DOI: 10.61782/fa.2025.0260

DEEP LEARNING FOR ULTRASOUND SURFACE ECHO DETECTION WITH MATRIX ARRAYS

M. Larrea[1*] **M. Muñoz**[1,2] **A. Rubio**[1,2]
J.F. Cruza[1] **G.Cosarinsky**[1]
[1] Ultrasound Systems and Technology Group (GSTU), Institute for Physical and Information Technologies (ITEFI), Spanish National Research Council (CSIC),c/Serrano 144, 28006 Madrid, Spain
[2] Electronics Department, Escuela Politécnica, Universidad de Alcalá de Henares, Ctra.Madrid-Barcelona, Km. 33,600, 28805 Madrid, Spain

ABSTRACT

In Non-Destructive Testing (NDT), imaging with array probes is frequently performed in a two propagation medium scenario, where the first acts as a coupling medium, and the second is the object under test. Therefore, the position and shape of the refracting interface between the two mediums must be known in order to compute the imaging focal laws. This geometrical information can be inferred from the arrival times of echoes reflected on the object's surface, which must be identified within the signals. A common approach for detecting these surface echoes is the first-threshold crossing method; however, it is susceptible to outliers, and more robust alternatives are needed. In a recent work, we developed and trained a 3D Convolutional Neural Network (DeepEcho3D) using an 11x11 matrix array to accurately and reliably detect surface echoes. While effective, the model is tailored to the specific array used. In this study, we explore extending DeepEcho3D to other matrix arrays with varying shapes and frequencies. Our findings demonstrate that it is unnecessary to start from scratch for each new array; instead, we can leverage transfer learning from the initial model.

Keywords: *Time of Flight (TOF), Surface Detection, Non-Destructive Testing (NDT), Convolutional Neural Network (CNN), Probe Location and Orientation (PLO), Deep Learning, Fine-Tuning*

**Corresponding author*: *marcelo.larrea@csic.es.*

1. INTRODUCTION

Ultrasonic Testing (UT) is a widely used Non-Destructive Testing (NDT) technique for internal imaging of industrial and structural components. It relies on proper coupling between the transducer and the test component, often achieved through immersion tests, that is, using water as coupling media.

The use of array probes [1] allows various imaging modes, including conventional Phased Array (PA), as well as more advanced techniques such as TFM (Total Focusing Method) [2] and PWI (Plane Wave Imaging) [3, 4], which offer improved resolution and versatility.

Inspection using the immersion technique presents difficulties in accurately calculating focal laws due to refracted rays. This, in turn, requires in-depth knowledge of both the test component surface and the probe location and orientation (PLO) relative to it.

The mentioned challenge has been addressed by adaptive imaging algorithms, which utilize ultrasonic signals to estimate the geometry before generating the final image, also referred to as auto-focusing techniques [5–7]. Among the various proposed algorithms for computing surface points, a subset of them relies on accurately detecting the time of flight (TOF) of surface echoes, making a robust TOF measurement method essential.

There are various methods for TOF detection [8, 9], being the simplest one the threshold crossing technique. Nevertheless, despite its efficiency, spurious echoes, as those from water bubbles or probe surface waves, can cause measurement errors. Improving these methods is essential, as TOF measurements is widely used not only in UT but also in numerous other acoustic signals applications.

Given the strong capabilities of machine learning and deep learning for data classification and image segmentation, as well as their successful application in NDT research [10, 11], we leveraged this potential in a recent study to address the aforementioned measurement issues. The designed network (DeepEcho3D) was trained using data acquired from a 3 MHz, 1 mm pitch, 11×11 matrix array and a set of simple yet diverse surfaces [12]. The model demonstrated strong performance in detecting surface echoes across a sufficiently distinct set of test geometries, evaluated using two customized metrics: the percentage of outliers and a magnitude of error relative to the ground truth for the acquired data.

This study explores the adaptation of the DeepEcho3D model to a different array configuration using transfer learning and fine-tuning techniques. The model was adapted to a 5 MHz, 0.9 mm, 8×16 matrix array. The adaptation process consisted of three key stages: **(a)** transferring the pre-trained model to the new array dimensions, **(b)** fine-tuning selected layers to improve performance, and **(c)** evaluating the model's stability across multiple training runs.

