DOC-DICAM: Domain Aware One Class Defect Identification in Composite Aerostructure Material

Austin Yunker
Data Science and Learning Division
Argonne National Laboratory
Lemont, USA
ayunker@anl.gov

Rajkumar Kettimuthu
Data Science and Learning Division
Argonne National Laboratory
Lemont, USA
kettimut@mcs.anl.gov

Zachary Kral
Spirit AeroSystems
Wichita, USA
zachary.t.kral@spiritaero.com

***Abstract*—Fiber-reinforced composites are a common material used in the design of aircraft structures due to their good tensile strength and resistance to compression. During the manufacturing process, these structures are thoroughly inspected for flaws and defects to ensure structural integrity during commercial use. Non-destructive testing (NDT) is a collection of inspection methods that allow inspectors to evaluate material without altering it. Due to the high safety standards in aerospace manufacturing, the NDT process is done manually and can be a significant bottleneck in the development workflow. In this paper, we develop an AI-based assistance tool to drastically reduce inspection time. Typical AI workflows require large amounts of annotated data, but defects rarely occur resulting in strong class imbalance. To overcome this, we formulate the problem of defect identification as an anomaly detection task in which our primary focus is learning non-defect characteristics. To do this, we develop a multi-task self-supervised learning framework that embeds problem specific domain knowledge into the deep learning model. We verify our method using fuselage data generated in a production environment. We show that our method can effectively identify defects and requires minimal training and inference time.**

***Index Terms*—Non-destructive testing (NDT), ultrasonic testing (UT), deep learning, composites, defects, anomaly detection**

I. Introduction

State-of-the-art aircrafts are designed using fiber-reinforced composites due to their light-weight material, good tensile strength and resistance to compression [1]. Specifically, carbon-fiber-reinforced plastic (CFRP) is applied to wing planks, fuselages, and sandwich panel skins [2]. These materials are continuously improving in product efficiency, cost-effectiveness, and are commonly used in applications to load-bearing structures outside the aerospace domain including wind turbines, transportation, medical equipment, and so on. Manufacturing of composite materials is a multivariable task, involving many procedures, where various types of defects may occur within a composite product, giving rise to significant safety concerns in service. Defects resulting in a loss of mechanical properties can occur in all composite structures and tend to increase in frequency with structural complexity. Robust and reliable non-destructive testing (NDT), such as ultrasonic testing (UT), of composites is essential in the production stage to catch and repair defects.

Advanced computer vision techniques, particularly those based on machine learning algorithms, can provide new perspectives on the high-level visual understanding of universal tasks. The power of these techniques suggests a new approach to evaluating UT images, in which machine learning algorithms rather than manual inspection is used for the general detection and classification of manufacturing imperfections. Applied to the field of NDI, an early work [3] built three artificial neural networks (ANNs) for location, quantification, and classification of structural damage using vibrational characteristics in aluminum cantilever beams. N. Saeed, et al. [4] implemented a convolutional neural network (CNN) to detect artificially created sub-surface defects in a CFRP sample as well as a deep feed-forward neural network to estimate the defect depth using thermograms. Recently, methods have been proposed that frame defect identification as an anomaly detection task to overcome the limited amount of available defect data. Researchers in [5] evaluate three deep learning anomaly detection methods for the nondestructive evaluation of steel blocks using ultrasonic images. In the field of X-ray radiography, Presenti, et al develop an anomaly detection workflow using autoencoders for the automatic detection of defects in radiographs [6].

Researches in [7] proposed OC-DICAM that uses self-supervised learning (SSL) to identify defects using only non-defects during training. To learn general ultrasonic signal characteristics, they train a model to classify modified signal waveforms into the following classes: unaltered, flipped, reversed, flipped-reverse. They show this training scheme allows the model to learn good temporal and spatial characteristics. Our approach draws inspiration from theirs but makes two critical changes. First, classifying all unaltered non-defects into the same class makes the assumption that the signals are homogeneous and share similar characteristics. Instead, we break down the non-defect signals into sub-classes based on different material characteristics. This includes the material thickness, the front-wall voltage, and the back-wall voltage. Doing this allows us to embed domain knowledge about the material directly into the model. Therefore, our model consists of a shared encoder followed by four classifier heads: material thickness classifier, front-wall voltage classifier, back-wall classifier, and synthetic defect classifier. Secondly, we change how defects are scored. In [7], they measure the distance to the center in the latent dimension and flag signals sufficiently far

