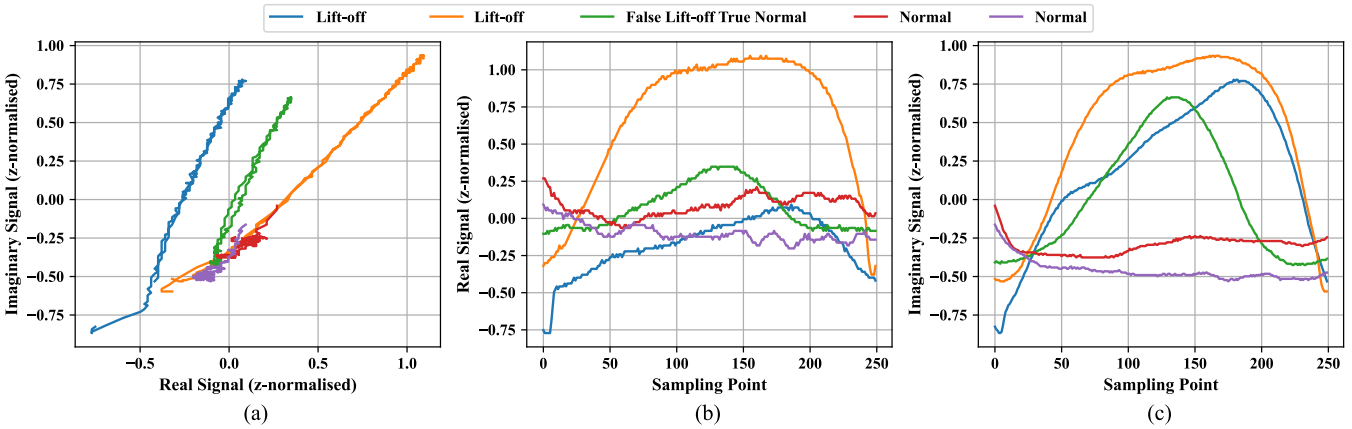

Fig. 17. Signals of true liftoff, true normal, and false liftoff (true "normal") samples in the test set. (a) Signals in a complex plane. (b) and (c) I/Q signals, respectively.

in the dataset, which is a noisy label. Some other noisily labeled samples are circled in the confusion matrix of test set in Fig. 15. Some surprising ones include a 1.6-mm defect that is misclassified as normal, a 1.9-mm defect as 1.0 mm, and a 0.9-mm defect as 1.9 mm. However, compared to the total number of samples, such noisily labeled samples are sparse.

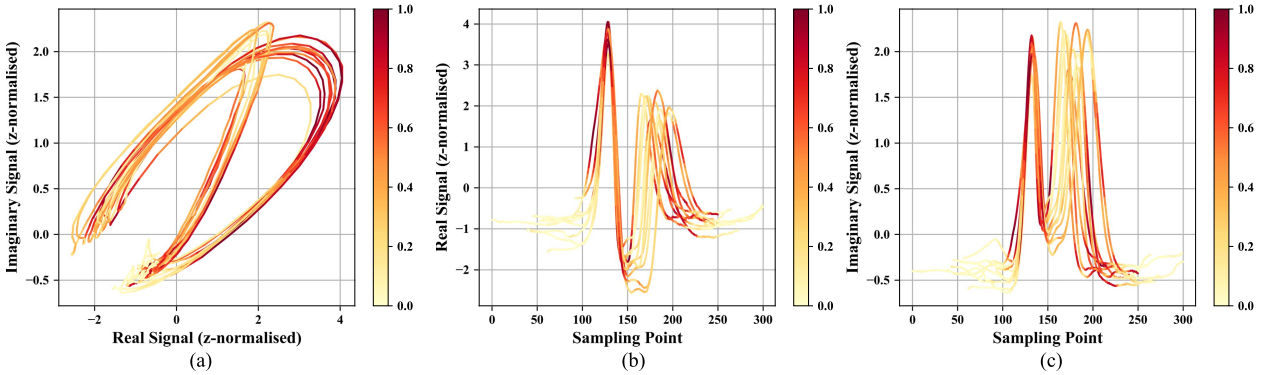

Fig. 18. Plots of the samples of the 2.0-mm defects in the test set with class activation mappings calculated based on the trained ResNeXt1D-38. The color represents the activation level of a sampling point with respect to the class of the 2.0-mm defect. (a) Sample signals in a complex plane. (b) and (c) I/Q signals with sampling point, respectively. In (b) and (c), the time series are aligned according to their first peak points in order to clarify the activation regions.

