Artificial Intelligence-based detection of exact and near duplicates of phased-array ultrasound data to prevent fraud

Étienne Drouin-Provencal[1], Richard Rheaume[1], Luc Laperrière[2], Lina Maria Strack[1], Yan-Éric Cossette[1]

[1]R&D, Ondia, Canada, [1]Mechanical Engineering, Université du Québec à Trois-Rivières, Canada

The detection of duplicate files in ultrasound scans of welds is a difficult task. The comparative analysis needed to detect duplicates is currently performed manually by certified analysts. However, due to the size of the files and their high number in projects, this task of cross comparing the files is error-prone. While there are automated tools to detect frauds and duplicates, none are appropriate in the case of phased-array ultrasound (PAUT) scans of welded joints, given the nature of the data. Neighboring welded joints are often very similar, and only small changes characterize each weld, so an automated detection tool needs to be precise and computationally efficient to be effective for Non-Destructive Testing (NDT) applications. The two most common types of duplicates that need to be identified in the NDT industry are exact and near duplicates. Exact duplicates, on one hand, are two sets of ultrasound data that are 100% identical. Near duplicates, on the other hand, are two sets of ultrasound data which are highly similar, most likely coming from the same weld and under the same conditions. Usually, their differences come from small transformations in the original data, such as shift normalization, shifts in the data and part suppression. Figure 1 shows examples of exact and near duplicates. The top line shows two exact duplicates of an S-Scan, and the bottom line shows near-duplicates. The near-duplicate S-Scans are different scans from the same stainless plate. While they may look very similar, they are different enough to trick most of the software currently used by the industry. Figure 1 Examples of duplicates of S-scans and PAUT: exact duplicates and near-duplicate The task of finding exact duplicates may be done either by comparing each point of two files or by randomly sampling the files and comparing these samples. The first option is highly inefficient and computationally expensive, because ultrasound data is comprised by millions of points in 4 dimensions. Analyzing sample points, selected randomly, would reduce the computational burden of the software while being statistically valid. However, this technique cannot find near-duplicates, as simply shifting one single entry in the data would cause this process to fail. Finding near-duplicates is therefore a much more difficult task, to which artificial intelligence, particularly deep learning, can contribute a great deal. Deep learning models can now successfully perform a large variety of tasks, including finding image duplicates with high accuracy in large datasets. Thus, the aim of this study was to leverage deep learning models to efficiently detect exact and near-duplicates in 3D PAUT scans. This paper demonstrates the efficiency of these models by using them to compare two sets of data, original and modified. From the original dataset, 60 ultrasound scans were randomly selected. The first 30 ultrasound scans were kept intact while the remaining 30 were altered slightly. Transformations included cropping, deleting random points, inverting the order of the data and increasing the average amplitude. The detection models were then applied to compare these datasets: the original with its exact duplicates, and the original with its 30 near-duplicates. Results show that our method can detect all 30 exact duplicates very quickly and is also very promising to find near-duplicates, even elaborate ones. A detailed discussion of the results will be provided in the complete paper.