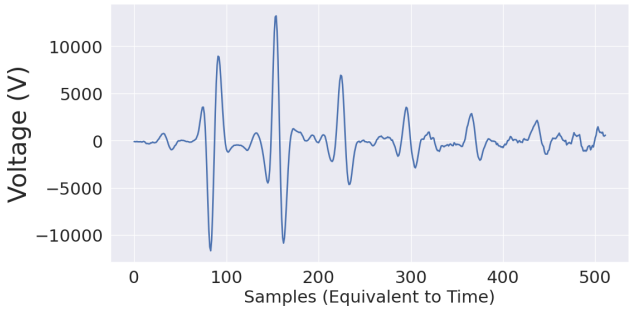
away. However, this assumes that during training the data is embedded into a tight hypersphere, yet they give no evidence that this occurs in practice. Instead, we show that measuring the entropy in the output of the synthetic defect classifier is a more effective scoring metric as defects tend to follow a more uniform distribution across the class outputs. We term our workflow as **D**omain Aware **O**ne **C**lass **D**efect **I**dentification in **C**omposite **A**erostructure **M**aterial (DOC-DICAM).

The remainder of this paper is organized as follows. Section 2 introduces the necessary materials and methods used. In Section 3, the corresponding results are reported and discussed. In Section 4, concluding remarks are made with possible directions for future work given.

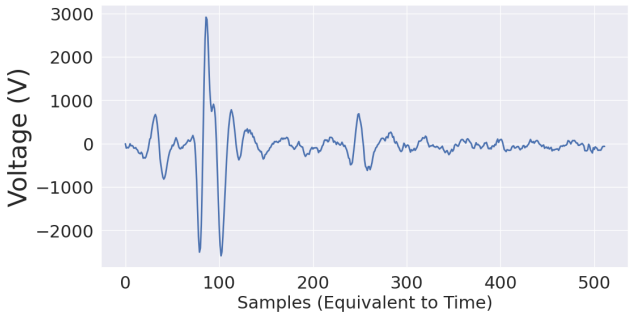
II. Materials and Methods

A. NDI Data

Ultrasonic testing (UT) is an NDI technique that is suitable for detecting and sizing flaws that are embedded within the surface by generating signal waveforms used for inspection [8]. Inspectors evaluate UT data visually using A-scans, B-scans, and C-scans. C-scans are used by default as they produce a view of the entire material with the pixel color representing a particular amplitude value. Figure 3a shows an example of a C-scan section of the fuselage. While Figure 3a shows the fuselage as a 2D image, the data is collected as a 3D volume with shape: height, width, and depth. The depth dimension corresponds to the recorded ultrasonic signal. Therefore, each pixel in Figure 3a has an associated signal waveform of 512 samples containing the measured voltage. Figure 1 shows the ultrasonic A-scan response from both a non-defect and porosity defect region from Figure 3a. The spikes in the non-defect region at time 80 and 160 are the recorded responses for the front and back wall echo, respectively. The smaller spikes seen past the back wall is the scanner picking up on the signal reflecting within the material due to the long scanning time. The defect A-scan records a response for the front wall but we can see it is significantly distorted in shape as well as reduced voltage compared to the non-defect. Furthermore, we can see the defect scatters the signal as seen in the attenuated back wall response.



