Deep learning for ultrasound surface echo detection

Guillermo Cosarinsky¹,²*, Mario Muñoz¹,², Jorge F. Cruza¹, Adrián Rubio¹ and Jorge Camacho³

¹Ultrasound Systems and Technology Group (GSTU), Institute for Physical and Information Technologies (ITEFI), Spanish National Research Council (CSIC), Madrid, Spain
²Non Destructive Testing Department, Comisión Nacional de Energía Atómica (CNEA), Buenos Aires, Argentina
³Electronics Department, Universidad de Alcalá, Alcalá de Henares, Madrid, Spain
*Contact: g.cosarinsky@csic.es

Abstract
Ultrasound array imaging techniques are often applied using a coupling medium to transmit waves from the transducer to the scanned component. In this type of tests, the echoes generated by the wave reflection at the component surface contain information that can be used to estimate the surface shape and its position relative to the probe, which are needed to compute the imaging focal laws. The most interesting feature to extract from surfaces echoes is the Time of Flight (TOF) for each reception channel. TOF measurement in acoustic signals is a problem that frequently arises not only in Non Destructive Testing (NDT), but also in medical imaging, geophysics and other fields.

The surface echo in a signal is usually identified as the first high amplitude peak, and is traditionally detected with standard methods as threshold crossing, peak search or matched filters. All these methods are easily affected by Signal to Noise Ratio, which might be high in some channels, for example if the ray arrives at the corresponding element at a low sensibility angle. Moreover, spurious echoes from bubbles, the emission signal tail or echoes from other parts of the component can be incorrectly labelled as surface echoes. This kind of outliers are difficult to filter by conventional means and, thus, there is a need for more robust surface estimation methods.

In this work, we study the application of a deep 3D Convolutional Neural Network (CNN) to detect surface echoes in Full Matrix Capture (FMC) ultrasound data. The CNN is trained using signals acquired with a matrix array and a reference component, using a robotic arm set-up to measure the probe position and orientation relative to the component. Thus, with knowledge of the set-up geometry, a model is used to compute the theoretical surface echo TOFs, which are used to label the data for training. The model was tested with a validation set, presenting good agreement with the ground truth. The results were compared with the threshold crossing method observing a significant reduction of outliers in the TOF estimation.

Keywords: ultrasound imaging, deep learning, surface detection, time of flight

1. Introduction
The use of a coupling medium in Ultrasound Testing (UT) produces signals with a strong echo produced by the reflection on the surface of the component under test. The detection of this surface echo and the measurement of its arrival time is a fundamental step for many signals processing tasks. In particular, the computation of focal laws for Phased Array or other imaging methods depends on the shape of the component surface. As ultrasound wave refract on that surface, the calculation of Times of Flight (TOF) from the array elements to the points inside the component requires the knowledge of the surface shape. This shape can be estimated using the surface echo TOFs [1-3].

Detection of strong echoes in signals is also fundamental for other disciplines. For example, in geophysics and seismology, fields in which extensive research was done on this subject [4]. As traditional methods are easily affected by background noise, the use of Machine Learning methods is an active research area [5-6].
In the field of medical ultrasound, deep learning methods were applied in [7] for automatic first arrival picking, applied to ultrasound sound-speed tomography. Deep learning methods are also being thoroughly researched for NDT applications [8]. However, to the best of the authors knowledge, no prior research has explored the application of these type of methods to the detection of surface echoes.

In this work we train a 3D Convolutional Neural Network (CNN) to detect surface echoes in an immersion testing set-up with a matrix array. The goal is to obtain a robust method for shape estimation, to be used in 3D auto-focusing algorithms.

2. Principles and methods

2.1 Statement of the problem

Let’s assume a matrix array performing a water immersion test on a component with surface S (Fig. 1). The array has $N_{el} = N_x N_y$ elements, where $N_x$ and $N_y$ are number of rows and columns respectively. Full Matrix Capture acquisition consist on firing one array element at a time, and, for each emission, receiving with all elements the array. This results in a matrix of A-scan signals $a(t, i, j)$ where $(i,j)$ are the linear indexes of the transmitter and receiver elements respectively. Waves reflected at the surface S produce echoes at times $t_s(i, j)$ in the signals. $t_s(i, j)$ is the TOF of the ray from element $i$ to element $j$ through the reflection point G.

