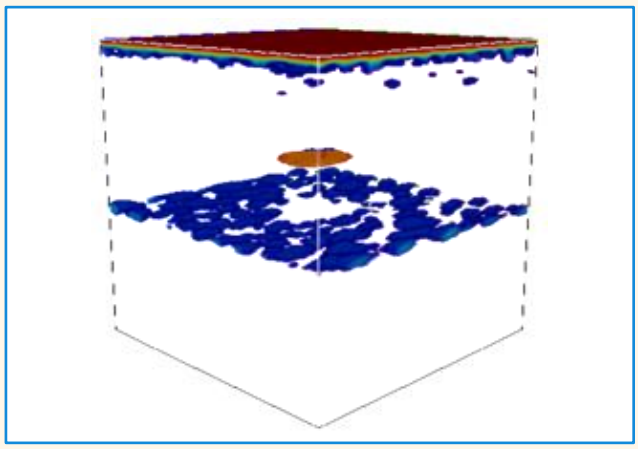
Advancing Carbon Fiber Composite Inspection: Deep Learning-Enabled Defect Localization and Sizing via 3-Dimensional U-Net Segmentation of Ultrasonic Data

Shaun McKnight[1*], Vedran Tunukovic[1], S. Gareth Pierce[1], Ehsan Mohseni[1], Richard Pyle[1], Charles N. MacLeod[1], Tom O'Hare[2]

***Abstract*— In Non-Destructive Evaluation (NDE), accurately characterizing defects within components relies on accurate sizing and localization to evaluate the severity or criticality of defects. This study presents for the first time a deep learning methodology using 3-Dimensional (3D) U-Net to localize and size defects in Carbon Fibre Reinforced Polymer (CFRP) composites through volumetric segmentation of ultrasonic testing data. Using a previously developed approach, synthetic training data closely representative of experimental data was used for the automatic generation of ground truth segmentation masks. The model's performance was compared to the conventional amplitude 6 dB drop analysis method used in industry against ultrasonic defect responses from 40 defects fabricated in CFRP components. The results showed good agreement with the 6 dB drop method for in-plane localization and excellent through-thickness localization, with Mean Absolute Errors (MAE) of 0.57 mm and 0.08 mm, respectively. Initial sizing results consistently oversized defects with a 55% higher mean average error than the 6 dB drop method. However, when a correction factor was applied to account for variation between the experimental and synthetic domains the final sizing accuracy resulted in a 35% reduction in MAE compared to the 6 dB drop technique.**

By working with volumetric ultrasonic data (as opposed to 2D images) this approach reduces pre-processing (such as signal gating) and allows for the generation of 3D defect masks which can be used for the generation of computer aided design files; greatly reducing the qualification reporting burden of NDE operators.



***Index Terms*— Ultrasonic Testing, Segmentation, Three-Dimensional, U-Net, Composite, Deep Learning, Defect Characterization**

I. Introduction

Composite materials find extensive application within the aerospace, marine and civil engineering sectors, with Carbon Fiber Reinforced Polymers (CFRP) being one of the most prominent. In recent commercial aircraft developments, such as the Airbus A350 and the Boeing 787, CFRP constitutes more than 50% by weight of the final structures. This prevalence extends to private jets and helicopters, where CRFP can even reach utilization of 70-80% by weight [1], [2]. CFRP's production involves the layering of multiple carbon ply sheets, preforming, and subsequent curing using a thermoset polymer in a mold. The inherent anisotropy of these composites, caused by the direction of fiber filaments in the fabrics' weaving patterns and components' layup sequences, allows for them to be engineered to meet precise structural requisites, rendering them ideal for high-performance applications at markedly reduced weight [3], [4], [5], [6], [7], [8], [9], [10], [11]. The complexities associated with manufacturing of these components can introduce defects, limiting both structural integrity and performance [3], [4], [6], [9], [10], [12], [13]. These defects encompass a broad spectrum, spanning from delamination and cracks to foreign object inclusions, ply stacking errors, fiber distortions, and porosities [8], [13]. Given the increasing use of composites in safety-critical components, the identification, characterization, and quantification of

This work was supported through Spirit AeroSystems/ Royal Academy of Engineering Research Chair for In-Process Non-Destructive Testing of Composites, RCSRF 1920/10/32.

Corresponding Author: Shaun McKnight

[1]*Sensor Enabled Automation, Robotics, and Control Hub (SEARCH), Centre for Ultrasonic Engineering (CUE), Electronic and Electrical Engineering Department, University of Strathclyde, Glasgow, UK*

[2]*Spirit AeroSystems, Belfast, UK*

Highlights

- **3D U-Net is employed to segment defects in volumetric ultrasonic responses, giving accurate sizing and 3D localization information.**
- **When accounting for the difference between the experimental and synthetic domains, the segmentation masks improved on the industrial 6 dB drop sizing method by a 35% reduction in MAE.**
- **This work provides a solution to address the industrial need for automated defect characterization, by using synthetic data to train a 3D U-Net.**

defects is of paramount importance [6].

Non-Destructive Evaluation (NDE) encompasses an array of methodologies employed for the examination of components, without impacting their structural integrity. Radiography, thermography, electromagnetic approaches, and ultrasound stand out as some of the most prevalent NDE techniques. The selection of the most fitting NDE technique hinges upon the inherent characteristics of the component in question, logistical requirements of the inspection, and the specific defects targeted for detection.

Ultrasonic Testing (UT) has found widespread adoption and standardization for volumetric inspections of composite components within the aerospace industry. UT's capabilities encompass the detection of a diverse array of volumetric defects [4], [8], [11], [12] whilst being comparatively straightforward to implement and free from hazards when compared to radiography. UT draws its foundation from the transmission, propagation, and reception of ultrasonic waves. It is generally used for volumetric inspection where a sound wave is excited on a component's surface. These waves travel through the material, with the resulting internal reflections and scatterings offering valuable insights into the component's volumetric integrity.

Phased arrays are often the technology of choice for transmission and reception of acoustic waves, primarily due to their operational flexibility, larger coverage area, and reduced inspection times. Phased arrays are constructed from independently controllable ultrasonic transducers positioned in an array or matrix formation, allowing for a range of electronic scanning and imaging capabilities such as beam steering, dynamic depth focusing, and variable sub-apertures [10]. The addition of a roller tire material enables the movement of the unit across the surface of different samples, achieving direct coupling without posing the risk of damaging the array. This is facilitated by positioning the ultrasonic transducers in the center of rotation of the tire, which is filled with liquid to promote ultrasonic propagation. Subsequently, phased arrays are widely used in industries such as aerospace and energy [14], [15].

With control of each individual element or sub-aperture within a linear phased array, depth-wise sectional images, referred to as B-scans, can be generated from a single array (Figure 1). When combined with mechanized scanning conducted perpendicular to the linear phased array's length, three-dimensional (3D) volumetric scan data can be produced. This is achieved by stacking multiple individual B-scans at known positions. Volumetric ultrasonic data is frequently summarized in the form of images. These images can be obtained by selecting a single B-scan or by generating amplitude and time-of-flight C-scans. In the case of C-scans, the image represents a top-down sectional view of the sample, with the choice of either displaying the maximum response amplitude or the time index of the maximum response amplitude within the volume [16].

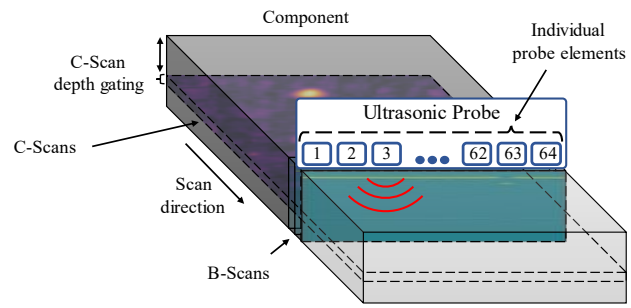


Figure 1: Demonstration of how individual probe elements comprise a linear phased array for production of B-scan and C-scan images.

The incorporation of robotics into NDE has ushered in a revolution in the inspection of large-scale components through increased automation [17]. Robotic scanning, compared to manual scanning, offers enhanced positional accuracy, repeatability, and substantially reduces the time required for scans, with previous systems demonstrating impressive scanning rates of 25.3 square meters per hour [18]. However, the analysis of results in industrial settings remains a laborious manual and time-intensive task. This necessitates the presence of highly trained and certified NDE operators to ensure results are in line with established standards [11], [19], [20], [21], [22], [23]. Human interpretation presents two drawbacks, namely, poor time efficiency and the risk of human error [21]. Therefore, research and development of automated data interpretation methods which can complement human interpretation would enhance the efficiency and reliability of NDE. By diminishing reliance on human interpretation, automation has the potential to bolster the consistency, repeatability, and traceability of NDE processes while concurrently trimming inspection durations and costs.

An automated NDE data interpretation and reporting pipeline can be broken down into distinct sections which in many cases will flow sequentially, as presented in Figure 2. In most settings, defect detection is the primary task to be completed. Once defects are detected it is important to evaluate their size against acceptance criteria. In some applications the acceptance criteria will vary depending on the defect type, orientation, and location, so the identification of different defect types and the location in the geometry is also important. In an industrial setting it is crucial to report on the NDE findings for traceability and to allow for downstream testing, design decisions, and re-work. It is therefore not acceptable to just detect a defect. Defects must be evaluated to extract as much information about

the defect as possible, such as sizing, type, localization etc. [24]. This is covered by the characterization and quantification section of the pipeline (Figure 2). Whilst there is potential for a single end-to-end model/system to complete the whole interpretation process, by breaking the tasks up it will allow for greater model testing and improved understanding and comparison/agreement to human operators. Breaking down the data interpretation process into separate tasks not only enables comprehensive model testing but also provides valuable insights into how errors propagate through the system. This decomposition allows for a more nuanced evaluation, shedding light on the strengths and limitations of each component of the analysis. Additionally, the isolation of specific tasks enhances the ability to discern how errors manifest at different stages, making it possible to facilitate a more effective comparison and agreement with human operators. Their expertise can be leveraged to refine and optimize each element. This stepwise methodology ensures a thorough assessment of the entire system and establishes a foundation for collaborative decision-making between automated systems and human operators. Moreover, this modular approach enhances the system's adaptability to different requirements, allowing for customization and optimization based on specific application needs.