(a) Non-Defect Signal



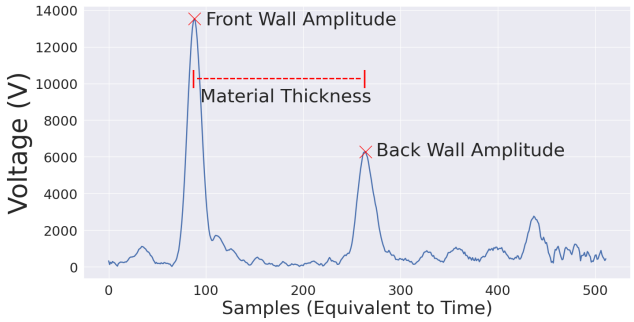
(b) Porosity Defect Signal

Fig. 1: A-scan Signal Waveform

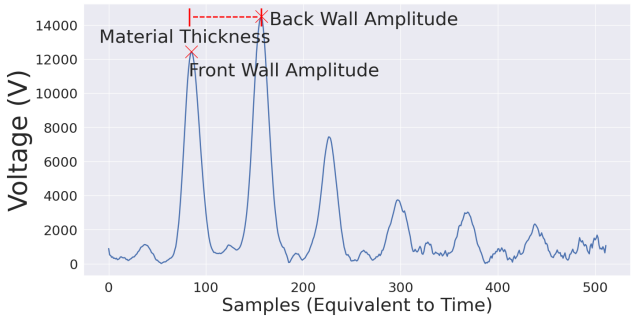
B. Automatic Labelling for Non-Defects

The proposed workflow requires having labelled information regarding the material thickness, front-wall voltage, and back-wall voltage. However, the data used here is only labelled as non-defect. Therefore, we devise an automatic labelling scheme necessary for the three tasks. The labels are determined using the information from the front and back wall calculated using the `find_peaks` method from the *SciPy* library [9]. The `find_peaks` method finds all local maxima and their time stamp. We assume that the two largest maxima correspond to the front and back wall peaks. To make identifying the peaks easier, we use the envelope of the signal instead of the original waveform shown in Figure 2. The front wall is assigned the voltage for the peak with the lower time stamp while the back wall is assigned the remaining voltage making use of the time-series nature of the data i.e., the front wall always occurs before the back wall. To calculate the material thickness, we take the difference between the back and front wall time stamps. An example of this labelling scheme is shown in Figure 2. We can see in Figure 2.a that the A-scan corresponds to a thicker region with a reduced back wall voltage compared to Figure 2.b showing the heterogeneity of the non-defect A-scans.

Finally, our approach is based on the classification of the three material characteristics; however, the above methods return real valued numbers. Therefore, to convert the information derived above into discrete labels for classification, we perform a binning operation. The thickness bins are based on intervals of 25 starting with [0-25] and ending with >400 giving 16 classes. Front wall voltage bins are based on intervals of 4000 starting at [0-4000] going up to >20000 giving 6 classes. Finally, the back wall voltage bins are based on intervals of 5000 starting at [0-5000] and ending at >30000 giving 7 classes. Figure 3 shows the output of applying this method to all the A-scans from the C-scan in Figure 3a. For the thickness label, regions in blue correspond to a thinner material than those in red. Similarly, blue regions in the front and and back wall label are those areas with a smaller voltage than the areas in red. Furthermore, we can see that the porosity defects significantly alter the generated labels for all three material characteristics. This is inline with how we expect defects to change the A-scans.



(a) Thick Material



(b) Thin Material

Fig. 2: Automatic Label Calculation for A-Scan Signal

C. DOC-DICAM

The overall framework is shown in Figure 4 which follows a similar approach to the work in [7]. Like [7], we make use of altered A-scans, referred to as synthetic defects throughout. To generate synthetic defects, we follow the approach in [10] that applies CutPaste, a data augmentation strategy that cuts an image patch and pastes it at a random location within the same image for image anomaly detection. Applied to NDT, we randomly cut a sequence from the A-scan and paste it back into the A-scan. Sequence lengths are randomly chosen between [64-256]. We then flip the A-scan by multiplying it by -1, further increasing the difference between non-defects and synthetic defects. We note that real defects may not always follow the above augmentation strategy. Therefore, we primarily use synthetic defects to regularize the model. Thus, in addition to classification, we incorporate the loss function proposed in [11] that encourages the model to output a uniform distribution for synthetic defects.

During training, we input both non-defects and synthetic defects into the model. For non-defects, the model classifies them based on the material thickness, back-wall voltage, front-wall voltage, and signal class. Synthetic defects are only used in the signal classifier. The flow for non-defects and synthetic defects through the model can be seen in Figure 4 by the green and red arrows, respectively. Our model consists of a shared encoder and four linear classifier heads making it a multi-task learning framework.