TABLE III
REGRESSION ERRORS FOR DEFECTS OF DIFFERENT DEPTHS

Depth	Δ (mm)	Depth	Δ (mm)
2.0 mm	0.000	1.0 mm	0.055
1.9 mm	0.108	0.9 mm	0.072
1.8 mm	0.074	0.8 mm	0.072
1.7 mm	0.066	0.7 mm	0.064
1.6 mm	0.130	0.6 mm	0.052
1.5 mm	0.046	0.5 mm	0.054
1.4 mm	0.074	0.4 mm	0.082
1.3 mm	0.055	0.3 mm	0.056
1.2 mm	0.063	Average	0.065
1.1 mm	0.061		

Apart from the noisily labeled samples, the majority of misclassifications are due to the high overlaps between the signals of two defects of similar depths. As shown in Fig. 15, most misclassified samples are predicted as the adjacent classes of their true classes. Without a strictly regularized scanning setup (e.g., with a stepper motor) and when scanning by human operators, the variance of signals of a defect is large, which compromises the distinction between the signals of defects of two similar depths. It is not a trivial task to nullify this type of classification error, even equipped with advanced DL algorithms

$$\Delta = \sqrt{\frac{\sum (d_{\text{predicted}} - d_{\text{true}})^2}{n}}. \tag{6}$$

The depth measurement accuracy can be evaluated based on the regression error Δ for each depth, which is calculated using (6) where n is the number of samples for a depth and $d_{\text{predicted}}$ and d_{true} are the predicted and true depths, respectively. The regression errors for different depths are summarized in Table III. It can be seen that the maximum regression error is 0.108 mm occurred in the 1.9-mm defect class. The average regression error is 0.065 mm.

A class activation mapping (CAM) can be calculated as per [38] to examine the contributions of every sampling points to the classification of a time series. As an illustration, CAMs for the samples of the 2.0-mm defect in the test set were calculated based on the trained ResNeXt1D-38 and are shown in Fig. 18. It can be seen that only some regions of the time series are activated and other regions are masked out. This can help explain why ResNeXt1D-38 is able to correctly differentiate samples of different classes that have overlapping parts because they have different activation regions for their own classes. Moreover, it is worth pointing out that the times series of the same class share similar activation regions, as can be clearly seen in Fig. 18.

VI. CONCLUSION

The evaluation of the depth of a surface defect of metallic materials is a major application of ECT, where recent DL-motivated methods commence to surpass the conventional ones. However, many existing approaches have not taken full advantage of the state-of-the-art DL techniques proposed in computer vision. In this article, we aim at addressing the problem of ECT-based surface defect depth estimation by using 1-D deep residual convolutional networks. First, a highly integrated and multifunctional portable ECT device is developed based on Zynq-7020 SoC, which provides fast data acquisition and I/Q demodulation. Second, a dataset, termed MDDECT, is constructed by 30 volunteer operators using the ECT device and consists of 48 000 samples of 20 classes in total. The MDDECT dataset is openly available and can be exploited as a testbed in order to promote new ECT algorithms. Third, 11 1-D residual networks of three different types are evaluated, and a 38-layer network ResNeXt1D-38 has achieved an accuracy of 93.58% in terms of estimating the depth of the surface defects from 0.3 to 2.0 mm with a depth resolution of 0.1 mm. In addition, the DL algorithms can reject liftoff signal and hence immune to liftoff noise. Future research directions would be to evaluate a different family of deep networks, such as recurrent networks and other learning strategies. Moreover, a continued effort should be made to enrich the MDDECT dataset with more diversities, scans, and specimens.

ACKNOWLEDGMENT

All research data supporting this publication are directly available within this publication.

REFERENCES

[1] A. Sophian, G. Y. Tian, D. Taylor, and J. Rudlin, "Electromagnetic and eddy current NDT: A review," *Insight*, vol. 43, no. 5, p. 6, 2001.
[2] J. García-Martín, J. Gómez-Gil, and E. Vázquez-Sánchez, "Non-destructive techniques based on eddy current testing," *Sensors*, vol. 11, no. 3, pp. 2525–2565, Feb. 2011.

[3] K. B. Ali, A. N. Abdalla, D. Rifai, and M. A. Faraj, "Review on system development in eddy current testing and technique for defect classification and characterization," *IET Circuits, Devices Syst.*, vol. 11, no. 4, pp. 338–351, Jul. 2017.