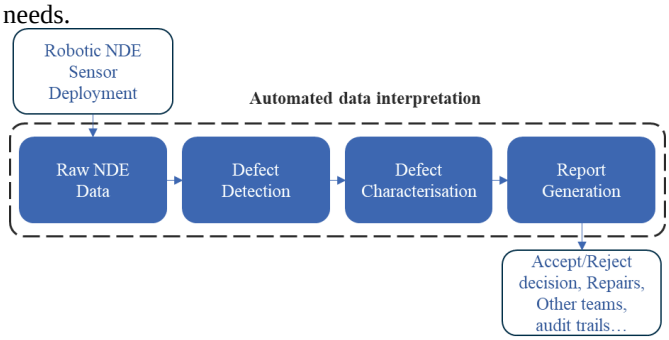


Figure 2: Potential automated data interpretation pipeline.

Much of NDE automation research focuses on defect detection, often overlooking the burden placed on NDE operators for report generation. By analyzing and segmenting complete ultrasonic volumes of a sample, this work focuses on defect sizing and 3D localization, which can be used directly when evaluating defects against accept/reject criteria and provide accurate defect positional information. This information is valuable in supporting downstream manufacturing processes such as potential material re-work to repair defects, and quality report generation through the automated creation of computer-aided design files or digital twins of inspected components [25], [26].

Accurate defect sizing is a key metric in determining if a component is safe against standardized acceptance criteria. The 6 dB drop method is a widely accepted method for defect sizing and is commonly used in industrial standards [27]. The technique relies on the utilization of a single transducer and the peak amplitude from the defect response to determine the boundaries of a defect response by detecting the point at which the transducer is directly over the edge of the defect as determined by a 50% energy dissipation from the reflector, manifesting as a 6 dB reduction in amplitude as indicated in Figure 3. The method benefits from being based on physical properties and is fully explainable. The 6 dB drop is often extended and applied to amplitude C-scans and phased arrays for thresholding defect areas [28], [29], [30]. Whilst the 6 dB drop method is widely established it does have limitations. Primarily, the defect must be larger than the acoustic beam to get an accurate value for the peak response amplitude [28]. In addition, real defect responses generally do not follow the ideal defect response curve, often leading to under sizing defects [31]. To combat this, alternative amplitude drop thresholds are used in different industrial settings [27], [32]. These are often component specific and require experimental determination.

Several works focused on sizing of defects captured with UT by using methods alternative to 6 dB drop. Li et al. proposed a method which utilized a generalized regression neural network and took additional features into account to provide dynamic thresholding of a C-scan image for more accurate defect sizing than the 6 dB drop method [33]. Niu and Srivastava [34] successfully sized and characterized cracks in load bearing structures using an DL approach from A-scan data. Study detailed in [35] also employed DL to characterize sizes and orientation of cracks in pipes from plane wave imaging data. Lardner et al. [36] proposed an automated approach to determine defect size and depth from ultrasonic data captured from pressure tubes used in nuclear sector. The sizing was achieved through a 6 dB drop while the depth estimation was done by examining data from several sensors. Whilst the 6 dB drop can be used for in-plane defect localization, depth-wise localization requires information from the time trace signal. Cheng et al. showed a promising method for depth localization of defects in CFRP panels using different Deep Learning (DL) approaches [37] with A-scan signals. They reported a minimum depth relative error of 8% for the hybrid CNN-LSTM. However, the aforementioned works base their analysis on parts of the captured UT signal instead of the full volumetric data.

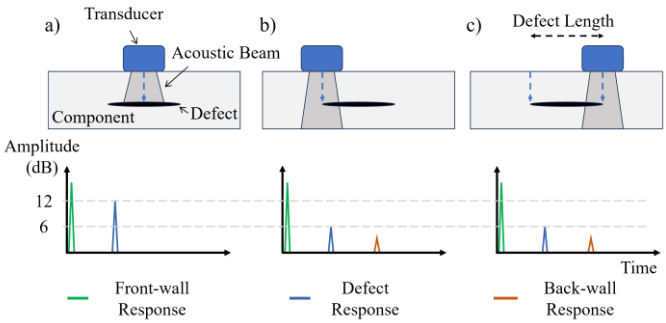


Figure 3: Demonstration of the 6 dB drop method for defect sizing. a) Finding maximum defect response. b) Using the 6 dB loss in maximum amplitude to locate one edge of the defect. c) The corresponding defect edge detected using the 6 dB drop to determine the defect length.

DL models present an avenue for advancing the automation of NDE data interpretation, and are becoming more prevalent in the literature, especially in the context of defect detection when dealing with images or A-scans [23]. DL is particularly well suited to addressing challenging automation tasks, where a traditional method may not be available, such as defect characterization and quantification. There are several examples of DL models demonstrating the ability to exceed human

performance in certain situations [38], [39]. DL solutions geared toward automating the interpretation of UT results, which integrate with robotic inspection systems, have the potential to elevate the quality and efficiency of inspections of large components. Such advancements would yield a reduction in data interpretation time and a broader adoption of automated NDE practices in aerospace and various other industries.

While the application of DL techniques in ultrasonic signal analysis for composite components holds promise, its widespread utilization has remained relatively constrained [23]. One of the primary hurdles impeding advancements in this field is the lack of training data. This scarcity, coupled with industry apprehensions regarding the interpretability and lack of relevant standards, has posed challenges to the effective integration of DL techniques.

While existing DL methods in the literature primarily utilize data formats that are readily interpretable by humans, such as images, it is important to note that DL algorithms are not constrained to image-level analysis alone. They have demonstrated remarkable capability in interpreting complex 3D volumetric data [40], [41], [42].

This work extends the authors previous work where a DL solution was utilized for the detection of defects from volumetric UT data. The previous work detailed data processing and the use of synthetic data generation along with domain specific augmentations to help bridge the simulation to experimental domain gap. Synthetic datasets play a pivotal role in DL to bolster small training datasets [43]. They have demonstrated their efficacy in UT of composites, particularly in the classification of C-scan images, yielding promising outcomes over simulated data alone [44]. In the previous volumetric defect detection work synthetic datasets are generated through simulations facilitated by semi-analytical physics-based software (CIVA [45]), known for its ability to generate experimentally precise defect responses [46], [47]. This software provides a more computationally efficient alternative to Finite Element Analysis (FEA), enabling the simulation of composite responses based on bulk material properties [48]. Whilst these concepts are introduced and discussed in this manuscript, for further details and visualizations the reader is referred to the prior work [42].

This paper presents an alternative method for defect sizing using 3D U-Net for volumetric segmentation of ultrasonic data, evaluated against the established 6 dB drop method. U-Net is a DL model, introduced in 2015, which proposed an architecture for medical two-dimensional (2D) image segmentation [49]. Despite advances in computer vision the U-Net architecture is still widely popular and shows impressive results in many different segmentation tasks [50]. Çiçek et al. [51] extended the original U-Net paper for 3D segmentation of highly variable kidney volumes, giving 3D U-Net, which showed impressive results. By incorporating algorithms with the capability to interpret volumetric data, it ensures that all spatial and depth wise information is preserved. This approach provides the model with more pertinent features to learn from and eliminates the necessity for image pre-processing and gating. Furthermore, it enables comprehensive 3D defect localization, a key focus of this study. To the best of the authors' knowledge, this marks the first utilization of a 3D U-Net for sizing and localization of defects in volumetric ultrasonic testing data, offering several advantages:

- Reduced pre-processing times as no thresholding, gating, or generation of images is needed.
- Results of the developed model trained exclusively on synthetic datasets, outperform industry standard 6 dB drop method for sizing by 35% on experimental test data.
- Complete localization of defects within 3D space. Which enables easy extraction of data for downstream processes i.e. testing with Finite Element Analysis.

Section II of the paper provides information on the collection of experimental test data, the generation of synthetic training data, and the creation of ground truth segmentation masks. It also covers signal processing and the augmentation applied during training. In Section III, the network architecture, training specifics, and the reference 6 dB drop method for sizing are detailed. The results and discussion are presented in Section IV, which is divided into localization and sizing.

II. Experimental Test Data and Synthetic Training Data

Synthetic data was used to train the 3D U-Net model in a fully supervised manner without the need for any experimental training data. The trained model's performance was evaluated against a fully experimental test dataset.

A. Experimental Data

Experimental ultrasonic data was acquired from two separate CFRP samples, with artificially introduced defects, to serve as test data. Delamination's are one of the most common defect types in composites [52] and a major life-limiting failure mode [53]. To imitate delamination defects and act as reference defects of known sizes, Flat-Bottom Holes (FBHs) were drilled from the backside of the samples up to known depths from the clean surface. FBHs are often used for this application, as detailed in [54].

Spirit AeroSystems provided composite samples with dimensions of 254.0 x 254.0 x 8.6 mm (width x depth x height). These samples were manufactured in compliance with the BAPS 260 specification, utilizing woven fabric and Cycom 890 resin through a vacuum-assisted resin transfer molding process. The first sample was characterized by the presence of 25 FBHs, each possessing diameters of 3.0, 4.0, 6.0, 7.0, and 9.0 mm. These holes were drilled to depths of 1.5, 3.0, 4.5, 6.0, and 7.5 mm from the front surface. These distinct defect sizes were spaced at intervals of 30 mm from one another, with variations in depth defects positioned 35 mm apart. The second sample featured 15 FBHs, all of which were drilled to the same depths as those in the first sample. However, the defect diameters for this sample were limited to 3.0, 6.0, and 9.0 mm. The manufactured defects adhered to tolerances of +/- 0.3 mm in depth and +/- 0.2 mm in diameter.

The acquisition of ultrasonic data involved an unfocused linear phased array, deployed on a robotic system. The array used was

the Olympus Inspection Solutions RollerFORM-5L64 [55], featuring a central frequency of 5 MHz, 64 elements with a 0.8 mm pitch and an elevation of 6.4 mm. A 4-element sub-aperture was employed during the electronic scan, resulting in a total of 61 beams (A-scans) sampled 100 MHz rate. Each array element was powered at 80 V with a receiver gain of 22.5 dB via a PEAK NDT MicroPulse [56].