Overall, the model is required to learn four classification tasks based on the characteristics of the scan as well as being regularized to synthetic defects. We do this using the standard cross entropy loss for all tasks. For regularization, we compare the cross entropy from the output of the signal classifier to the uniform distribution when inputting the synthetic defects. We find that both predicting the synthetic defects as a separate class as well as regularizing against them performs better than only doing either one. Therefore, the total model loss is given as

$$L_{total} = L_{THK} + L_{FWV} + L_{BWV} + L_S + L_R, \quad (1)$$

where THK is the thickness classifier loss, FWV is the front wall voltage classifier loss, BWV is the back wall voltage classifier loss, S is the signal classifier loss, and R is the regularization loss for synthetic defects. We give an equal weight of 1 to each loss term during training.

During inference, we score test signals based on the class entropy using the signal classifier shown in the bottom half of Figure 4. Therefore, for a given test signal $\boldsymbol{x} \in X$, we define anomaly score s as

$$s(x) = -\sum_i p_i log(p_i), \quad (2)$$

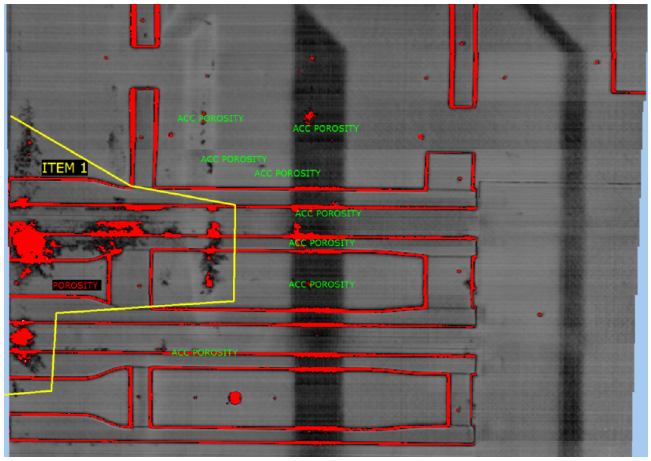
where we sum over the number of classes and p is the class probabilities. Since the model was trained to output a uniform distribution for synthetic defects, we hypothesize that real defects will have a higher entropy than non-defects. To flag defects, we set a threshold on the distribution of scores for the test scan.

A final remark is on the threshold selection. A common approach is to set the threshold based on the False Positive Rate (FPR)-N. The FPRN metric is the probability that a non-defect signal raises a false alarm when $N\%$ of defects are detected. However, this requires a dataset with defects. To overcome this, we use our non-defect validation dataset to create a distribution of entropy scores. We expect some noise in the dataset so we set a threshold, τ, based on three standard deviations above the mean entropy score. Thus, for model ϕ, we define decision function G to flag defects based on:

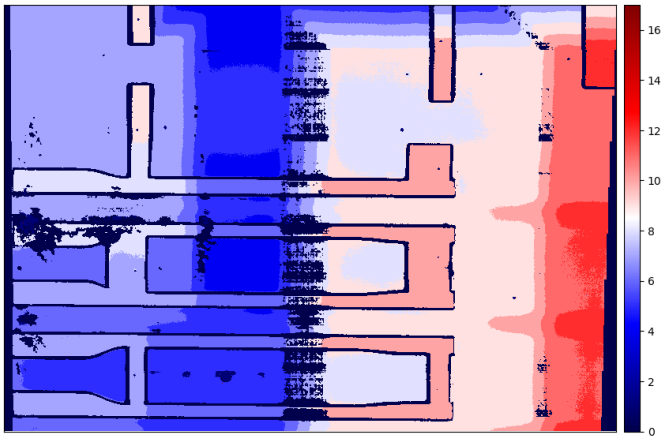
$$G(\boldsymbol{x}; \phi, \tau) = \begin{cases} 0 & \text{if } s(\boldsymbol{x}) < \tau \\ 1 & \text{if } s(\boldsymbol{x}) \geq \tau \end{cases} \quad (3)$$