2. METHODS

The adaptation process involved loading the pre-trained DeepEcho3D model, modifying its architecture to accommodate the new input shape (8×16), and transferring the original model's weights to the corresponding layers in the new network. To fine-tuning the transferred network, the dataset was obtained from a plane surface. The purpose is to use a single component instead of a set of different shapes, as in [12]. And finally, performance evaluation was conducted using customized metrics originally defined for the original net.

2.1 Data Acquisition

Data was generated through the Full Matrix Capture (FMC) [1] technique across 178 PLOs. For each PLO, a partial FMC was obtained using 9 elements as emitters, and all 128 elements as echo receivers, resulting in a total of 1602 (178 × 9) images. These images were then randomly divided into training (60%), testing (20%), and validation (20%) subsets

Essential for the network retraining stage, image labeling was automated during the post-processing of the acquired data. The Ground Truth (GT) is a binary mask, $GT(i, j)$, being i the time index and j the array element index. The theoretical TOF of the surface echo has time index $Idx_{true}(j)$, such that:

$$GT(i,j) = \begin{cases} 0, & i > Idx_{true}(j) \\ 1, & i < Idx_{true}(j) \end{cases} \tag{1}$$

$Idx_{true(j)}$ is computed using the law of specular reflection, as described in [13]. Fig. 1 shows an example of Idx_{true} in its upper graph (green line).

Three extra sets of data, two convex cylindrical surfaces (35 and 12 mm diameter), and one concave-shaped cylinder (40 mm diameter), were further acquired to assess the fine-tuned models performance under different geometries.

2.2 Experimental Setup

2.2.1 *Preliminary Testing*

The first dataset came from a plane surface, providing an initial test for the transferred model. The initial assessment revealed that, although the model generalized well, certain segmentation inaccuracies persisted. Fig. 1 and Fig. 2, illustrate, respectively, **1)** an example of satisfactory performance, where the model correctly segmented the image (lower plot) based on the ground truth shown in the upper graph; and **2)** a challenging case for the new model, with imprecise segmentation. Both images correspond to a single emitter element, and what is seen is the 2D representation of the 3D array acquisition (upper plot), and the transferred model segmentation output (lower plot). Now, to mitigate these segmentation inaccuracies, a fine-tuning strategy was implemented, retraining specific layers on the new dataset.

2.2.2 *Section-Wise Exploration*

DeepEcho3D consists of an *autoencoder* architecture, which can be divided into three main sections: (1) the encoder, which extracts hierarchical features and compresses input data into a lower-dimensional representation. It consists of two internal blocks, each containing

[1] FMC: Technique where each element in an ultrasound array transmits a pulse one at a time while all elements in the array record the returning echoes. This process is repeated for each element in the array.