[4] F. Lingvall and T. Stepinski, "Automatic detecting and classifying defects during eddy current inspection of riveted lap-joints," *NDT & E Int.*, vol. 33, no. 1, pp. 47–55, 2000.

[5] A. Sophian, G. Y. Tian, D. Taylor, and J. Rudlin, "A feature extraction technique based on principal component analysis for pulsed eddy current NDT," *NDT & E Int.*, vol. 36, no. 1, pp. 37–41, 2003.

[6] J. Kim, G. Yang, L. Udpa, and S. Udpa, "Classification of pulsed eddy current GMR data on aircraft structures," *NDT & E Int.*, vol. 43, no. 2, pp. 141–144, Mar. 2010.

[7] Y. He, M. Pan, F. Luo, D. Chen, and X. Hu, "Support vector machine and optimised feature extraction in integrated eddy current instrument," *Measurement*, vol. 46, no. 1, pp. 764–774, Jan. 2013.

[8] J. A. Buck, P. R. Underhill, J. E. Morelli, and T. W. Krause, "Simultaneous multiparameter measurement in pulsed eddy current steam generator data using artificial neural networks," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 3, pp. 672–679, Mar. 2016.

[9] W. Deng, B. Ye, J. Bao, G. Huang, and J. Wu, "Classification and quantitative evaluation of eddy current based on kernel-PCA and ELM for defects in metal component," *Metals*, vol. 9, no. 2, p. 155, Feb. 2019.

[10] S. Hosseini and A. A. Lakis, "Application of time–frequency analysis for automatic hidden corrosion detection in a multilayer aluminum structure using pulsed eddy current," *NDT & E Int.*, vol. 47, pp. 70–79, Apr. 2012.

[11] R. Smid, A. Docekal, and M. Kreidl, "Automated classification of eddy current signatures during manual inspection," *NDT & E Int.*, vol. 38, no. 6, pp. 462–470, Sep. 2005.

[12] L. S. Rosado, F. M. Janeiro, P. M. Ramos, and M. Piedade, "Defect characterization with eddy current testing using nonlinear-regression feature extraction and artificial neural networks," *IEEE Trans. Instrum. Meas.*, vol. 62, no. 5, pp. 1207–1214, May 2013.

[13] T. Chen, G. Y. Tian, A. Sophian, and P. W. Que, "Feature extraction and selection for defect classification of pulsed eddy current NDT," *NDT & E Int.*, vol. 41, no. 6, pp. 467–476, 2008.

[14] B. Liu and Z. Li, "Study on the automatic recognition of hidden defects based on Hilbert Huang transform and hybrid SVM—PSO model," in *Proc. Prognostics Syst. Health Manage. Conf. (PHM-Harbin)*, Jul. 2017, pp. 1–7.

[15] 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, Sep. 2008.

[16] N. H. Jo and H.-B. Lee, "A novel feature extraction for eddy current testing of steam generator tubes," *NDT & E Int.*, vol. 42, no. 7, pp. 658–663, Oct. 2009.

[17] G. D'Angelo and S. Rampone, "Shape-based defect classification for non destructive testing," in *Proc. IEEE Metrol. Aerosp. (MetroAeroSpace)*, Jun. 2015, pp. 406–410.

[18] G. D'Angelo, M. Laracca, and S. Rampone, "Automated eddy current non-destructive testing through low definition Lissajous figures," in *Proc. IEEE Metrol. Aerosp. (MetroAeroSpace)*, Jun. 2016, pp. 280–285.

[19] G. D. Angelo, M. Laracca, S. Rampone, and G. Betta, "Fast eddy current testing defect classification using Lissajous figures," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 4, pp. 821–830, Apr. 2018.

[20] L. Yin *et al.*, "A novel feature extraction method of eddy current testing for defect detection based on machine learning," *NDT & E Int.*, vol. 107, Oct. 2019, Art. no. 102108.

[21] Y. Tao, H. Xu, J. R. S. Avila, C. Ktistis, W. Yin, and A. J. Peyton, "Defect feature extraction in eddy current testing based on convolutional sparse coding," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (IMTC)*, May 2019, pp. 1–6.

[22] J. Bao, B. Ye, X. Wang, and J. Wu, "A deep belief network and least squares support vector machine method for quantitative evaluation of defects in titanium sheet using eddy current scan image," *Frontiers Mater.*, vol. 7, Sep. 2020, Art. no. 576806.

