Novel Multi Data Acquisition and Hybrid Neural Network for Pipe Inspection and Imaging

Guang An Ooi
Electrican and Computer Engineering
King Abdullah University of Science and Technology
Thuwal, Saudi Arabia
guang.ooi@kaust.edu.sa

Mehmet Burak Ozakin
Electrican and Computer Engineering
King Abdullah University of Science and Technology
Thuwal, Saudi Arabia
mehmet.ozakin@kaust.edu.sa

Tarek M. Mostafa
Electrican and Computer Engineering
King Abdullah University of Science and Technology
Thuwal, Saudi Arabia
tarek.mostafa@kaust.edu.sa

Moutazbellah Khater
Electrican and Computer Engineering
King Abdullah University of Science and Technology
Thuwal, Saudi Arabia
moutazbellah.khater@kaust.edu.sa

Ahmed Al-Jarro
Research & Develoment Center
Aramco Research Center at KAUST
Saudi Aramco
Thuwal, Saudi Arabia
ahmed.aljarro@aramco.com

Hakan Bagci
Electrican and Computer Engineering
King Abdullah University of Science and Technology
Thuwal, Saudi Arabia
hakan.bagci@kaust.edu.sa

Shehab Ahmed
Electrican and Computer Engineering
King Abdullah University of Science and Technology
Thuwal, Saudi Arabia
shehab.ahmed@kaust.edu.sa

***Abstract*—A new machine learning-based non-destructive testing (NDT) technique for the examination of conductive objects is presented. NDT of objects behind barriers can utilize the defect-induced distortions on alternating electromagnetic (EM) fields to detect flaws in the structure of inspected targets. Such distortions are highly non-linear with respect to defect properties, requiring significant amounts of data for training neural networks to generate accurate predictions. To this end, a massively parallelized data generation framework is proposed in conjunction with a multi-frequency hybrid neural network (MF-HNN), to create a physics-informed inversion AI model. The performance of the resulting inversion algorithm is applied on casings, where tubular conductive pipes are inspected. For data generation, physics-based solvers are employed to simulate the EM field distribution resulting from pipes with defects. The large-scale distribution of this step leads to 43 times faster execution time than a single CPU, with big and diverse datasets. This allows the MF-HNN to achieve significantly improved generalization performance and to generate the desired high-resolution cross-sectional images of the inspected pipelines.**

Keywords—Machine learning, massively parallelized simulations, hybrid neural network, NDT, EM-based inspection.

I. Introduction

In the past decade, the field of machine learning has seen significant increase in the complexity of artificial neural network (ANN) architectures [1]. The availability of specialized hardware and big data have further allowed for advancements in addressing various challenges, including the development of physics-informed ANNs [2]. ANNs incorporate various types of layers that can be organized in parallel and/or cascading combinations, creating hybrid neural networks (HNNs). They benefit from the extraction of features from spatial and/or temporal variations of data.

In data fusion scenarios, multiple sensors at deployed at various locations to provide measurement data, generating large datasets. An example of such problems is the deployment of electromagnetic (EM) sensors for the inspection of tubulars, such as carbon steel pipelines. In the oil and gas industry, the harsh environments can lead pipes to suffer from loss of metal as a result of both corrosion and physical damages. Such metal losses potentially threaten the structural integrity of pipes, requiring reliable early detection mechanisms to help prevent economic and environmental damages. For example, transmitter (TX) coils emit alternating EM fields that induce eddy currents in the pipes, in turn generating secondary EM fields in their surroundings. The distribution of these secondary fields is distorted by the presence of metal losses. Large arrays of EM sensors, installed on inspection tools that scan along the interior of the pipes, measure such distortions. In this work, the measured data is fed into novel AI-driven inversion algorithms to infer the severity of metal losses. This inspection technique results in large spatiotemporal datasets. In addition to the multitude of noise sources, the inherent non-linearity of EM fields with respect to metal loss properties renders such data challenging to interpret. Moreover, the diverse variety of operating and environmental conditions, coupled with parameters of the inspection tools, requires the ANN inversion models to process large and diverse datasets for acceptable accuracy.

In many supervised learning problems, the collection and processing of real data require considerable amount of time and effort. The parameters of sensors data, such as cameras, thermometers, and EM receivers also require pre-definition, rendering them intolerant to changes in the data acquisition protocol. In addition, such obtained data may not be sufficiently large, diverse and/or evenly distributed, causing overfitting in the trained ANNs. To address these limitations, simulated data are considered one of the most feasible and attractive methods to provide these data needs.

To simulate EM-related problems, numerical solution of Maxwell equations is required. However, these numerical methods can consume significant computing power, especially for problems where high spatial resolution is required. In this work, a massively parallelized simulation framework is designed to accelerate the generation of such simulated datasets. This approach is applied as part of an EM casing inspection method, where the EM field distributions in pipes with various types of metal losses are considered. TX signals of multiple frequencies are generated to detect inner, outer, and full losses. The resulting datasets are used to train a bespoke multi-frequency hybrid neural network (MF-HNN) that uses simulated EM data as input to generate pipe cross-sectional images; revealing shapes, dimensions, and locations of defects. The MF-HNN is shown to reliably identify metal losses, using a limited number of magnetic sensors.