Data collection was performed using a pulse repetition frequency configured to obtain a B-scan every 0.8 mm, and the scanning speed was set at 10 mm/s. The imaging step size of 0.8 mm was chosen to align with the pitch of the employed phased array, ensuring that the captured data results in equal spatial aspect ratios. While reducing the imaging step size could theoretically lead to a more detailed and precise 3D scan, it comes with certain trade-offs. A smaller step size would result in increased data volume, leading to high correlation between individual B-scans, minimal practical improvements in capturing defective areas, an adjustment for skewed aspect ratios, and greater data storage requirements. Conversely, increasing the imaging step size beyond 0.8 mm would lead to less data being captured, potentially resulting in the oversight of smaller defects, and decreased overall scan accuracy. The scanning process was under the precise control of a fully automated robotic system based on a KUKA KR 90 R3100 extra HA industrial robot [57]. This method allowed for the concatenation of encoded B-scans to generate volumetric ultrasonic data. For this study, the errors associated with robotic movement and positioning were deemed negligible as the utilized industrial manipulator boasts a pose repeatability of ± 0.06 mm according to the ISO 9283 standard. This level of precision far exceeds the best resolution achievable with the UT setup, which is limited to 0.8 mm due to the pitch between individual ultrasonic elements.

To maintain a consistent coupling of the roller-probe's tire to the component's surface and to ensure uniform transfer of acoustic wave energy into the sample at each scan position, a force-torque sensor's real-time feedback was employed for robotic pose compensation. This mechanism regulated the distance (through contact force) between the sample's surface and the probe using feedback from the force axis perpendicular to the sample. It was executed through the integration of a Schunk GmbH & Co. FTN-GAMMA-IP65 SI-130-10 force-torque sensor, installed between the robot's flange and the roller-probe [57]. This setup targeted a constant scanning force of 70 N, preserving complete tire compression during the entire scanning process. Throughout the data acquisition process, water sprayed between the roller form's tire and the component's surface served as the acoustic couplant. This data collection setup aligns with setups in industry and has been widely adopted for the acquisition of data on large composite aerospace components [15].

B. Synthetic Data

Due to the scarcity of available experimental training data, a simulated dataset was generated for model training. This process was facilitated using CIVA, a semi-analytical physics based commercial NDE simulation software [45]. CIVA possesses the capability to accurately simulate wave propagation and its interactions with defects, a feature that has been experimentally validated for UT [46], [47]. Moreover, it offers computational efficiency when contrasted with alternative methods such FEA. Specifically, while a single ultrasonic volume was generated in less than 2 minutes using CIVA, the same task took over 60 times longer with FEA. The software allowed for comprehensive control over the simulated domain, thus enabling the modelling of defects and material properties closely resembling those found in the experimental domain. However, it is important to note that the utilization of semi-analytical software like CIVA came with certain limitations, such as restrictions to the defect geometry that can be modelled using a parametric study. Furthermore, the software, lacked the capacity to simulate responses arising from interactions within ply layers and was incapable of reproducing the noise observed in experimental data [44]. Consequently, disparities emerged between the simulated and measured experimental domains. To mitigate these disparities, a set of synthetic data generation procedures were employed to better align the synthetic and experimental domains. For further details into the distinctions between the simulated and experimental domains and the underlying rationale for the generation of more accurate synthetic data, please refer to prior work dedicated to this subject [44].

The simulation was configured with the composite material's individual layers created and employed to compute the equivalent homogeneous material properties resembling those of the experimental CFRP samples. A single ply layer was synthesized and replicated with alternating orientations of 0, 45, -45, and 90 degrees. This replication was intended to closely mimic the composition of the experimental sample. The outcome of this process was a multilayer structure, which was further homogenized to emulate a uniform medium with mechanical properties consistent with those of the multi-ply composite. To achieve this consistency, the fiber density was adjusted to 50%, matching it as closely as possible with the experimental sample's density of 1440 kg/m³.

To conduct multiple sequential simulations, a parametric study was established. In this study, the composite's bulk properties, which were previously computed, were kept constant, while the diameter and depth of defects were systematically varied. Specifically, FBH defects were simulated with diameters ranging from 3.0 mm to 15.0 mm, incremented by 0.5 mm. The depths of these defects varied from 1.5 mm to 7.0 mm from the surface, increasing in increments of 1.5 mm. In the simulations, both the reflections from the front and back walls were considered. The entire set of simulations were completed within a relatively short timeframe, requiring less than 15 hours to run on a workstation equipped with a 24-Core 3.79 GHz CPU and 128 GB of memory.

Previous research findings have underscored the limitations of relying solely on semi-analytical simulated data to faithfully replicate the experimental domain signals for model training [44]. Consequently, there arises a critical need for approaches that can effectively bridge the gap between simulated and experimental data. Fully statistical methods for introducing noise offer distinct advantages as they allow for continuous re-sampling, ensuring the generation of unique noise profiles that closely mirror real experimental data. Figure 4 provides a visual representation of the process of introducing synthetic

noise to an entire ultrasonic volume. The statistical characteristics of the noise distributions in the A-scans were derived from an independent holdout sample featuring the same layup and thickness as the test samples.

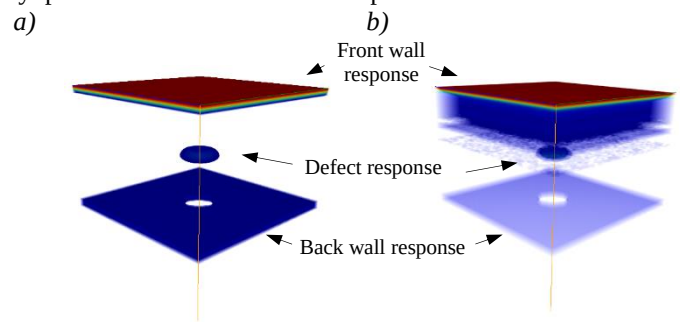


Figure 4: a) Complete ultrasonic volume of simulated A-scans for a defect response. b) the corresponding synthetically noised volume for the same defective response (opacity has been adjusted to allow for visualization of the defect through the noise). Color mapping and axes are given in Figure 5.

Table 1 provides a summarized description of the datasets created from both the experimental and synthetic data sources.

Table 1: Summary of the datasets produced.

Data source	Dataset	Number of datapoints
Simulated defect responses (300 Flat-Bottom Holes)	Synthetic defective || *Train (80%)*	240
	Synthetic defective || *Validation (20%)*	60
Experimental defect reference sample (25+15 Flat-Bottom Holes)	Sample 1 ~ Diameters: 3, 4, 6, 7, 9 mm || *Test*	25
	Sample 2 ~ Diameters: 3, 6, 9 mm || *Test*	15

C. Signal Processing

The resolution of the UT data in the array dimensions was constrained by the number (61) of acoustic beams. Padding was used to expand this dimension to 64 to amount to a power of two. To match this, 64 B-scans were selected in the scan dimension to create cuboidal datasets. The distance between the array elements was 0.8 mm, and the robotic scanning speed was regulated with the pulse repetition frequency to ensure a B-scan offset of 0.8 mm. Consequently, this approach enabled the generation of volumetric data with square voxels in the spatial domains, along the probe axis and the scanning direction. This resulted in a standardized volumetric resolution that remained consistent throughout the entire dataset.

The data obtained from both the experimental and simulated domains was in the form of radio frequency A-scans, also known as amplitude scans. To transform these signals into volumetric datasets, a series of signal processing procedures were applied. Initially, the A-scans were adjusted to have a center at zero amplitude and were then enveloped by calculating the absolute value of the Hilbert transform. This envelope is a valuable tool for determining the instantaneous response within a time series and is widely used when generating an envelope for the creation of C-scan images from ultrasonic A-scans [58]. After enveloping the signal, each volumetric dataset was normalized to fit within a range from 0 to 1. This was achieved by dividing every data point by the maximum peak amplitude present from each sample.

Following envelope calculation and normalization, corrections were applied in the time domain to align the peak front wall response with the temporal origin. This step ensured temporal alignment of features and accommodated any variations in the acoustic path length between individual transducers and the sample's surface caused by the probe's varying tire compression. For further details on the time-shifting process for an individual A-scan, in conjunction with the application of the envelope transformation, please refer to previous work which has demonstrated this [42]. Figure 5 shows the effect of this alignment on the complete ultrasonic volume.

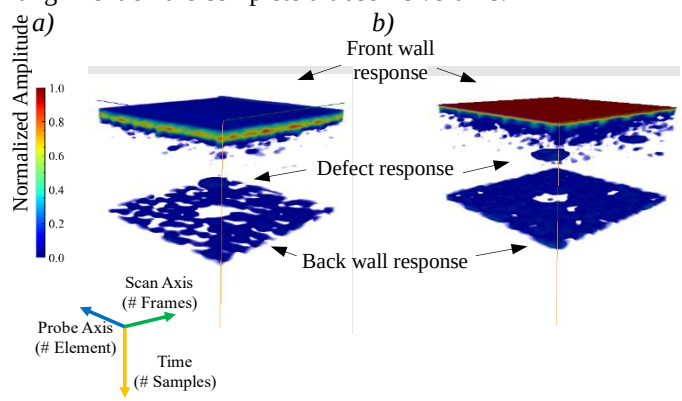


Figure 5: a) Enveloped volumetric data only. b) Volumetric data with time shifting to the central response of the front wall peak. Both figures have been thresholded to remove the lowest 10% of amplitudes to aid in visual clarity.

D. Mask Generation

An important advantage of employing simulated data as the foundation for training datasets lies in the capacity to have full control over the simulation input parameters. This control can be harnessed for the automated generation of ground truth masks during the training process, a task that would pose challenges when training models on experimental data. In the context of this study, defect diameter and depth were utilized to create segmentation masks for defects with a nominal thickness. Figure 6 provides an illustration of a ground truth defect mask alongside its corresponding simulation.