The aim of this study is to establish a technique for quantifying $t_s$. As referenced in [9], the most common approach is threshold crossing. Here, we denote $t_s,thr$ as the minimum $t$ value satisfying $|a(t)| > thr$, where $thr$ represents a positive value. For the sake of notation simplicity, the indices $(i,j)$ have not been explicitly expressed. The term $|a(t)|$ denotes the envelope of the signal, or its absolute value when the analytical signal is unavailable.

![Fig. 1 Schematic of an array probe and the surface of a component under test](image-url)
Selecting an appropriate \( thr \) value is not an easy task. The value should be greater than background noise, but if it is too large, some low amplitude surface echoes might be undetected. Fig. 2 shows an example of this problem, with the signals received when a single element was excited in a 11x11 array (Fig. 2.b). Signals of each 11 element row form a B-scan image, and the B-scans of all rows are stacked from left to right. Surface echoes are observed forming low intensity stripes, while a much stronger indication comes from the component back-wall. It can be observed in Fig. 2.a that in some channels the surface echo is not correctly detected, as the threshold is not crossed. Instead, a part of the back-wall is incorrectly labeled as a surface echo. This might be solved by using a lower threshold. However, a lower threshold might be crossed by noise, or by the excitation pulse tail. Additionally, a \( thr \) value adjusted to work fine for one image, might not be adequate for another image. That is, searching a globally optimal threshold is not a practical solution.

In order to implement a more robust method to measure \( t_s \), we study in this article the possibility of training a CNN to achieve the measurement. For training the model, a set of labeled images is needed. The availability of such a set is one of the biggest problems for the application of machine learning methods in NDT. In our case, we propose a method for automatically labeling the images acquired in an experiment with a known geometry. This is described in the next section.

![Fig. 2 Problems of the threshold method. (a) B-scan type image with surface echoes and back-wall echoes, (b) Array elements and its linear indexes](image)
2.2 Data set creation and automatic labelling method

If the component under test has a known shape, and probe position and orientation (PLO) relative to the component is also known, the theoretical $t_s$ can be computed for each pair of array elements. This is achieved by finding a point $G$ on $S$ that satisfies the reflection law. For an arbitrary surface, this requires of numerical root finding methods. However, for some simple shapes, closed-form formulas can be applied (at least approximately). In [10] we developed an approximation method for cylinders and spheres, which is applied in this work to compute the ground truth used to train and test the CNN.

The data set was constructed by an experiment using a 3 MHz 11x11 array with 1 mm pitch. The component used was a 35 mm diameter cylinder tested in immersion. The probe was positioned by a 6 degrees of freedom robotic arm (Universal Robots, Denmark), shown in Fig. 3. FMC acquisitions were taken for a set of 180 different PLOs, using a SITAU ultrasound equipment by DASEL SL, Spain.

Fig. 4 shows two example acquisitions along the $t_s$ computed with the model presented in [10] in Fig. 2, the images present the signals received by all elements for one emitter. Each FMC matrix is composed of 121 images, one for each emitter. Thus, for each PLO, 121 images are obtained. The data set has then $180 \times 121 = 21780$ elements.

![Robotic arm and probe](image-url)
The signals, acquired with a 40 MHz sampling frequency, have 1200 samples, thus each element in the data set is a 121 x 1200 pixel image. The corresponding label is an equally sized binary image, with ones only in an M pixel thickness stripe below the theoretical TOF (black curve in Fig. 4), in this case, M=30.

The computed theoretical TOF presents in some case an offset relative to the measured surface echo. This is caused by systematic errors present in the determination of the PLO. However, we applied a compensation procedure, based on finding an optimal shift $\Delta$ for each image. We define a function $F(\Delta K)$ is defined as:

$$F(\Delta K) = \sum_{i=1}^{121} \sum_{k=K_{\text{g}}}^{k_{\text{s}}+W} |a(k + \Delta K, i)|$$

(1)

where $k$ is the sample index, $k_s$ is the sample index of $\tau_s$, $i$ is the receiver linear index, and $W$ is the width of the time window where the signal envelope is summed. This function measures the amount of signal “energy” inside a window linked to the theoretical TOF, as a function of an offset $\Delta K$. We select an optimal offset as the one that maximizes $F$. 
2.3 Neural network architecture

The chosen neural network architecture is a 3D auto-encoder based on the V-Net network, originally designed for three-dimensional segmentation in medical imaging [11].