D. Model Architecture

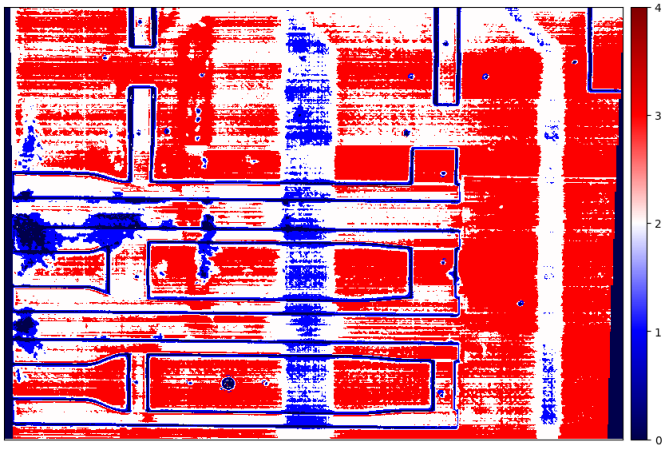
The proposed workflow requires the classification of A-scan signals based on the four learning tasks which can be framed as a time series classification task. This problem has been extensively studied with numerous model architectures developed. Since our work seeks to address the shortcomings of OC-DICAM [12], we reuse their model which was based on the EISATC-Fusion model originally developed in [13] for motor imagery decoding using electroencephalography (EEG) signals. The researchers in [13] found that the EISATC-Fusion model was capable of extracting time-domain, space-domain, multi-scale, and long-time dependent features of EEG signals for accurate end-to-end intra subject and inter subject motor



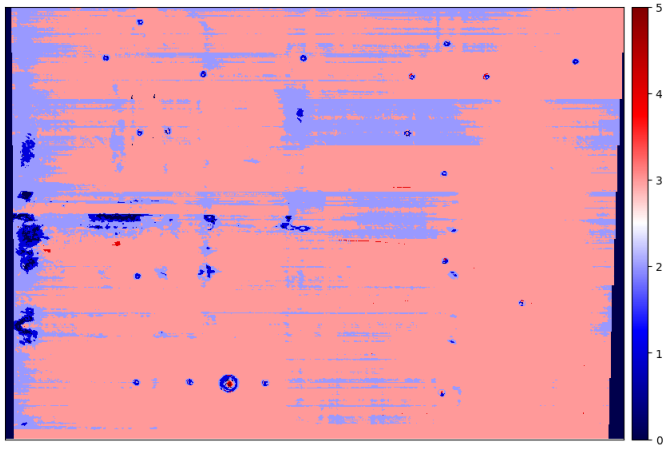
(a) Original Scan



(b) Material Thickness Label



(c) Back Wall Voltage Label



(d) Front Wall Voltage Label

Fig. 3: Automatic Labels for the Test Scan

imagery decoding. We briefly describe the model architecture below. A full description can be found in [13].

The EISATC-Fusion model consists of four sequential modules within the encoder: the EEGNet-Inception (EEGInc) module, multi-head self-attention (MSA) module, temporal convolutional network (TCN) module, and fusion module. The core structure of the EEGInc module is composed of three convolutional layers and two average pooling layers. The first layer is a temporal convolutional, the second is a channel convolutional using depth-wise convolution, and the last layer is the inception block with three paths composed of depth-wise convolution and a residual path. The MSA module applies the standard self-attention mechanism based on the query, key and value components calculated from the input sequence. The TCN module uses casual convolution that restricts the convolutional kernel to the current and past time steps, ensuring the model cannot learn future information. Dilated convolutional is used to increase the receptive field of the model without increasing the number of convolutional layers. The fusion module consists of two parts: feature fusion and decision fusion. The feature fusion combines the output of different levels in the model to extract the hidden information of the input data and improve the representation ability of the feature. The decision fusion combines the outputs of multiple classifiers to reduce the uncertainty and errors of individual classifiers, improving the information integration ability of the model and allowing the model to obtain more reliable decisions. The output of the encoder portion is shared among the four classification tasks and used as input into four fully connected linear modules.