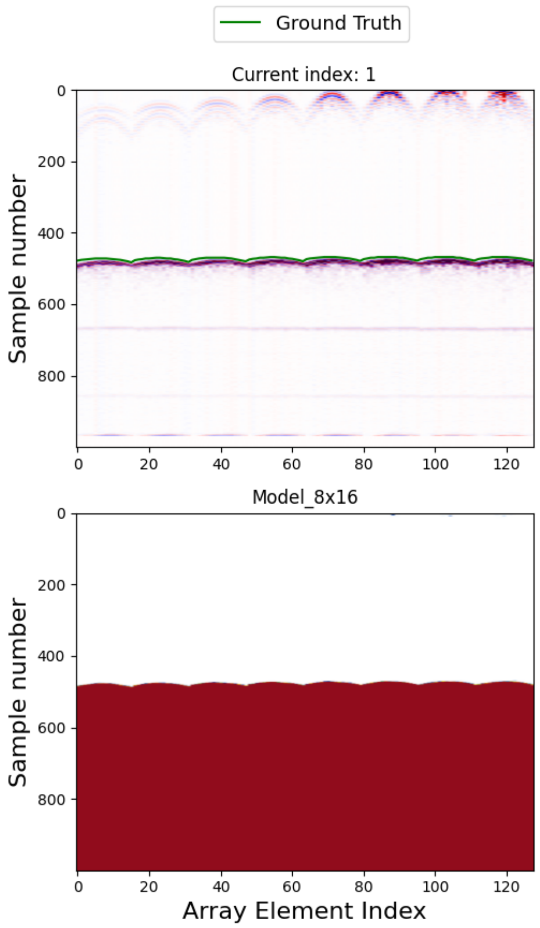


Figure 1: Instance of acceptable performance for initial model adaptation

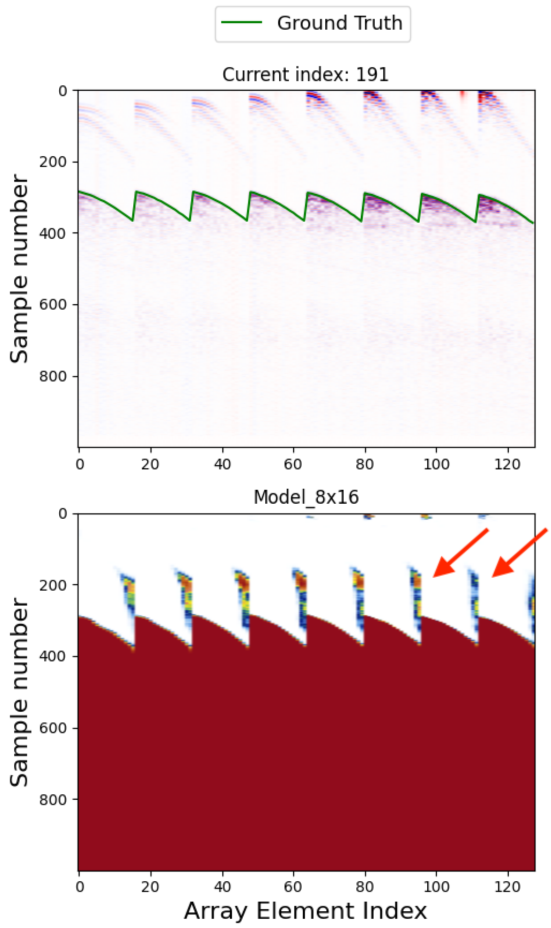


Figure 2: Instance of challenging case for initial model adaptation. Red arrows highlight the regions where the network's segmentation is inaccurate.

two convolutional layers. (2) The bottleneck, which incorporates a reduced set of features representing the most important information from the input; and (3) the decoder, that reconstructs the original data from the latent space, aiming to preserve essential spatial and feature details.

The exploration was structured into the network's three main sections: encoder, bottleneck, and decoder, with fine-tuning applied only to convolutional layers, excluding batch normalization and other layers. Additionally, the two internal encoder blocks were added as supplementary configurations to be explored. This led to five different fine-tuning setups. **1)** Encoder (4 convolutional layers); **2)** Bottleneck (2 convolutional layers); **3)** Decoder (5 convolutional layers); **4)** Encoder 1st block (2 convolutional layers); and **5)** Encoder 2nd block (2 convolutional layers)

Convolutional layers are primarily responsible for feature extraction and generalization across different datasets [14–16], which motivated the chosen approach.

2.2.3 *Fine-Tuning Implementation*

During the fine-tuning process, the following optimized hyperparameters were used (see Table 2), originally selected as the best-performing ones for the 11×11 model [12]. This approach aimed to retain the effectiveness of the original model while allowing it to generalize to different array configurations.

Table 1: DeepEcho3D optimized hyperparameters.

Hyperparameter	**Value**
Number of filters	16
Number of layers	12
Convolutional kernel size	(3, 3, 12)
Pooling size	(1, 1, 8)
Epochs	10
Batch size	16
Learning rate	0.001
Steps per epoch	61
Loss function	Binary cross-entropy

2.3 Evaluation

Regarding the metrics used to assess the performance of the resulting models, two customized evaluation metrics were employed, defined as follows.

Index Error: This metric is defined as the maximum of the absolute value of the numerical difference between the predicted indexes and the true indexes, over the entire set of array elements [2], see Eqn. (2). Thus, for each A-scan within an FMC acquisition, a theoretical TOF index is determined and compared against the index predicted by the model.

$$Idx_{MaxError} = \max(|Idx_{pred} - Idx_{true}|) \quad (2)$$

Outlier Rate: This metric is particularly important for evaluating models reliability in practical applications. It is based on the computed error of indexes ((3))

$$Idx_{err} = |Idx_{pred} - Idx_{true}| \quad (3)$$

It measures how frequently the Idx_{err} exceeds a predefined threshold. The total count of such occurrences, relative to the dataset size (N), defines the outlier percentage.