The implemented network (Figure 5) was adapted to the peculiarities of our problem, and is composed of 2 convolutional blocks in the encoder part and 2 convolutional blocks in the decoder part. As shown in Fig. 5, Skip connections are used to provide spatial information sharing between blocks and improving the final prediction. Each convolutional block is composed of the following layers: Conv3D + BatchNorm + Dropout + Conv3D + BatchNorm, using Rectified Linear Units (ReLU) as activation functions. With these layers it is possible to extract the spatial features of the ultrasonic signals, preserving the spatial information of the input layer.

In order to adapt the input data to the proposed network architecture, the network input is reshaped to be 12x12x1200. This is achieved by replicating the signals of the last array element in each direction (x and y). The last convolutional layer has a 1x1x1 kernel with Sigmoid activation function, and its output has the same size as the input volume.

3. Results and discussion

The data set of 21780 examples were split in 60% train set and 40% test set, with randomly chosen elements. The training procedure was carried out in a NVIDIA RTX3080 Graphic Processing Unit (GPU) using 100 epochs with a Binary Cross-entropy loss function. Due to high memory requirements, batches of size 8 elements were used.

Fig. 6 shows an example from the test set. Fig. 6.a shows the image along with the points detected with a manually chosen threshold (thr = 50), the ground truth, and the prediction made with the customized V-Net. Fig. 6.b shows the V-net output image. The predicted TOF corresponds to 0.5 threshold on that output image, which has a range of values in [0, 1].
It can be seen in Figure 6.a that the threshold method fails for some channels, resulting in an undetected echo, or generating an outlier. On the other hand, the proposed method achieves a result almost equal to the ground truth. It is worth to note how the customized V-net filters out the back-wall echoes (Fig. 6.b)

In order to measure the overall performance on the whole test set, the error was computed for all channels in each image. This error is the difference (in number of samples) between the measured TOF and the ground truth. The error histograms are shown in Figures 6.a (threshold method with \( \text{thr} = 50 \)), and Figure 6.b (proposed CNN with 0.5 threshold). Table 1 shows error statistics, where a point was considered outlier if its error is more than 30 samples. It is observed that the threshold method has significantly larger systematic error, dispersion and outlier number than the customized V-net.

One of the limitations of using the proposed method compared to the threshold method is the inference time. While the inference time using the 3D CNN is 6 ms, enough to be performed in real time, in the case of the threshold method requires 1.7 ms. The increase in computational cost can be justified since the response is much more robust using the proposed method. It should be added that the designed network architecture could be optimized in future work, which would improve these results.
Table 1. Error statistics

<table>
<thead>
<tr>
<th></th>
<th>Mean (number of samples)</th>
<th>Standard deviation (number of samples)</th>
<th>Number of outliers (error &gt; 30 samples)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Threshold</td>
<td>-10.2</td>
<td>15.8</td>
<td>502</td>
</tr>
<tr>
<td>Custom V-net</td>
<td>1.1</td>
<td>2.7</td>
<td>25</td>
</tr>
</tbody>
</table>

Fig. 6 TOF error histograms. (a) Threshold method, (b) Custom V-net

4. Conclusions

With the obtained results, the effectiveness of employing Deep-Learning techniques for accurate response in TOF calculation has been shown. It demonstrated to produce less outliers and systematic errors than the traditional threshold method, while it does not require any adjustment parameter.

Despite the limitations of the study, the findings of this work can be a good starting basis for future work on the application of Deep-Learning tools for Non-Destructive Testing (NDT) area. Future work includes the study of the behavior of the method with other probe and component geometries, and the optimization of the inference time for high-speed real-time implementations.

Acknowledgement

This research was supported by the projects PDC2022-133504-I00 founded by MCIN/AEI/10.13039/501100011033 (Spain) and “Next Generation EU”/PRTR (European
For a detailed list of references, please see the text in the image.