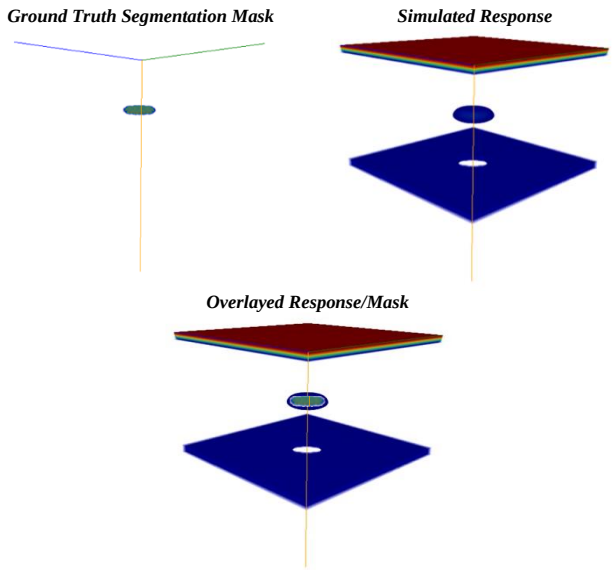


Figure 6: (a) The ground truth segmentation mask and (b) the corresponding simulated defect response. The overlay of both the mask and response is shown in (c). Color mapping and axes are given in Figure 5.

E. Augmentation

The generalizability of ML models is a critical aspect of their performance and accounts for differences in the source and target domain. Augmenting the training data improves model generalizability by adding noise at the training stage, reducing the likelihood of overfitting. This often improves performance in the target domain particularly when the target (experimental) domain is different from the source (synthetic) domain.

In this study, two domain-specific augmentation techniques that have shown their effectiveness in previous research [42] were employed during training. Standard computer vision augmentation methods (e.g. mix-up, cut mix etc.) do not translate directly to UT data as they would impact the underlying signal response. The first type of augmentation is concerned with response magnitude. Magnitude can vary due to various factors unaccounted for in the simulation, such as manufacturing variances in the sample and the UT array probe, wear on the probe and its electrical connections, and the complexities of sound wave scattering between layers. To replicate these variations while maintaining the appropriate data normalization, the amplitude of each A-scan was adjusted by a constant factor beyond the front wall. This factor was randomly selected from a uniform distribution, resulting in a scaling factor ranging from 80% to 120%.

The second augmentation method aims to replicate phase aberration - the variations in ultrasonic travel time between elements [59]. These variations can result from a range of factors, including fluctuations in the sound speed of composite materials due to their anisotropic properties, and deviations from the central frequency for specific elements. To simulate phase aberration, a 1-D interpolation technique was employed to randomly stretch or compress the signal in the time domain. The extent of dilation was randomly determined from a uniform distribution for each A-scan, allowing for dilation of up to ±300 ns.

The objective in implementing these augmentation methods is to enhance the models' ability to generalize effectively within the experimental domain. The convenience of real-time augmentations allows for their integration into the training process, eliminating the need for additional data collection or preprocessing steps. To maintain consistent data length in the time domain, each A-scan was extended by zero padding, resulting in a length of 1024 samples during training. Subsequently, to mitigate computational demands, each volume was down sampled in the time domain by a factor of 4.

III. Segmentation Methods

A. Model: Architecture and Training

In this paper, volumetric segmentation was carried out through the training and deployment of a customized 3D U-Net architecture. The design of the architecture drew inspiration from [51] but extended to five convolutional blocks with a sigmoid layer applied to the output. Models of varying number of blocks were tested, and the inclusion of an additional convolutional block resulted in a 40.8% reduction in validation loss. Further optimization of hyperparameters and architecture may result in a performance increase, however this was outside the scope of this work. A graphical representation of the overarching architectural design is presented in Figure 7.

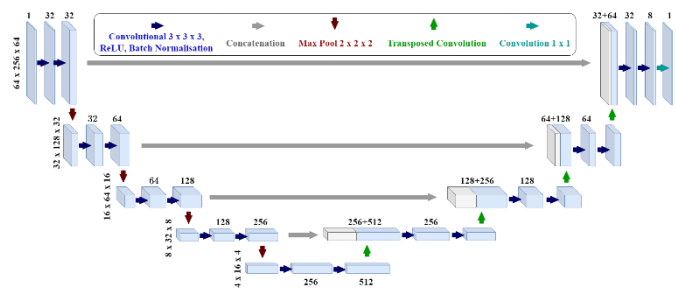


Figure 7: Architecture diagram for the 3D U-Net. Blue boxes depict the feature maps. The number of channels and dimensions of the data (probe × scan × time) are denoted above and to the side of each feature map respectively.

The model was trained on synthetic datasets with the corresponding defect masks generated from the previously defined parameter space (Figure 6). The model was trained with a batch size of 8, a patience of 3 epochs, and a learning rate of 0.003. During training, binary cross-entropy loss was employed in conjunction with an Adam optimizer [60], which was parameterized with the default parameters given in the original paper (β1 set to 0.9 and β2 to 0.999) and a learning rate of 0.003. Training and testing of the models were conducted using a desktop Windows 11 PC with Nvidia RTX 3090 Ti GPU, 128 GB RAM, and two Intel® Xeon® Gold 6428 2.50 GHz CPUs. The PyTorch 1.13.1 library and Python 3.10.8 were used for coding. Overall training took less than 20 hours. For the subsequent testing phase, the model exhibiting the lowest validation loss at the 47th epoch.

B. Reference Metric: 6 dB Drop Sizing

The 6 dB drop criterion represents the prevailing industrial methodology for defect sizing, as documented in the literature [29], [30]. The underlying principle of this method relies on the utilization of a single transducer to pinpoint the edge of a defect response by detecting the moment when exactly 50% of the

energy is reflected by the defect, corresponding to a 6 dB decrease in amplitude [27]. The technique is repeated on the opposing boundary of the defect, and the resultant displacement results in the measured defect's length (Figure 3). This fundamental principle can be further extended to encompass the sizing of defects from amplitude C-scan images produced from employing phased array transducers [28], [29], [30].

In this research, sizing in the fiber plane using the segmentations predicted by the U-Net model are compared with the established 6 dB drop. Given the prior knowledge of the manufactured defects present in the reference sample, recognized as circular FBHs, and to mitigate variations in diameters through the component thickness, the defect diameters were computed based on the maximum segmented area through the depth of the sample, based on pixel summation, using the formula outlined in equation 1.

$$Diameter = 2\sqrt{\frac{\sum_{pixels}\left(\underset{through\ depth}{\mathrm{Max}}(Segmented\ Volume)\right)}{\pi}} \quad (1)$$

IV. Results and Discussion

Figure 8 presents examples of experimental defect responses and their corresponding segmentation masks as generated from the 3D U-Net. Along with demonstrating results, these visualizations could be used by human operators to sense-check the models' predictions and reject inaccurate model predictions easily, providing the possibility for the method to contribute to a human-in-the-loop semi-autonomous NDE system.

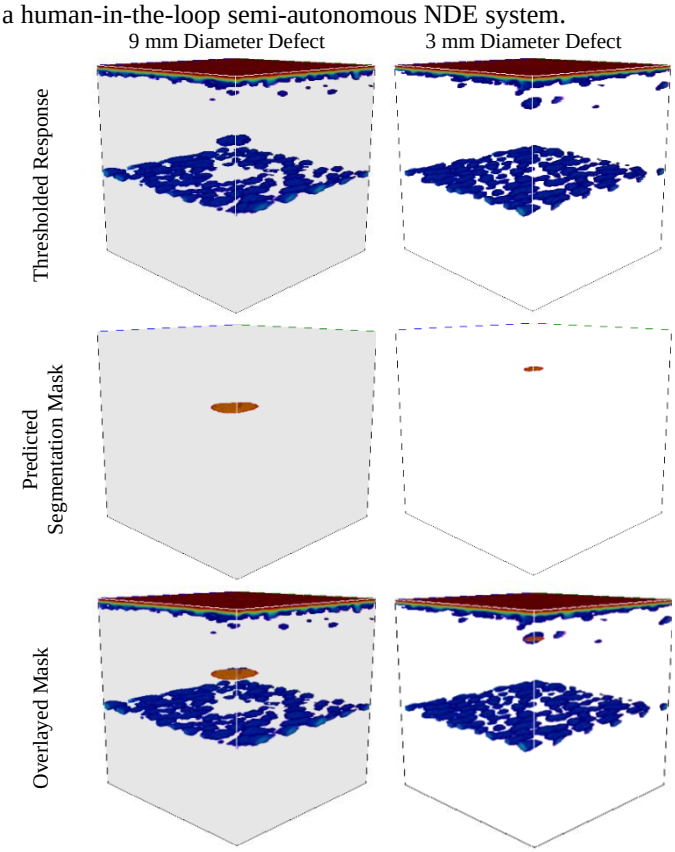


Figure 8: For two example defects their thresholded experimental ultrasonic responses (amplitudes >10% of maximum response for visualization of the defects), and their corresponding predicted segmentations. Color mapping and axes are given in Figure 5.

A. Localization

The localization of defects in 3D can be deconstructed into two primary components: in-plane localization and through-thickness depth-wise localization. It is important to note that the widely adopted 6 dB drop criterion only addresses in-plane localization and it does not provide information regarding depth-wise localization.

The 6 dB drop method can produce inaccuracies in defect sizing as discussed in section B. However, the circular shape of the test defects ensures that any errors in sizing, which might cause changes in diameter, will have minimal impact on the position of the defect's centroid. Obtaining an accurate ground truth for in-plane localization less than 1.0 mm is infeasible due to the cumulative positional errors introduced throughout the experimental setup. Therefore, this research uses the 6 dB drop criterion as a reference standard, to validate the agreement between the 6 dB and U-Net's in-plane localization. By comparison, experimental through-thickness depth measurements for defects are considerably easier to acquire, which allows for a direct assessment.

1) Depth

The segmentation of volumetric ultrasonic data offers a distinct advantage compared to the 6 dB drop method due to its capacity for depth-wise localization. This eliminates the need to employ multiple data types for characterization, such as amplitude and time-of-flight C-scans. The depth-wise position determined from the segmented volume is represented by the mean segmented depth. In Figure 9 the predicted defect depth is compared to the true measured depth of the reference defects. The segmented volumes demonstrate an excellent level of accuracy in depth-wise localization, as evidenced by a Mean Absolute Error (MAE) of 0.08 mm. This precision can be attributed to the substantially higher sampling rate in the temporal domain in contrast to the spatial domains, resulting in superior temporal resolution when compared to spatial resolution.