[2] For instance, with 989 test images, this would correspond to $N = 989 \times 128 = 126592$ index-error values

Finally, given the stochastic nature of deep learning optimization, a stability assessment was carried out on each of the fine-tuned configurations, via multiple runs. Specifically, 10 runs per model were performed, and both evaluation metrics were computed for each, aiming to identify the configuration with the least deviation from its previous results. Therefore, the ideal models are those that exhibit minimal dispersion and remain close to the origin in the metric space.

Performance was assessed using data acquired from the three previously mentioned cylindrical components. These geometries were previously included in the test set for the original 11×11 model, making them suitable benchmarks to evaluate fine-tuned models. In particular, the concave-shaped 40 mm cylinder posed the greatest challenge, as it represents a complex case for the segmentation process.

3. RESULTS

Fig. 3 presents a comprehensive comparative analysis, not only between the fine-tuned models and the transferred network (Model_8x16) but also among themselves. It shows that the selective retraining of each set of convolutional layers enhances the capabilities, to a greater or lesser extent, of the transferred model, which has 14.43% outliers.

The most significant improvement correspond to the first block of the encoder section, highlighted in Fig. 3 (with 2.33% of outliers), followed by the full encoder (5.53 %); the bottleneck (6.56 %); the decoder (8.13 %); and finally the second block of the encoder (8.37 %).

Fig. 4 illustrates how fine-tuning enhances performance in a challenging case where the initial model encounters difficulties with image segmentation, where the test image exhibits low-amplitude surface echoes.

With the model using only the transferred weights, it fails to accurately identify and segment these surface echoes (Fig. 4a). However, in the encoder's first-block fine-tuned version, the model demonstrates a notable ability to capture very low-amplitude surface echoes while effectively suppressing superficial waves generated by the transducer (Fig. 4b). This highlights the efficacy of fine-tuning in improving the model's image segmentation performance.

Fig. 5 represents the results of the stability analysis described earlier. As observed, this analysis confirms that the first block of the encoder is the most consistent and stable configuration, exhibiting the least dispersion and