II. Methodology

The proposed framework is shown in Figure 1. The EM solvers are run on the Ibex HPC Cluster [2] of the King Abdullah University of Technology and Science, KAUST, Supercomputing Laboratory, KSL. It hosts a combination of Skylake, Cascade Lake, and AMD Rome CPUs, as well as Nvidia GPUs. A batch of 700 simulations is concurrently executed, with an average overall completion time of eight hours. Available processors are automatically assigned new jobs to maximize efficiency. Since metal losses occur in myriad shapes, locations, dimensions, and combinations, the parameters of metal loss are randomized at the beginning of each batch of simulation. Once a batch is completed, the resulting distribution of EM fields and metal loss cross-section image is recorded, and the solver is reinitialized with new random parameters.

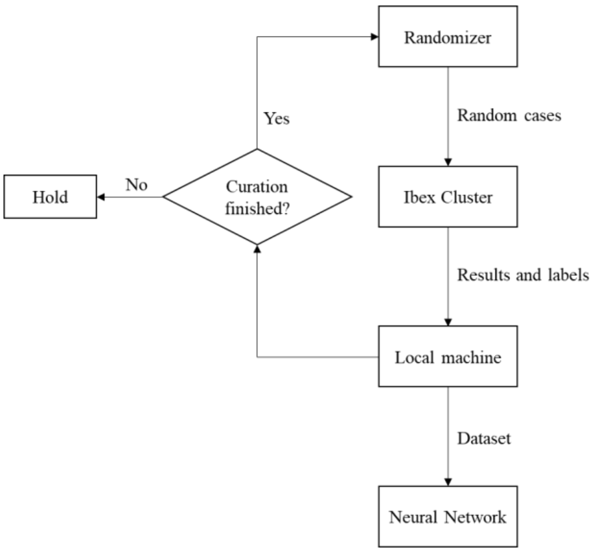

Figure 1: The framework of simulation parallelization.

In the simulations, the defects are configured to be either full losses (holes), partial losses on the inner or outer surface of the pipe wall, or random combinations of them. A random number of individual metal losses are created on the pipe, with the first metal loss having a 100% chance of appearance, the second 25%, the third 6.25%, and so on. Multiple defects are allowed to fully or partially overlap with one another to create more complex defect shapes that possess varying sizes and levels (depths) of metal loss. A total of 100,000 different scenarios are simulated, and the results are compiled. The obtained datasets are then split into 60%, 20%, and 20% to train, validate, and test the MF-HNN, respectively.

III. Results

A local machine with an i9-9980, 2.40 GHz processor and 64 GB RAM completes one simulation in about 30 minutes. In comparison, a parallelized batch of 700 simulations is completed within 8 hours on average. To generate the 100,000 scenarios, the parallelized framework reduces the runtime from 2083 days to 48 days when using 145 batches. The outcome of the MF-HNN model are presented in Figure 2. It shows the cross-sectional inversion image of a pipe with the same defect, i.e. one record from the test dataset. It is evident that there are significant improvements in quality of prediction, middle column, in comparison to ground truth, left column, as the amount of training data is increased, top to bottom. The difference errors plotted in the right column.

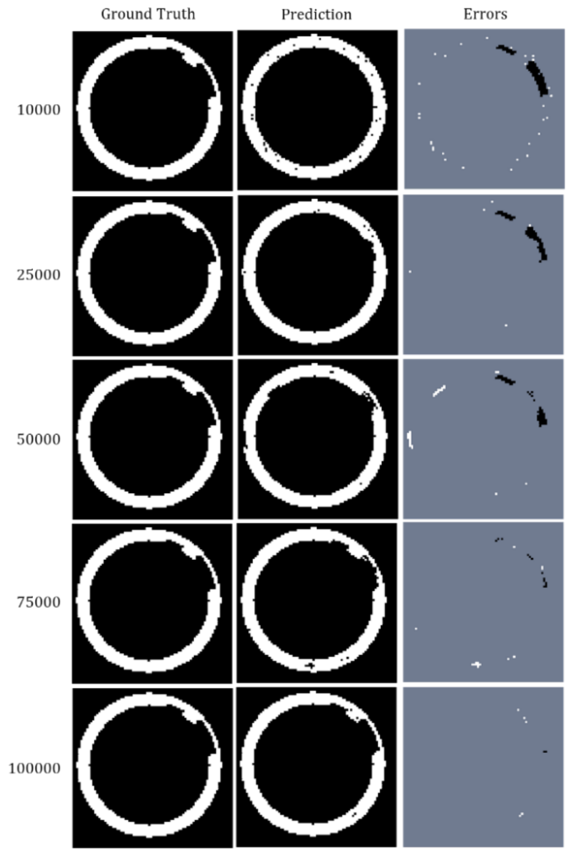

Figure 2: Predictions of the MF-HNN on the same test case, trained with increasing amounts of data points, top to bottom.

To evaluate the performance of the AI model, the accuracy and specificity of the MF-HNN on the overall test datasets are calculated, with the resulting values are shown in Table I.

Table I: Performances of the MF-HNN.

Data Count	Accuracy	Specificity
10000	0.8322	0.0159
25000	0.8574	0.1053
50000	0.8618	0.2264
75000	0.9236	0.6345
100000	0.9790	0.9081

The generalization performance in the proposed MF-HNN AI-model demonstrates the importance of a suitably effective data generation framework. As shown in this study, an appropriate and diverse dataset is required for a reliable pipe inspection NDT approach. The simulated data is generated within a feasible amount of time via modern HPC clusters, achieving up to 43x speedup in comparison to a serial run.

IV. REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning. Nature, 521(7553), 436–444, 2015.

[2] Ibex HPC Cluster - https://www.hpc.kaust.edu.sa/ibex