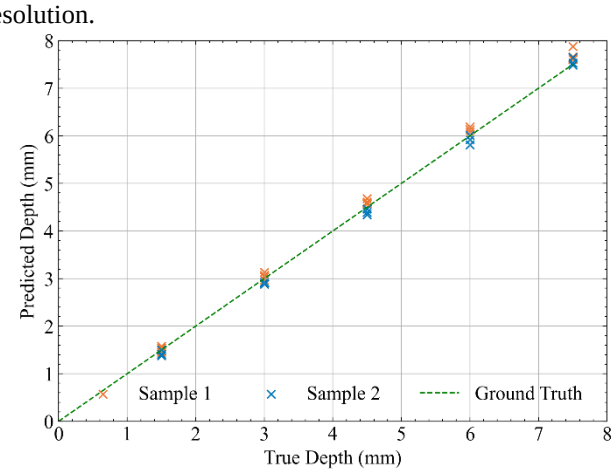


Figure 9: Depth localization results.

2) In plane

The performance of the models' in-plane localization is quantified by measuring their deviation from the centroid area compared to the 6 dB drop. Figure 10 visually presents the

centroid deviation with reference to defect sizes and the pitch of array elements. As depicted in the figure, 75% of the variations (30 out of 40 defects) are below the 0.8 mm array pitch.

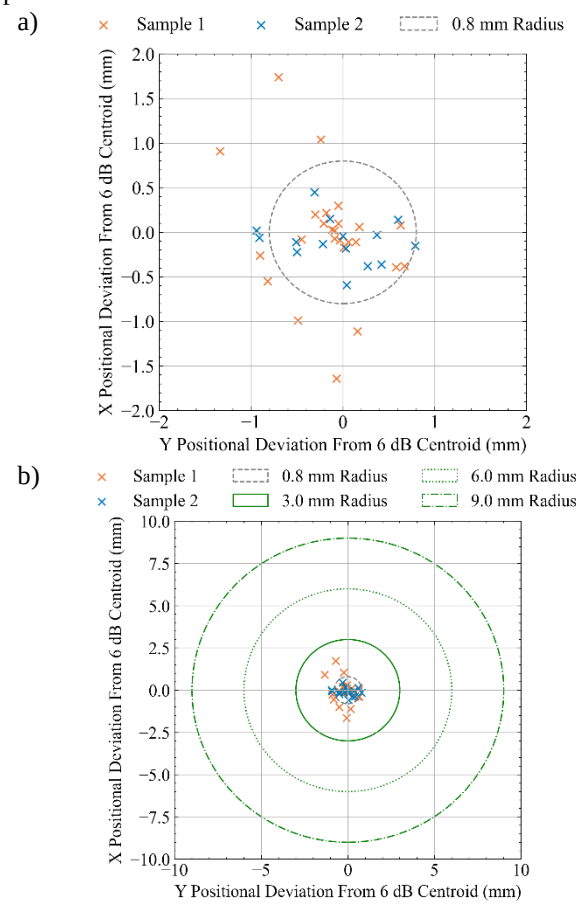


Figure 10: In plane localization results compared to 6 dB drop with reference to defect diameters (a) and the expanded (b), which shows the reference to the array pitch more clearly.

Table 2 in the appendix, provides a comprehensive overview of the in-plane localization outcomes, including the absolute distance between the 6 dB criterion and the centroid determined by the model. The MAE of 0.57 mm demonstrates a substantial concordance with the established industrial benchmark represented by the 6 dB drop criterion for in-plane localization. Notably, the MAE, being less than the 0.8 mm array pitch, which establishes the spatial resolution, underscores the robust agreement between the model-based in-plane localization and the standard reference.

B. Sizing

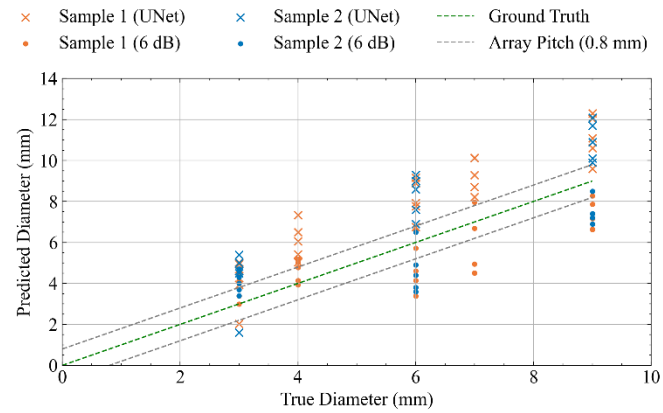


Figure 11: Sizing results for the 6 dB drop method and U-Net predictions.

Figure 11 presents a summary of the defect diameter predictions from the original segmented from the 6 dB drop and U-Net areas for each defect compared to the known ground truth area.

1) 6 dB Drop Method

With a MAE of 1.35 mm, our findings demonstrate a reasonable degree of accuracy, which, when coupled with suitable safety factors, is likely to be deemed adequate in industrial settings. Nevertheless, it is crucial to acknowledge that real-world responses often deviate from the ideal defect response, leading the 6 dB drop approach to systematically underestimate defect sizes [31].

This limitation has prompted the utilization of alternative amplitude drop methods for sizing in industry [27], [32], wherein the threshold values are frequently determined through experimental calibration. The experimental data collected for this research corroborates this tendency for under sizing defect responses (Figure 11), particularly for defects exceeding 4.0 mm in diameter. Which exhibit a mean undersize of 1.37 mm. Conversely, our results reveal a tendency to oversize defects of 3.0 and 4.0 mm in diameter.

It is essential to recognize that any sizing method relying on maximum amplitude necessitates the defect to be substantially larger than the acoustic beam to accurately ascertain the maximum acoustic response. The experimental setup employed in this work, which utilized a 4-element sub-aperture with an element pitch of 0.8 mm, resulted in an effective transducer width of 3.2 mm. Considering this in combination with the spatial resolution limitations imposed by the fixed 0.8 mm pitch for each beam step, along with accounting for any beam spread, it was determined that the experimental setup was inadequate for the precise sizing of defects measuring 4 mm or less when employing an amplitude drop method. This leads to the average oversizing of 3.0 and 4.0 mm defects of 0.82 mm, counter to the expectation of the 6 dB amplitude drop under sizing defects. The under sizing of defects larger than the diameter of the acoustic beam can likely be attributed to the curved edges of the defects. Since these defects do not maintain orthogonality to the sizing axis, the result is a diminished reflector when accounting for the three-dimensional nature of the response. This leads to a 6 dB decrease in acoustic energy closer to the center of the defect, rather than at the true defect edge. The

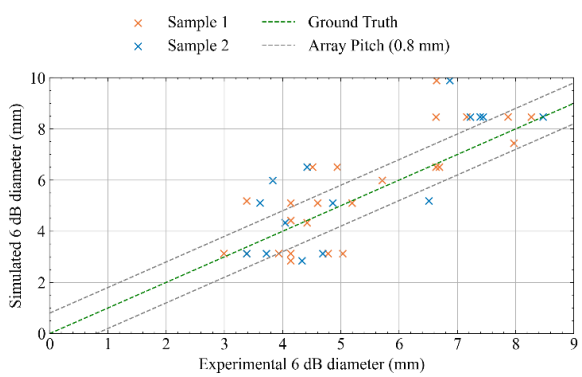
inconsistency of this method and the need for varying amplitude drop thresholds adds to the complexity of consistent defect sizing in industry and could be a concern for safety critical parts.

2) U-Net

The initial segmentation of U-Net masks yielded a MAE of 2.09 mm, which represents a 55% increase in error compared to the 6 dB drop method. As depicted in Figure 11, there was a consistent tendency to overestimate defect sizes across the range of diameters. This observation underscores the U-Net approach's reduced reliance on absolute peak amplitudes and its ability to deliver more consistent performance across a range of defect sizes, even when the defect size is not greater than the width of the acoustic beam. Moreover, it is worth noting that in numerous industrial applications for safety-critical components, it is preferable to overestimate rather than underestimate defect sizes.

It is imperative to delve into the reasons for the model's consistent trend of oversizing defects. Given that the model exhibited strong convergence during training on synthetic datasets, the overestimation observed during testing hints at a domain disparity between training on synthetic data and testing on experimental data. To elucidate this distinction, a comparative analysis using the 6 dB drop method was conducted between the responses derived from the experimental data and those generated by synthetic data for corresponding defect sizes and depths, as illustrated in Figure 12 (a). As previously noted, defects of 3 and 4 mm in diameter were too small to be accurately sized using this experimental setup and the 6 dB drop method, and thus, were excluded from this analysis (Figure 12 (b)).

a)



b)

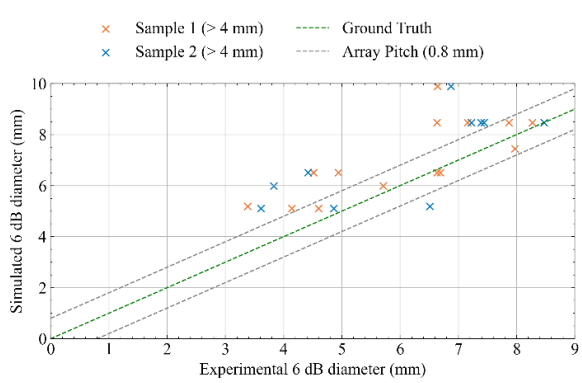


Figure 12: Comparison of sizing for synthetic and experimental data for all defect sizes (a) and defect diameters above 4 mm (b).

The comparison reveals that synthetic responses tend to yield larger defect sizes than experimental responses when employing the 6 dB drop method. Since our model's ground truth during training was based on synthetic response masks, it becomes apparent why there exists a propensity to overestimate defect sizes in our model; simulated responses tend to produce spatially larger defect responses than experimental. This disparity can be attributed to a combination of factors, CFRP acting as an acoustic collimator is likely the most impactful cause. Due to the anisotropic nature of CFRP, attenuation increases significantly as the propagation angle to normal increases [61]. This effect was not captured in the semi-analytical simulations, as it did not account for the anisotropic acoustic attenuation. However, due to the infeasible timescales of the full FEA study, the approach was deemed promising with the addition of augmentations to bridge the gap between synthetic and real domains.