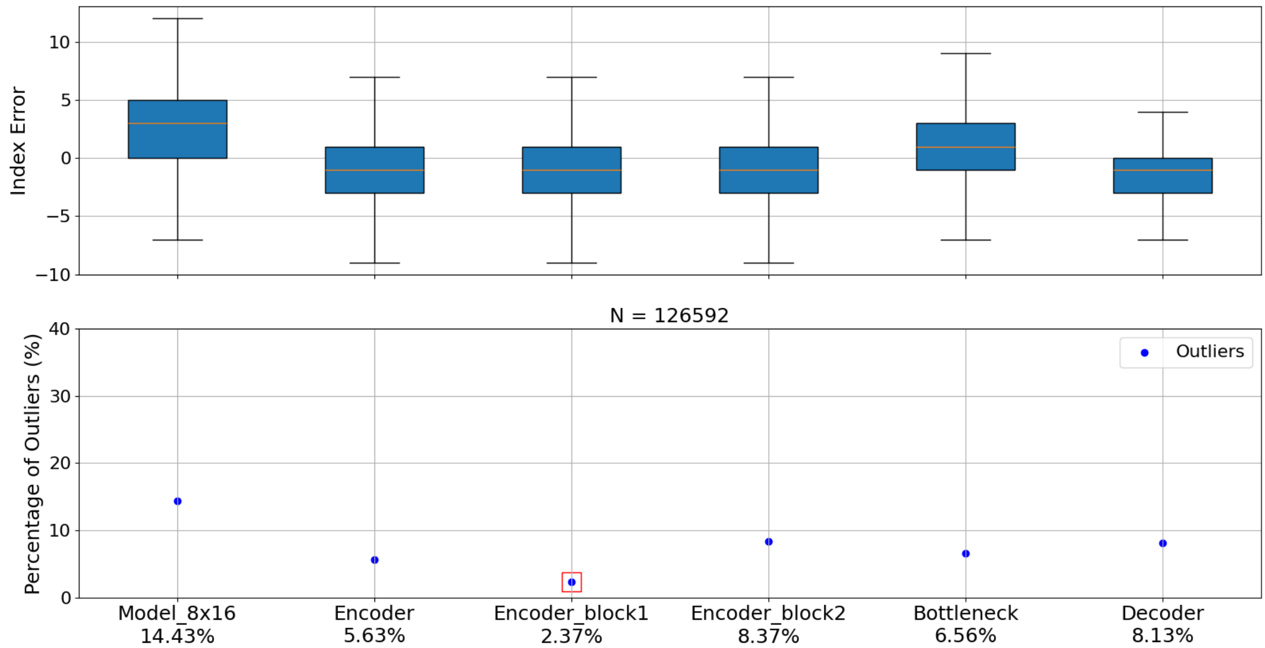


Figure 3: Comparison of Model Performance by customized metrics

the best performance in the metric space (with an average outlier rate of 2.6%, and an average $Idx_{MaxError}$ of 6.4%).

4. DISCUSSION

This study confirmed that knowledge learned from one array setup can be successfully leveraged for another, with fine-tuning improving segmentation accuracy. However, several aspects of this adaptation approach call for further discussion.

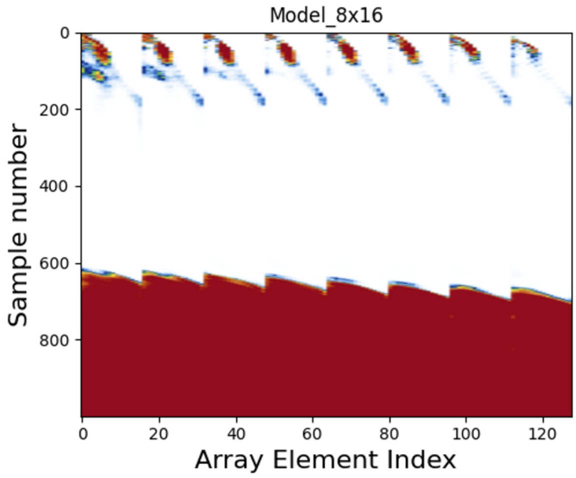
One of the key challenges in transfer learning is determining which layers should be fine-tuned to maximize performance while maintaining computational efficiency. This study evaluated various fine-tuning configurations, with results indicating that retraining the first encoder block yielded the most stable and accurate outcomes. However, the success of the first encoder block in this specific setting does not necessarily generalize to all ultrasound imaging problems, and a more thorough analysis could reveal alternative fine-tuning strategies with even better results. For instance, investigating the impact of fine-tuning different types of layers, beyond just convolutional ones, or employing techniques such as progressive fine-tuning [17], may further refine model adaptation.

While the study successfully adapted the model to an 8x16 array, the performance of this transfer learning approach could be explore further across different ultrasound imaging conditions, such as variations in material properties, noise levels, or different probe manufacturers.

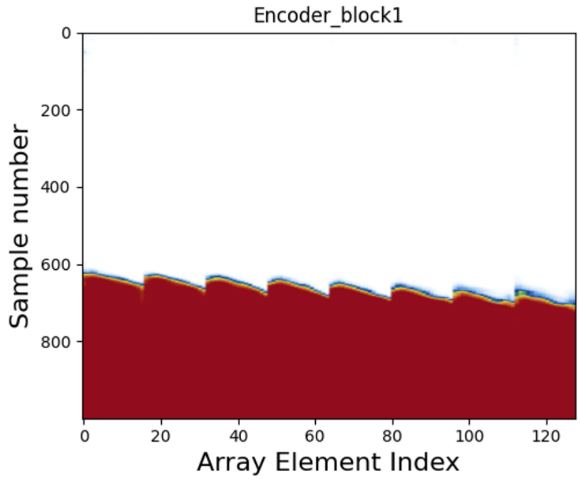
To enhance model adaptation even more, future research could explore adaptive fine-tuning strategies that automatically determines the optimal layer to update based on the dataset and task complexity [18]. Additionally, conducting experiments with a wider variety of geometries and different acquisitions conditions would help validate the robustness of the approach.