E. Performance Evaluation Criteria

To evaluate the model performance, metrics commonly reported are accuracy, F1-score, precision, and recall. For defect detection, recall is usually preferred over the others reflecting the increased importance in having a low false negative i.e., missed defects. However, the scan shown in Figure 3a only has text labels indicating the general defect region. Inspectors rarely provide pixel perfect annotations as doing so drastically slows down the NDI process. Therefore, we are unable to calculate metrics based on the pixel level. Furthermore, we are primarily concerned with identifying defects as a cluster of

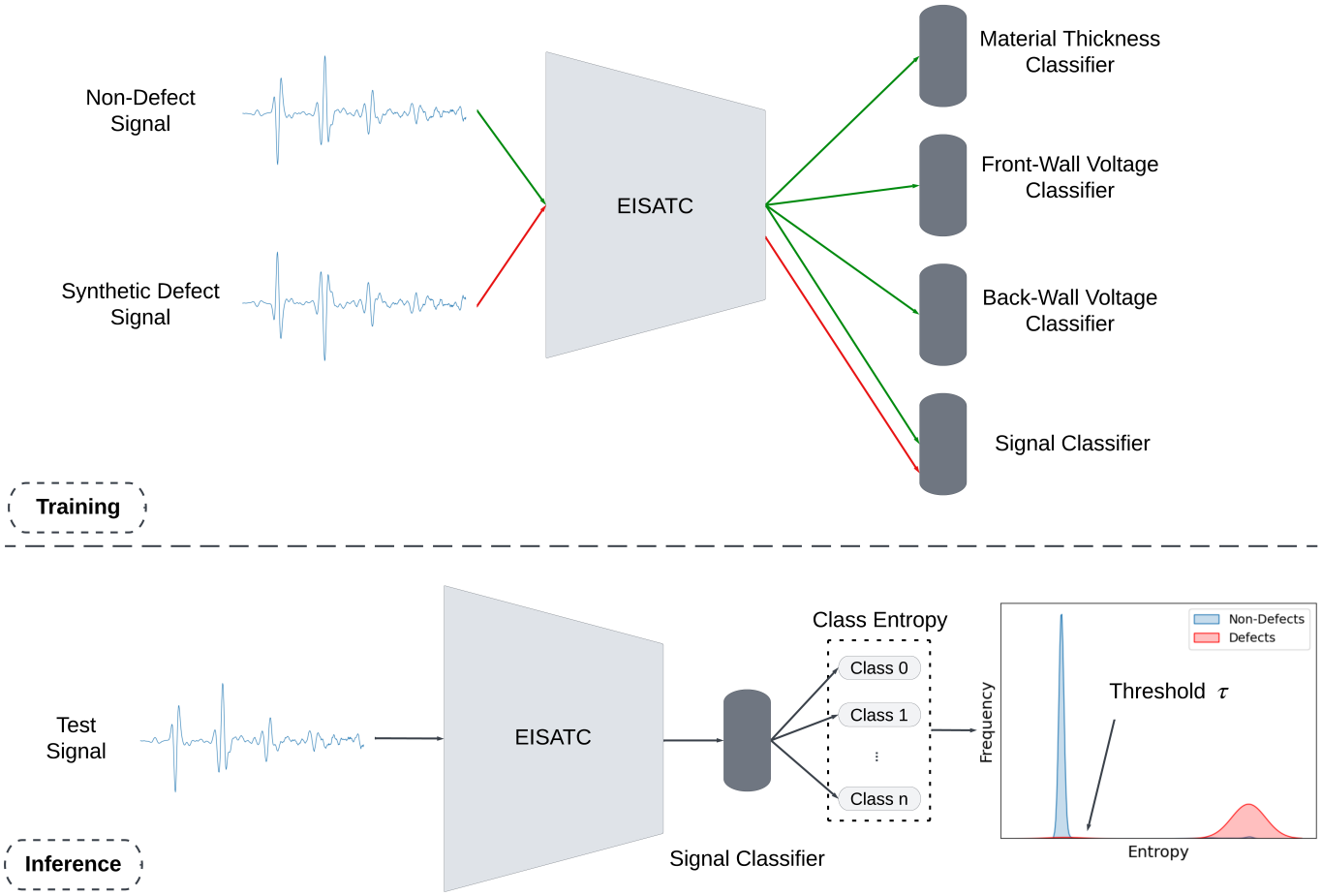


Fig. 4: DOC-DICAM Workflow

pixels and not individual pixels, so we can consider a success as a majority of the pixels being identified qualitatively.