Consequently, these simulations resulted in increased acoustic beam spread and larger defect responses in the simulated domain compared to the experimental domain.

To rectify this domain disparity, a constant R (given as the ratio between synthetic and experimental response diameters) was computed. The calculation for R was performed based on the disparities between sizes for the experimental and synthetic defect responses using the 6 dB drop method, as detailed Table 3 in the appendix. The resulting correction factor was determined by the mean R calculated across defects with diameters exceeding 4 mm. The resulting correction factor is 1.21 (as previously noted, defects with diameters ≤ 4 mm were deemed too small to be accurately sized with this experimental setup and the 6 dB drop method). When this correction factor is applied to the defect sizes determined from the model predictions, a greatly improved agreement with the known defect sizes is achieved, as shown in the Figure 13. This correction results in a reduction of the MAE by 58%, bringing it down to 0.88 mm. Consequently, defect sizing exceeds the accuracy of the 6 dB drop method substantially by 35%.

As the model is fully convolutional, it can scale to accept different sized inputs. For this application, the maximum defect size is limited by the inspected volume of 51.2 mm^2 in the spatial domain. While the architecture can adjust for larger volumes, this was not investigated due to computational cost and the sizes of the defects considered. Conversely, the minimum defect size is constrained by the ultrasonic setup, with the smallest defect size tested in this study being 3.0 mm in diameter. Initial trials with 1.0 mm diameter FBHs were unsuccessful due to the current measurement setup's inability to capture ultrasonic responses from them due to the probes pitch. However, by achieving good results on the 3.0 mm FBHs, it is expected that the same approach could be applied to even smaller defects given the UT setup is adjusted for that inspection scenario.

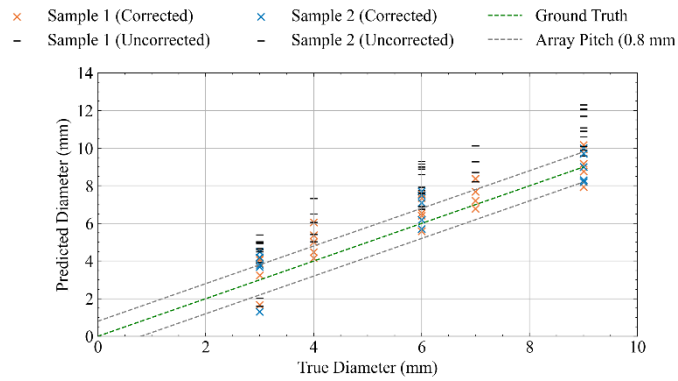


Figure 13: Corrected sizing results for the U-Net predictions.

C. Out-of-Distribution Testing

The model utilized in this study underwent supervised training, a process commonly employed in ML where the model learns patterns and relationships from labelled training data. Typically, supervised models excel in predicting examples that fall within the distribution of the training data. In this research, the training data was generated from simulated data, designed to replicate the geometry and characteristics of FBHs observed during testing. However, it's important to note that naturally occurring defects may exhibit significant variations in geometry. While defects in composite materials often manifest in-plane, their characteristics can vary widely. Training a model to generalize across such diverse conditions would likely necessitate a substantially larger and more varied training dataset.

To give an insight into the model's generalizability to different defect types and geometries, its sizing and in-plane localization performance was also evaluated on 15 square, 6 mm wide, Polytetrafluoroethylene (PTFE) inserts from a different sample. These different defect types can be considered a semantic distribution shift from the FBHs seen in training, and therefore categorized as “near Out-of-Distribution” [62]. 5 of the defects were located within 2 plies of the front wall surface, 5 were in the middle of the component and 5 were located within 2 plies of the back surface. This highlighted a limitation of the model, in that defects very near to front or backwall surfaces proved too challenging to segment. To account for this the front and back wall responses were removed. The results of this evaluation are detailed in Table 4, providing valuable insights into the model's performance for defects out of its training distribution.

Whilst this does not serve to test the wide variety of defect geometries which occur naturally, the analysis provides insights into the model's efficacy to generalize to defects seen outside of the training distribution. A MAE of 1.50 mm for defect width was observed after the correction factor was applied. This represents a 71% increase in sizing error compared to defects within the training distribution. In-plane localization performed much better with an MAE of 0.86 mm, a 51% increase in localization error however, this is still in line with the element pitch. As these inserts were embedded pre-cure it is not possible to extract a true measured ground truth of through-thickness localization as is the case for FBHs, this analysis has therefore been omitted. These results highlight a limitation of the proposed model and training regime. Synthetic data effectively addresses the challenge of acquiring labelled training data and can train an effective model for expected defects. However, for defects outside of the training distribution, there is a significant drop in performance. Furthermore, for edge cases such as defects near geometric features, additional pre-processing steps may be required; further limiting the generalizability of the model. Most DL work applied to NDE has a very specific and controlled application, and there's a notable challenge in finding models that have demonstrated effective generalizability across various materials, defect types, and component geometries. In future work the authors hope to further expand the synthetic training data to encompass a far wider range of defect types and geometries whilst also simulating different component geometries, accounting for edge cases such as near front and back wall responses. With a much larger synthetic training set it is hoped that a far more generalizable model can be trained.

V. Conclusion

This paper proposes the use of a 3D U-Net to size and localize defects in CFRP by segmenting volumetric ultrasound data. Defect sizing is a crucial piece of information for evaluating defects against standards and acceptance criteria whilst accurate localization is beneficial if re-work is required. A key benefit of the approach is that the use of volumetric ultrasound data allowed for through-thickness and in-plane localization whilst removing the requirement for gating and reducing pre-processing. Such gating and amplitude threshold selection is often performed manually by the NDE operator leading to data errors if gates and thresholds are incorrectly set.

Simulations were used to generate synthetic data and ground truth segmentation masks for training. This was a key requirement and allowed for the training of a segmentation model in a fully supervised manner. Experimentally collected ultrasonic responses from manufactured reference defects were used for testing. Sizing and in-plane localization were evaluated against the widely accepted 6 dB drop standard, and through thickness localization was compared to the measured ground truth.

The models’ depth-wise localization showed excellent results with a MAE of 0.08 mm. In-plane localization had good agreement with the accepted 6 dB drop standard with a MAE of 0.57 mm. The significant resolution differences in the spatial and temporal domains resulted in differences of error scales for in-plane and depth wise localization. This is a limitation of using a fixed pitch array, but the errors in-plane are reasonable when compared to the array pitch of 0.8 mm, which is the limiting factor for spatial resolution.

The 6 dB drop consistently undersized experimental defects greater than 4 mm in diameter whilst the U-Net produced segmentation masks that consistently oversized defects. The U-Net’s oversizing was consistent across defect sizes, showing that it was less reliant in maximum amplitude, making it more robust to sizing defects smaller than the width of the acoustic beam, which proved inconsistent for amplitude drop methods. In industrial settings other factors such as parallelism of defects to the inspection surface would also impact the maximum

signal response, introducing further inaccuracies to amplitude-based sizing methods. Upon investigation of the synthetic and experimental data domains it was evident that the experimental responses gave rise to smaller defect responses. By correcting for this disparity between the source and target domain using a single correction factor it became possible to reduce the MAE for defect sizing from 2.09 mm to 0.88 mm. The corrected defect sizing from U-Net gave a 35% reduction in MAE sizing compared to the commonly accepted 6 dB drop method. Despite this the 6 dB drop method is based on physical understanding of defect responses and whilst it has limitations the results are directly explainable. Whilst the generation of a segmentation map aides in providing some explanation for defect sizing and is more interrogatable than a regression model, it still relies on a deep learning approach which is less explainable than an amplitude-based threshold. Even so, a clear benefit of the volumetric segmentation is that it can be translated directly into a computer-aided design file which could dramatically increase the efficiency of subsequent report generation and simulation-based testing of components.

While the study demonstrated promising results in defect sizing and localization, it's essential to acknowledge a current limitation: the absence of real defects from industrial manufacturing processes for testing. Although out-of-distribution cases were examined using PTFE inserts, evaluating performance on naturally occurring defects would be advantageous. The out-of-distribution testing underscored the necessity for a more extensive distribution of training data. Future work aims to address this by expanding the training dataset, with different defect geometries, incorporating different probe frequencies, and testing on naturally occurring defects with irregular shapes. Furthermore, the objective is to integrate this research with previous work on defect detection to develop an end-to-end system for automated NDE data processing in industrial manufacturing environments.

VI. Appendix

Table 2: Complete model centroid deviation results from the 6 dB drop.

Sample 1			
Defect Diameter (mm)	**Deviation in X (mm)**	**Deviation in Y (mm)**	**Absolute Displacement (mm)**
9	-0.05	0.10	0.11
	0.58	-0.39	0.70
	-0.49	-0.99	1.10
	0.18	0.06	0.19
	0.01	-0.17	0.18
7	0.63	0.08	0.63
	-0.12	0.03	0.13
	-0.07	-1.64	1.64
	-0.18	0.22	0.29
	-0.09	0.02	0.09
6	-0.03	-0.09	0.09
	0.16	-1.11	1.12
	0.67	-0.38	0.77
	-0.24	1.04	1.06
	-0.09	-0.07	0.12
4	-1.34	0.91	1.62
	-0.45	-0.08	0.46
	-0.05	0.30	0.30
	-0.70	1.74	1.88
	-0.90	-0.26	0.94
3	-0.82	-0.55	0.98
	0.14	-0.11	0.18
	-0.21	0.10	0.23
	-0.30	0.20	0.36
	0.06	-0.11	0.13
MAE Sample 1			0.61
Sample 2			
9	0.42	-0.36	0.55
	-0.50	-0.22	0.54
	-0.94	0.02	0.94
	-0.31	0.45	0.55
	-0.14	0.15	0.21
6	0.37	-0.03	0.37
	-0.51	-0.11	0.52
	-0.91	-0.06	0.91
	0.00	-0.04	0.04
	-0.22	-0.13	0.25
3	0.27	-0.38	0.47
	0.04	-0.59	0.59
	0.79	-0.15	0.80
	0.03	-0.18	0.18
	0.60	0.14	0.61
MAE Sample 2			0.50
Total MAE			**0.57**

Table 3: Comparison in sizing of experimental and synthetic responses with the 6 dB drop method. Where R is the ratio of synthetic response to experimental response. $\overline{R}$ gives the mean average of R for a given defect diameter.