[23] W. Deng, J. Bao, and B. Ye, "Defect image recognition and classification for eddy current testing of titanium plate based on convolutional neural network," *Complexity*, vol. 2020, pp. 1–10, Oct. 2020.

[24] S. Li *et al.*, "Learning to reconstruct crack profiles for eddy current nondestructive testing," in *Proc. Workshop Mach. Learn. Phys. Sci.*, Vancouver, BC, Canada, Dec. 2019, pp. 3003–3008.

[25] X. Fu, C. Zhang, X. Peng, L. Jian, and Z. Liu, "Towards end-to-end pulsed eddy current classification and regression with CNN," in *Proc. IEEE Int. Instrum. Meas. Technol. Conf. (IMTC)*, May 2019, pp. 1–5.

[26] K. Demachi, T. Hori, and S. Perrin, "Crack depth estimation of non-magnetic material by convolutional neural network analysis of eddy current testing signal," *J. Nucl. Sci. Technol.*, vol. 57, no. 4, pp. 401–407, Apr. 2020.

[27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 248–255.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2012, pp. 1097–1105.

[29] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. CVPR*, Jun. 2015, pp. 1–9.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015, pp. 2201–2214.

[31] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Highway networks," 2015, *arXiv:1505.00387*. [Online]. Available: https://arxiv.org/abs/1505.00387

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

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Computer Vision* (Lecture Notes in Computer Science), B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds. Cham, Switzerland: Springer, 2016, pp. 630–645.

[34] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2017, pp. 5987–5995.

[35] H. I. Fawaz, G. Forestier, J. Weber, L. Idoumghar, and P.-A. Muller, "Deep learning for time series classification: A review," *Data Mining Knowl. Discovery*, vol. 33, no. 4, pp. 917–963, Mar. 2019.

[36] L. N. Smith, "Cyclical learning rates for training neural networks," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2017, pp. 464–472.

[37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2015, *arXiv:1412.6980*. [Online]. Available: https://arxiv.org/abs/1412.6980

[38] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, "Learning deep features for discriminative localization," in *Proc. 29th IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2921–2929.

Tian Meng received the B.Eng. degree in electrical and electronic engineering from The University of Manchester, Manchester, U.K., in 2021, where he is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering working with Wuliang Yin.

His current research interests include artificial intelligence, machine learning, and electromagnetics.

Yang Tao received the B.Sc. degree in automation and the M.Sc. degree in control science and engineering from Tianjin University, Tianjin, China, in 2010 and 2013, respectively, and the Ph.D. degree in electrical and electronic engineering from The University of Manchester, Manchester, U.K., in 2018.

He worked as a Research Associate under the University of Manchester Impact Accelerator Account Project since 2018. Since 2020, he has been a Research and Development Engineer at Mettler Toledo Safeline, Manchester, specialized in artificial intelligence and deep learning. His research interests include electromagnetic instrumentation, metal detection, multifrequency power inverter, electromagnetic tomography, inverse problem, sparse representation, and deep learning.

Ziqi Chen received the B.Sc. degree in electrical and electronic engineering from The University of Manchester, Manchester, U.K., in 2018, where she is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering.

Her current research interests include developing hardware for electromagnetic sensing techniques, such as electromagnetic tomography (EMT) and multifrequency electromagnetic testing.

Jorge R. Salas Avila received the B.S. degree in electronic and computer engineering and the M.Sc. degree in informatics technology from Monterrey Institute of Technology and Higher Education, Monterrey, Mexico, in 2010 and 2013, respectively, and the Ph.D. degree in electrical and electronic engineering from The University of Manchester, Manchester, U.K., in 2020.

His current research interests include instrumentation, electromagnetic sensors, field-programmable gate array (FPGA)-based digital instruments, and nondestructive testing.

Qiaoye Ran received the B.Sc. degree in electrical engineering and automation from North China Electric Power University, Beijing, China, in 2018, and the B.Sc. degree in electrical and electronic engineering from The University of Manchester, Manchester, U.K., in 2018, where he is currently pursuing the Ph.D. degree in electrical and electronic engineering.

His current research interests include electromagnetic (EM) field modeling, EM sensor design, metal detection, EM interference immunity, and digital signal processing.

Yuchun Shao is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, The University of Manchester, Manchester, U.K., working with Wuliang Yin, mainly focusing on electromagnetic technology method (EMT) sensor design, for nondestructive testing (NDT).