III. Results and Discussion

A. Experimental Details

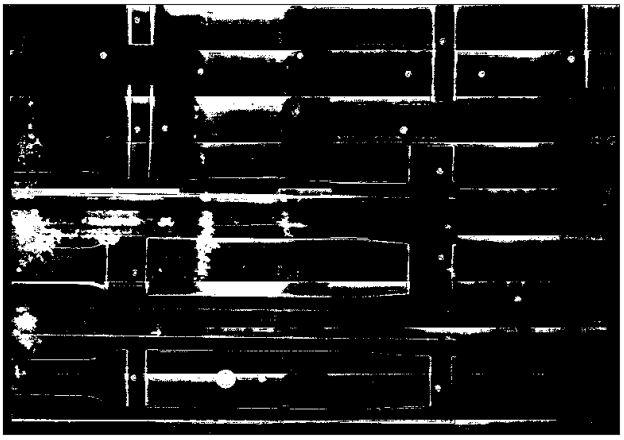
The EISATC-Fusion model was built using the PyTorch 1.12 deep learning framework with Python 3.7 and trained using a single Nvidia Tesla V100-SXM2-32GB card [14]. The model was optimized using the Adam optimizer using a learning rate of 0.001 with the final model saved based on the smallest validation error. After training, we save the cutoff threshold calculated using the validation dataset. Since we are utilizing a time series based model we can flatten the first and second dimension and add a dimension for the channel information. Figure 3a has a 3D shape of (540, 774, 512) which after reshaping gives (417960, 1, 512) where 1 represents the single channel consisting of the ultrasonic waveform. The training and validation data consists of previously archived fuselage data. Finally, we preprocess the data by scaling it between [0-1].

B. Results

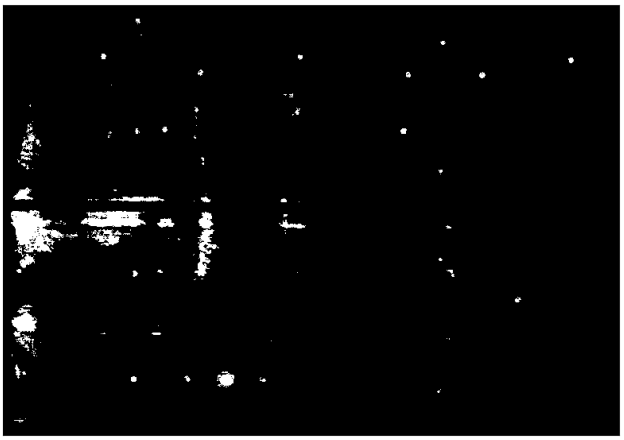
Figure 5 shows the OC-DICAM and DOC-DICAM prediction for the test scan. Both workflows clearly highlight the large areas of porosity on the left region of the scan as well as the smaller regions in the center marked as "ACC POROSITY" giving a low false negative rate. However, the OC-DICAM workflow predicts a large amount of false positives with most of those areas corresponding to the stringers within the scan. The method proposed here does produce some false positive areas but upon further investigation we see that these areas correspond to the fastener holes which are benign anomalies. Finally, the DOC-DICAM prediction shows that it is able to clearly identify the cluster centers but may not precisely identify the boundary between the defect and non-defect area. This may reflect the severity at which porosity distorts the A-scan making certain regions easier to identify than others.

IV. Conclusion

In this work, we proposed the DOC-DICAM workflow. This multi-task self-supervised learning workflow explicitly embeds material specific domain knowledge into the deep learning model. We do this by having the model classify the material thickness, front wall voltage, and back wall voltage. By incorporating relevant domain information about the scan characteristics into our workflow, we are able to overcome the limitations of OC-DICAM. When applied to real fuselage data,



(a) OC-DICAM Prediction



(b) DOC-DICAM Prediction

Fig. 5: Model Prediction for Porosity Scan

we are able to accurately identify large regions of porosity with false positives attributed to benign anomalies. Finally, our method is extremely efficient requiring only a few seconds to perform the inference step.