Sample 1				
Defect Diameter (mm)	**6 dB Drop Diameter**		$R = \frac{synth}{exp}$	$\overline{R}$
	Experimental Response	**Synthetic Response**		
9	6.63	8.47	1.28	1.21
	7.87	8.47	1.08	
	7.16	8.47	1.18	
	8.27	8.47	1.02	
	6.64	9.89	1.49	
7	4.51	6.51	1.44	1.22
	4.94	6.51	1.32	
	4.51	6.51	1.44	
	6.69	6.51	0.97	
	7.97	7.44	0.93	
6	3.38	5.19	1.53	1.18
	4.14	5.11	1.23	
	4.60	5.11	1.11	
	5.71	5.99	1.05	
	6.63	6.51	0.98	
4	4.78	3.13	0.65	0.82
	3.93	3.13	0.80	
	5.03	3.13	0.62	
	4.14	4.42	1.07	
	5.19	5.11	0.98	
3	3.38	3.13	0.93	0.88
	4.14	3.13	0.76	
	2.99	3.13	1.05	
	4.14	2.85	0.69	
	4.42	4.33	0.98	
Sample 2				
9	8.47	8.47	1.00	1.18
	7.22	8.47	1.17	
	7.39	8.47	1.15	
	7.44	8.47	1.14	
	6.87	9.89	1.44	
6	6.51	5.19	0.80	1.26
	4.86	5.11	1.05	
	3.61	5.11	1.41	
	3.83	5.99	1.56	
	4.42	6.51	1.47	

3	3.38	3.13	0.93	0.83
	3.72	3.13	0.84	
	4.69	3.13	0.67	
	4.33	2.85	0.66	
	4.04	4.33	1.07	

Table 4: Sizing and in-plane localization results for out of distribution test defects.

	Width (mm)				Deviation in X (mm)	Deviation in Y (mm)	Absolute Displacement (mm)
True	**Predicted**	**Predicted Error**	**Corrected**	**Corrected Error**			
6	5.18	-0.82	4.28	-1.72	-0.18	-1.6	1.61
	6.69	0.69	5.53	-0.47	-0.24	-0.14	0.28
	8.98	2.98	7.42	1.42	0.08	0.47	0.48
	8.98	2.98	7.42	1.42	0.06	-0.45	0.45
	6.65	0.65	5.49	-0.51	-0.09	-0.18	0.20
	6.60	0.60	5.45	-0.55	-0.15	-0.21	0.26
	1.96	-4.04	1.62	-4.38	2.22	1.86	2.90
	7.11	1.11	5.88	-0.12	-0.44	0.07	0.45
	8.80	2.80	7.27	1.27	-0.03	0.51	0.51
	4.23	-1.77	3.50	-2.50	-0.42	-1.23	1.30
	5.99	-0.01	4.95	-1.05	-1.44	-0.69	1.60
	3.10	-2.90	2.56	-3.44	-0.82	-0.6	1.02
	8.80	2.80	7.27	1.27	0.09	0.1	0.13
	6.50	0.50	5.37	-0.63	-0.5	-0.07	0.50
	5.06	-0.94	4.18	-1.82	-1.04	-0.85	1.34
MAE		**1.71**		**1.50**			**0.86**

VII. References

[1] J. Zhang, G. Lin, U. Vaidya, and H. Wang, 'Past, present and future prospective of global carbon fibre composite developments and applications', *Composites Part B: Engineering*, vol. 250, p. 110463, Feb. 2023, doi: 10.1016/j.compositesb.2022.110463.

[2] V. Giurgiutiu, 'Chapter 1 - Introduction', in *Structural Health Monitoring of Aerospace Composites*, V. Giurgiutiu, Ed., Oxford: Academic Press, 2016, pp. 1–23. doi: 10.1016/B978-0-12-409605-9.00001-5.

[3] B. Djordjevic, 'Non Destructive Test Technology for the Composite', p. 7, Jan. 2009.

[4] C. Meola, S. Boccardi, G. M. Carlomagno, N. D. Boffa, E. Monaco, and F. Ricci, 'Nondestructive evaluation of carbon fibre reinforced composites with infrared thermography and ultrasonics', *Composite Structures*, vol. 134, pp. 845–853, Dec. 2015, doi: 10.1016/j.compstruct.2015.08.119.

[5] A. M.-E. Dorado, 'Composite Material Characterization using Acoustic Wave Speed Measurements', p. 5.

[6] Ley, O. and V. Godinez, 'Non-destructive evaluation (NDE) of aerospace composites: application of infrared (IR) thermography', doi: 10.1533/9780857093554.3.309.

[7] A. Kokurov and D. Subbotin, 'Ultrasonic detection of manufacturing defects in multilayer composite structures', *IOP Conference Series: Materials Science and Engineering*, vol. 1023, p. 012013, Jan. 2021, doi: 10.1088/1757-899X/1023/1/012013.

[8] D. K. Hsu, '15 - Non-destructive evaluation (NDE) of aerospace composites: ultrasonic techniques', in *Non-Destructive Evaluation (NDE) of Polymer Matrix Composites*, V. M. Karbhari, Ed., in Woodhead Publishing Series in Composites Science and Engineering. , Woodhead Publishing, 2013, pp. 397–422. doi: 10.1533/9780857093554.3.397.

[9] F. Heinecke and C. Willberg, 'Manufacturing-Induced Imperfections in Composite Parts Manufactured via Automated Fiber Placement', *J. Compos. Sci.*, vol. 3, no. 2, p. 56, Jun. 2019, doi: 10.3390/jcs3020056.

[10] I. Papa, V. Lopresto, and A. Langella, 'Ultrasonic inspection of composites materials: Application to detect impact damage', *International Journal of Lightweight Materials and Manufacture*, vol. 4, no. 1, pp. 37–42, Mar. 2021, doi: 10.1016/j.ijlmm.2020.04.002.

[11] L. Séguin-Charbonneau, J. Walter, L.-D. Théroux, L. Scheed, A. Beausoleil, and B. Masson, 'Automated defect detection for ultrasonic inspection of CFRP aircraft components', *NDT & E International*, vol. 122, p. 102478, Sep. 2021, doi: 10.1016/j.ndteint.2021.102478.

[12] S. Gholizadeh, 'A review of non-destructive testing methods of composite materials', *Procedia Structural Integrity*, vol. 1, pp. 50–57, 2016, doi: 10.1016/j.prostr.2016.02.008.

[13] M. Jolly *et al.*, 'Review of Non-destructive Testing (NDT) Techniques and their Applicability to Thick Walled Composites', *Procedia CIRP*, vol. 38, pp. 129–136, Jan. 2015, doi: 10.1016/j.procir.2015.07.043.

[14] E. Duernberger, C. MacLeod, D. Lines, C. Loukas, and M. Vasilev, 'Adaptive optimisation of multi-aperture ultrasonic phased array imaging for increased inspection speeds of wind turbine blade composite panels', *NDT & E International*, vol. 132, p. 102725, Dec. 2022, doi: 10.1016/j.ndteint.2022.102725.

[15] C. Mineo *et al.*, 'Flexible integration of robotics, ultrasonics and metrology for the inspection of aerospace components', *AIP Conference Proceedings*, vol. 1806, no. 1, p. 020026, Feb. 2017, doi: 10.1063/1.4974567.

[16] S. Maack, V. Salvador, and S. David, 'Validation of artificial defects for Non-destructive testing measurements on a reference structure', *MATEC Web of Conferences*, vol. 199, p. 06006, Jan. 2018, doi: 10.1051/matecconf/201819906006.

[17] P. Gardner *et al.*, 'Machine learning at the interface of structural health monitoring and non-destructive evaluation', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 378, no. 2182, p. 20190581, Oct. 2020, doi: 10.1098/rsta.2019.0581.

[18] C. Mineo *et al.*, 'Robotic Geometric and Volumetric Inspection of High Value and Large Scale Aircraft Wings', in *2019 IEEE 5th International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, Jun. 2019, pp. 82–86. doi: 10.1109/MetroAeroSpace.2019.8869667.

[19] 'Introduction to non-destructive testing', Aerospace Testing International. Accessed: Nov. 17, 2021. [Online]. Available: https://www.aerospacetestinginternational.com/features/introduction-to-non-destructive-testing.html

[20] F. W. Margrave, K. Rigas, D. A. Bradley, and P. Barrowcliffe, 'The use of neural networks in ultrasonic flaw detection', *Measurement*, vol. 25, no. 2, pp. 143–154, Mar. 1999, doi: 10.1016/S0263-2241(98)00075-X.

[21] J. Ye, S. Ito, and N. Toyama, 'Computerized Ultrasonic Imaging Inspection: From Shallow to Deep Learning', *Sensors (Basel)*, vol. 18, no. 11, p. 3820, Nov. 2018, doi: 10.3390/s18113820.

[22] B. Valeske, A. Osman, F. Römer, and R. Tschuncky, 'Next Generation NDE Sensor Systems as IIoT Elements of Industry 4.0', *Research in Nondestructive Evaluation*, vol. 31, no. 5–6, pp. 340–369, Nov. 2020, doi: 10.1080/09349847.2020.1841862.

[23] S. Cantero-Chinchilla, P. D. Wilcox, and A. J. Croxford, 'Deep learning in automated ultrasonic NDE -- developments, axioms and opportunities', *arXiv:2112.06650 [eess]*, Dec. 2021, Accessed: Jan. 12, 2022. [Online]. Available: http://arxiv.org/abs/2112.06650

[24] M. Gower, G. Sims, R. Lee, S. Frost, M. Stone, and M. Wall, 'Measurement Good Practice Guide', no. 78.

[25] Y. Wang, F. Tao, Y. Zuo, M. Zhang, and Q. Qi, 'Digital-Twin-Enhanced Quality Prediction for the Composite Materials', *Engineering*, vol. 22, pp. 23–33, Mar. 2023, doi: 10.1016/j.eng.2022.08.019.