Ruochen Huang is currently pursuing the Ph.D. degree working with Wuliang Yin with the School of Electrical and Electronics Engineering, The University of Manchester, Manchester, U.K., with a focus on finite-element method (FEM) modeling software packages for electromagnetic (EM) simulations.

Yuedong Xie (Member, IEEE) received the Ph.D. degree in electrical and electronic engineering from The University of Manchester, Manchester, U.K., in 2016.

After that, he worked as a Research Associate with The University of Manchester and a Research Fellow with the University of Warwick, Coventry, U.K. He is currently an Associate Professor with the School of Instrumentation and Opto-Electronic Engineering, Beihang University, Beijing, China. His research interests include electromagnetic nondestructive testing (NDT), ultrasonic-phased array, electromagnetic acoustic transducers, nanodevices, and energy harvesting.

Dr. Xie was a recipient of the Best Student Paper Award (First Place) and the Student Travel Grant Award from the IEEE International Instrumentation and Measurement Technology Conference in 2016.

Qian Zhao received the B.Sc. degree from the College of Engineering, Qufu Normal University, Jining, Shandong, China, in 2009, and the M.Sc. and Ph.D. degrees in electronic measurement and instruments from Tianjin University, Tianjin, China, in 2011 and 2014, respectively.

She is currently an Associate Professor with Qufu Normal University. Her current research interests include nondestructive eddy current sensing, machine learning, and electroencephalogram (EEG) signal processing.

Zhijie Zhang is currently a Professor with the National Key Laboratory for Electronic Measurement Technology, North University of China, Taiyuan, China. His research interests are thermocouple sensors, wireless sensors, eddy current testing, and field-programmable gate array (FPGA)-based temperature measurements.

Hujun Yin (Senior Member, IEEE) received the B.Eng. degree in electronic engineering and the M.Sc. degree in signal processing from Southeast University, Nanjing, China, in 1983 and 1986, respectively, and the Ph.D. degree in neural networks from the University of York, York, U.K., in 1997.

He is currently a Professor of artificial intelligence at The University of Manchester, Manchester, U.K. He has received research funding from U.K. research councils, EPSRC, BBSRC, Innovate U.K., and industries and has had over 25 funded projects. His research areas include signal/image processing, pattern recognition, machine and deep learning, time series modeling, and bioinformatics/neuroinformatics. He has supervised over 20 Ph.D. students in these areas and published over 200 peer-reviewed articles.

Dr. Yin has been a member of the EPSRC Peer Review College since 2006 and a Turing Fellow since 2018. He has served or has been serving as an Associate Editor for the IEEE TRANSACTIONS ON NEURAL NETWORKS, the IEEE TRANSACTIONS ON CYBERNETICS, and the *International Journal of Neural Systems*. He has also served as the General or Program Chair for a number of international conferences, such as the International Conference on Intelligent Data Engineering and Automated Learning and the International Symposium on Neural Networks.

Anthony J. Peyton received the B.Sc. degree in electrical engineering and electronics and the Ph.D. degree in medical instrumentation from the University of Manchester, Manchester, U.K., in 1983 and 1986, respectively.

After various positions in the scientific instrumentation industry and academia, he was appointed as a Professor in electromagnetic engineering in May 2004. He has over 30 years of experience in a diverse range of electromagnetic sensor systems and has been a Principal Investigator of numerous national, international, and industry-funded projects, totaling several £10M. He has been a partner on many major EU projects and a coauthor of over 160 international journal articles related to electromagnetics and sensing.

Wuliang Yin (Senior Member, IEEE) was appointed as an MT Sponsored Lecturer with the Department of Electrical and Electronic Engineering, School of Engineering, The University of Manchester, Manchester, U.K., in 2012, promoted to Senior Lecturer in 2016 and Reader in 2020. He has authored one book, more than 300 articles, and was granted more than ten patents in the area of electromagnetic (EM) sensing and imaging.

Dr. Yin was a recipient of the National Scientific and Technical Progress Award by the Chinese Government in 1999, the Excellent Paper Award–Agilent Measurement Forum in 2000, the Williams Award from The Institute of Materials, Minerals and Mining in 2014 and 2015, the 2016 Best Graduate Student Paper from the IEEE I2MTC Conference as a supervisor, the 2018 IEEE Instrumentation and Measurement Society Graduate Fellowship as a Supervisor, the 2020 Best Application Award from the IEEE Instrumentation and Measurement Society, and the IEEE gold medal as the most productive reviewer in the U.K. for IEEE TRANSACTIONS ON INSTRUMENTATION AND MEASUREMENT in 2020.