5. CONCLUSIONS

The results demonstrate that transfer learning plus fine-tuning was a viable approach for adapting the DeepEcho3D model for a different array configuration, without requiring a complete retraining process, and by utilizing only a single training component. The initial transferred model already showed promising performance, confirming that knowledge learned from one array setup can be effectively leveraged for another. Despite this, fine-tuning



(a) Segmentation for initial transferred model without fine-tuning



(b) Segmentation after fine-tuning the first block of the encoder

Figure 4: Comparison between Transfer learning vs Transfer learning + Fine-tuning

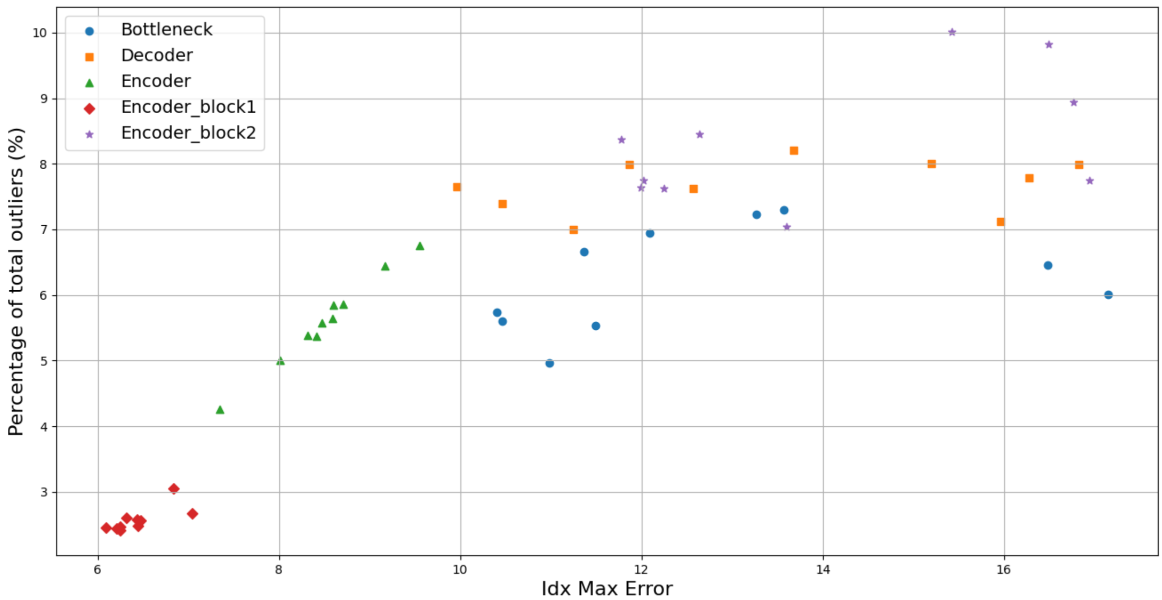


Figure 5: Stability Analysis

specific network sections enhanced segmentation accuracy.

The exploration of different fine-tuning strategies revealed that focusing on the first block of the encoder provides the most significant improvements, offering the best performance and greatest stability within the metric space. This suggests that early layers of the encoder, which capture more fundamental features, may be particularly critical for adapting the model to new array configurations.

6. ACKNOWLEDGMENTS

This research was supported by the project PID2022-143271OB-I00, funded by MCIN/AEI /10.13039/501100011033/FEDER, UE., the project PLEC2024-011165, funded by MICIU/AEI/10.13039/501100011033/ FEDER, UE, and by the European Commission – NextGenerationEU, through Momentum CSIC Programme: Develop Your Digital Talent. The funding for these actions/grants and contracts comes from the European Union's Recovery and Resilience Facility-Next Generation, in the framework of the General Invitation of the Spanish Government's public business entity Red.es to participate in talent attraction and retention programmes within Investment 4 of Component 19 of the Recovery, Transformation and Resilience Plan.

G. Cosarinsky staff is hired under the Generation D initiative, promoted by Red.es, an organisation attached to the Ministry for Digital Transformation and the Civil Service, for the attraction and retention of talent through grants and training contracts, financed by the Recovery, Transformation and Resilience Plan through the European Union's Next Generation funds.