[26] D. G. Puttaraju and H. G. Hanumantharaju, 'Finite element analysis and validation of tensile properties of carbon fiber reinforced polymer matrix composites', *Materials Today: Proceedings*, vol. 62, pp. 2800–2807, Jan. 2022, doi: 10.1016/j.matpr.2022.02.188.

[27] M. V. Felice and Z. Fan, 'Sizing of flaws using ultrasonic bulk wave testing: A review', *Ultrasonics*, vol. 88, pp. 26–42, Aug. 2018, doi: 10.1016/j.ultras.2018.03.003.

[28] A. Hauffe, F. Hähnel, and K. Wolf, 'Comparison of algorithms to quantify the damaged area in CFRP ultrasonic scans', *Composite Structures*, vol. 235, p. 111791, Mar. 2020, doi: 10.1016/j.compstruct.2019.111791.

[29] S. Barut, V. Bissauge, G. Ithurralde, and W. Claassens, 'Computer-aided analysis of ultrasound data to speed-up the release of aerospace CFRP components', *e-Journal of Nondestructive Testing*, vol. 17, no. 07, Jul. 2012, [Online]. Available: https://www.ndt.net/search/docs.php3?id=12429&msgID=0&rootID=0

[30] S. Barut and N. Dominguez, 'NDT Diagnosis Automation: a Key to Efficient Production in the Aeronautic Industry', *e-Journal of Nondestructive Testing*, vol. 21, no. 07, Jul. 2016, [Online]. Available: https://www.ndt.net/search/docs.php3?id=19184&msgID=0&rootID=0

[31] S. Kumaran and S. Rani, 'Application of 6db Drop Technique to Estimate the Width of Sub Assembly Ring Top Using Pulse Echo Ultrasonic Technique', *International Journal of Engineering and Technology*, vol. 5, pp. 4771–4775, Jan. 2013.

[32] P. Ciorau, 'Comparison Between -6 DB and -12 DB Amplitude Drop Techniques for Length Sizing'.

[33] X. Li, Y. Wang, P. Ni, H. Hu, and Y. Song, 'Flaw sizing using ultrasonic C-scan imaging with dynamic thresholds', *insight*, vol. 59, no. 11, pp. 603–608, Nov. 2017, doi: 10.1784/insi.2017.59.11.603.

[34] S. Niu and V. Srivastava, 'Simulation trained CNN for accurate embedded crack length, location, and orientation prediction from ultrasound measurements', *International Journal of Solids and Structures*, vol. 242, p. 111521, May 2022, doi: 10.1016/j.ijsolstr.2022.111521.

[35] R. J. Pyle, R. L. T. Bevan, R. R. Hughes, R. K. Rachev, A. A. S. Ali, and P. D. Wilcox, 'Deep Learning for Ultrasonic Crack Characterization in NDE', *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 68, no. 5, pp. 1854–1865, May 2021, doi: 10.1109/TUFFC.2020.3045847.

[36] T. Lardner, G. West, G. Dobie, and A. Gachagan, 'Automated sizing and classification of defects in CANDU pressure tubes', *Nuclear Engineering and Design*, vol. 325, pp. 25–32, Dec. 2017, doi: 10.1016/j.nucengdes.2017.09.029.

[37] X. Cheng, G. Ma, Z. Wu, H. Zu, and X. Hu, 'Automatic defect depth estimation for ultrasonic testing in carbon fiber reinforced composites using deep learning', *NDT & E International*, vol. 135, p. 102804, Apr. 2023, doi: 10.1016/j.ndteint.2023.102804.

[38] I. Virkkunen, T. Koskinen, O. Jessen-Juhler, and J. Rinta-aho, 'Augmented Ultrasonic Data for Machine Learning', *J Nondestruct Eval*, vol. 40, no. 1, p. 4, Jan. 2021, doi: 10.1007/s10921-020-00739-5.

[39] P. M. Cheng and H. S. Malhi, 'Transfer Learning with Convolutional Neural Networks for Classification of Abdominal Ultrasound Images', *J Digit Imaging*, vol. 30, no. 2, pp. 234–243, Apr. 2017, doi: 10.1007/s10278-016-9929-2.

[40] Y. Zhou *et al.*, 'Multi-task learning for segmentation and classification of tumors in 3D automated breast ultrasound images', *Medical Image Analysis*, vol. 70, p. 101918, May 2021, doi: 10.1016/j.media.2020.101918.

[41] Y. Liu, '3D Image Segmentation of MRI Prostate Based on a Pytorch Implementation of V-Net', *J. Phys.: Conf. Ser.*, vol. 1549, no. 4, p. 042074, Jun. 2020, doi: 10.1088/1742-6596/1549/4/042074.

[42] S. McKnight *et al.*, '3-Dimensional Residual Neural Architecture Search for Ultrasonic Defect Detection', *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, pp. 1–1, 2024, doi: 10.1109/TUFFC.2024.3353408.

[43] A. Figueira and B. Vaz, 'Survey on Synthetic Data Generation, Evaluation Methods and GANs', *Mathematics*, vol. 10, no. 15, Art. no. 15, Jan. 2022, doi: 10.3390/math10152733.

[44] S. McKnight *et al.*, 'A comparison of methods for generating synthetic training data for domain adaption of deep learning models in ultrasonic non-destructive evaluation', *NDT & E International*, vol. 141, p. 102978, Jan. 2024, doi: 10.1016/j.ndteint.2023.102978.

[45] 'EXTENDE, Experts in Non Destructive Testing Simulation with CIVA Software'. Accessed: Nov. 07, 2022. [Online]. Available: https://www.extende.com/

[46] S. Lonné, L. D. Roumilly, L. L. Ber, S. Mahaut, and G. Cattiaux, 'EXPERIMENTAL VALIDATION OF CIVA ULTRASONIC SIMULATIONS', 2006, [Online]. Available: https://www.semanticscholar.org/paper/EXPERIMENTAL-VALIDATION-OF-CIVA-ULTRASONIC-Lonn%C3%A9-Roumilly/16b85af3b6a96d4657c9902ca8652fbbf93cbf2e

[47] M. Darmon *et al.*, 'VALIDATION OF AN ULTRASONIC CHARACTERIZATION TECHNIQUE FOR ANISOTROPIC MATERIALS: COMPARISON OF EXPERIMENTS WITH BEAM PROPAGATION MODELLING', presented at the 2019 INTERNATIONAL CONGRESS ON ULTRASONICS, Bruges Belgium.

[48] K. Jezzine, D. Segur, R. Ecault, and N. Dominguez, 'Simulation of ultrasonic inspections of composite structures in the CIVA software platform', *e-Journal of Nondestructive Testing*, vol. 21, no. 07, Jul. 2016, [Online]. Available: https://www.ndt.net/search/docs.php3?id=19438&msgID=0&rootID=0

[49] O. Ronneberger, P. Fischer, and T. Brox, 'U-Net: Convolutional Networks for Biomedical Image Segmentation', in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., in Lecture Notes in Computer Science. Cham: Springer International Publishing, 2015, pp. 234–241. doi: 10.1007/978-3-319-24574-4_28.

[50] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen, and K. H. Maier-Hein, 'nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation', *Nat Methods*, vol. 18, no. 2, Art. no. 2, Feb. 2021, doi: 10.1038/s41592-020-01008-z.

[51] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, '3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation'. arXiv, Jun. 21, 2016. doi: 10.48550/arXiv.1606.06650.

[52] D. Medak, L. Posilovic, M. Subasic, M. Budimir, and S. Loncaric, 'Automated Defect Detection From Ultrasonic Images Using Deep Learning', *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 68, no. 10, pp. 3126–3134, Oct. 2021, doi: 10.1109/TUFFC.2021.3081750.

[53] Y. Fu and X. Yao, 'A review on manufacturing defects and their detection of fiber reinforced resin matrix composites', *Composites Part C: Open Access*, vol. 8, p. 100276, Jul. 2022, doi: 10.1016/j.jcomc.2022.100276.

[54] P. Blain *et al.*, 'Artificial defects in CFRP composite structure for thermography and shearography nondestructive inspection', presented at the Fifth International Conference on Optical and Photonics Engineering, A. K. Asundi, Ed., Singapore, Singapore, Jun. 2017, p. 104493H. doi: 10.1117/12.2271701.

[55] 'RollerFORM: Phased Array Wheel Probe'. Accessed: Jun. 06, 2023. [Online]. Available: https://www.olympus-ims.com/en/rollerform/

[56] 'Ultrasonic Inspection Solutions for NDT | Peak NDT'. Accessed: Feb. 28, 2024. [Online]. Available: https://www.peakndt.com/

[57] M. Vasilev *et al.*, 'Sensor-Enabled Multi-Robot System for Automated Welding and In-Process Ultrasonic NDE', *Sensors*, vol. 21, no. 15, Art. no. 15, Jan. 2021, doi: 10.3390/s21155077.

[58] R. Drai, F. Sellidj, M. Khelil, and A. Benchaala, 'Elaboration of some signal processing algorithms in ultrasonic techniques: application to materials NDT', *Ultrasonics*, vol. 38, no. 1, pp. 503–507, Mar. 2000, doi: 10.1016/S0041-624X(99)00082-7.

[59] M. Sharifzadeh, H. Benali, and H. Rivaz, 'Phase Aberration Correction: A Convolutional Neural Network Approach', *IEEE Access*, vol. 8, pp. 162252–162260, 2020, doi: 10.1109/ACCESS.2020.3021685.

[60] D. P. Kingma and J. Ba, 'Adam: A Method for Stochastic Optimization'. arXiv, Jan. 29, 2017. doi: 10.48550/arXiv.1412.6980.

[61] K. Ono and A. Gallego, 'Attenuation of lamb waves in CFRP plates', *Journal of Acoustic Emission*, vol. 30, pp. 109–124, Jan. 2012.

[62] J. Yang, K. Zhou, and Z. Liu, 'Full-Spectrum Out-of-Distribution Detection', *Int J Comput Vis*, vol. 131, no. 10, pp. 2607–2622, Oct. 2023, doi: 10.1007/s11263-023-01811-z.