Moving forward we seek to extensively verify the method developed here. The scan considered here is only one of many from the fuselage. Furthermore, there are other defect types such as foreign object debris (FOD) and delamination. We also note that while giving equal weight to each loss function during training is simple, it is naive to think that this is the optimal weighting strategy. Recent methods in the field of multi-task learning have investigated dynamically changing weights during training reflecting model uncertainty for a particular task as well as how to best combine the gradients such that tasks don't conflict with one another during the model update step. Finally, a more optimal method may exist in setting the threshold to flag defects that adequately balances the trade off between false positives and negatives.

ACKNOWLEDGEMENTS

This work was supported in part by the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy (EERE), under contract DE-AC02- 06CH11357.

LICENSE

REFERENCES

[1] Giurgiutiu, Victor. *Introduction*. Academic Press, London (2016): .

[2] Ibrahim, M.E. "Nondestructive evaluation of thick-section composites and sandwich structures: A review." *Composites Part A: Applied Science and Manufacturing* Vol. 64 (2014): pp. 36–48. DOI https://doi.org/10.1016/j.compositesa.2014.04.010.

[3] Carlos M. Ferregut, Roberto A. Osegueda and Ortiz, Jamie. "Artificial Neural Networks for Structural Damage Detection and Classification." *SPIE Proceedings* (1995)DOI https://doi.org/10.1117/12.207718.

[4] Numan Saeed, Zafar Said, Nelson King and Omar, Mohammed A. "Automatic defects detection in CFRP thermograms, using convolutional neural networks and transfer learning." *Infrared Physics Technology* Vol. 102 (2019): p. 103048. DOI https://doi.org/10.1016/j.infrared.2019.103048.

[5] Posilović, Luka, Medak, Duje, Milković, Fran, Subašić, Marko, Budimir, Marko and Lončarić, Sven. "Deep learning-based anomaly detection from ultrasonic images." *Ultrasonics* Vol. 124 (2022): p. 106737.

[6] Presenti, Alice, Liang, Zhihua, Alves Pereira, Luis F, Sijbers, Jan and De Beenhouwer, Jan. "Automatic anomaly detection from X-ray images based on autoencoders." *Nondestructive Testing and Evaluation* Vol. 37 No. 5 (2022): pp. 552–565.

[7] *OC-DICAM: One Class Defect Identification in Composite Aerostructure Material*, Vol. ASME 2024 Aerospace Structures, Structural Dynamics, and Materials Conference of *Aerospace Structures, Structural Dynamics, and Materials Conference* (2024). DOI 10.1115/SSDM2024-121466. URL `https://doi.org/10.1115/SSDM2024-121466`.

[8] Felice, Maria V. and Fan, Zheng. "Sizing of Flaws Using Ultrasonic Bulk Wave Testing: A Review." *Ultrasonics* Vol. 88 (2018): pp. 26–42. URL `https://doi.org/10.1016/j.ultras.2018.03.003`.

[9] Jones, Eric, Oliphant, Travis, Peterson, Pearu et al. "SciPy: Open source scientific tools for Python." (2001–). URL `http://www.scipy.org/`.

[10] Li, Chun-Liang, Sohn, Kihyuk, Yoon, Jinsung and Pfister, Tomas. "Cutpaste: Self-supervised learning for anomaly detection and localization." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*: pp. 9664–9674. 2021.

[11] Hendrycks, Dan, Mazeika, Mantas and Dietterich, Thomas. "Deep anomaly detection with outlier exposure." *arXiv preprint arXiv:1812.04606* (2018).

[12] Austin Yunker, Rami Lake, Rajkumar Kettimuthu and Kral, Zachary. "Comparative Study on Deep Learning Methods for Defect Identification and Classification in Composite Aerostructure Material." *Journal of Nondestructive Evaluation, Diagnostics and Prognostics of Engineering Systems* (2023).

[13] Liang, Guangjin, Cao, Dianguo, Wang, Jinqiang, Zhang, Zhongcai and Wu, Yuqiang. "EISATC-Fusion: Inception Self-Attention Temporal Convolutional Network Fusion for Motor Imagery EEG Decoding." (2023).

[14] Paszke, Adam and others. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. Curran Associates Inc., Red Hook, NY, USA (2019): .