7. REFERENCES

[1] B. W. Drinkwater and P. D. Wilcox, "Ultrasonic arrays for non-destructive evaluation: A review," *NDT & e International*, vol. 39, no. 7, pp. 525–541, 2006.

[2] C. Holmes, B. W. Drinkwater, and P. D. Wilcox, "Post-processing of the full matrix of ultrasonic transmit–receive array data for non-destructive evaluation," *NDT & e International*, vol. 38, no. 8, pp. 701–711, 2005.

[3] G. Montaldo, M. Tanter, J. Bercoff, N. Benech, and M. Fink, "Coherent plane-wave compounding for very high frame rate ultrasonography and transient elastography," *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, vol. 56, no. 3, pp. 489–506, 2009.

[4] L. Le Jeune, S. Robert, and C. Prada, "Plane wave imaging for ultrasonic inspection of irregular structures with high frame rates," in *AIP Conference Proceedings*, vol. 1706, AIP Publishing, 2016.

[5] Z.-b. Luo, Z.-h. Liu, F.-l. Li, and S.-j. Jin, "Defects imaging in corner part with surface adaptive ultrasonic and focusing in receiving (fir) strategy," *Journal of Nondestructive Evaluation*, vol. 43, no. 2, p. 49, 2024.

[6] J. Camacho, J. F. Cruza, J. Brizuela, and C. Fritsch, "Automatic dynamic depth focusing for ndt," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 61, no. 4, pp. 673–684, 2014.

[7] G. Cosarinsky, J. Cruza, M. Muñoz, and J. Camacho, "Optimized auto-focusing method for 3d ultrasound imaging in ndt," *NDT & E International*, vol. 134, p. 102779, 2023.

[8] F. Bai, D. Gagar, P. Foote, and Y. Zhao, "Comparison of alternatives to amplitude thresholding for onset detection of acoustic emission signals," *Mechanical Systems and Signal Processing*, vol. 84, pp. 717–730, 2017.

[9] S. Pan, Z. Qin, H. Lan, and J. Badal, "Automatic first-arrival picking method based on an image connectivity algorithm and multiple time windows," *Computers & Geosciences*, vol. 123, pp. 95–102, 2019.

[10] S. Cantero-Chinchilla, P. D. Wilcox, and A. J. Croxford, "Deep learning in automated ultrasonic nde–developments, axioms and opportunities," *Ndt & E International*, vol. 131, p. 102703, 2022.

[11] H. Hervé-Côte, F. Dupont-Marillia, and P. Bélanger, "Automatic flaw detection in sectoral scans using machine learning," *Ultrasonics*, vol. 141, p. 107316, 2024.

[12] M. Muñoz Prieto, A. Rubio Martínez, M. Larrea Álvarez, J. F Cruza, J. Camacho, and G. Cosarinsky, "Three-dimensional convolutional neural network for ultrasound surface echo detection," *Available at SSRN 4979182*.

[13] G. Cosarinsky, J. F. Cruza, M. Muñoz, and J. Camacho, "Automatic estimation of surface and probe location for 3d imaging with bidimensional arrays," *NDT & E International*, vol. 141, p. 102990, 2024.

[14] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?,"

Advances in neural information processing systems, vol. 27, 2014.

[15] C. Linse, B. Brückner, and T. Martinetz, "Enhancing generalization in convolutional neural networks through regularization with edge and line features," in *International Conference on Artificial Neural Networks*, pp. 432–446, Springer, 2024.

[16] C.-L. Fan, "Multiscale feature extraction by using convolutional neural network: Extraction of objects from multiresolution images of urban areas," *ISPRS International Journal of Geo-Information*, vol. 13, no. 1, p. 5, 2023.

[17] T. Kadeethum, D. O'Malley, Y. Choi, H. S. Viswanathan, and H. Yoon, "Progressive transfer learning for advancing machine learning-based reduced-order modeling," *Scientific Reports*, vol. 14, no. 1, p. 15731, 2024.

[18] G. Vrbančič and V. Podgorelec, "Transfer learning with adaptive fine-tuning," *IEEE Access*, vol. 8, pp. 196197–196211, 2020.