CLOSING THE SIM-TO-REAL GAP IN GUIDED WAVE DAMAGE DETECTION WITH ADVERSARIAL TRAINING OF VARIATIONAL AUTO-ENCODERS

Ishan D. Khurjekar, Joel B. Harley

University of Florida
Dept. of Electrical and Computer Engineering
Gainesville, FL 32608

ABSTRACT

Guided wave testing is a popular approach for monitoring the structural integrity of infrastructures. We focus on the primary task of damage detection, where signal processing techniques are commonly employed. The detection performance is affected by a mismatch between the wave propagation model and experimental wave data. External variations, such as temperature, which are difficult to model, also affect the performance. While deep learning models can be an alternative detection method, there is often a lack of real-world training datasets. In this work, we counter this challenge by training an ensemble of variational autoencoders only on simulation data with a wave physics-guided adversarial component. We set up an experiment with non-uniform temperature variations to test the robustness of the methods. We compare our scheme with existing deep learning detection schemes and observe superior performance on experimental data.

Index Terms— Damage detection, guided waves, sim-to-real, variational auto-encoder, adversarial training

1. INTRODUCTION

Infrastructures with daily utility, such as airplanes, bridges, and buildings, need to be monitored regularly for their structural integrity. Guided wave based testing (GWT) is a popular approach for monitoring structural health as these waves can travel over long distances and are sensitive to structural defects [1]. A GWT setup consists of a spatially distributed sensor array, that can transmit and receive waves, placed on the structure to be investigated for the presence of damage.

In this paper, we focus on the task of structural damage detection using a GWT setup. A number of target detection schemes, such as matched filtering [2], energy detectors [3], among others, can be applied to damage detection in ideal conditions. On the other hand, external variations, such as temperature affect wave amplitude, phase, and velocity [4, 5]. This leads to a mismatch between the theoretical wave propagation model and the experimental data. For methods based on matched filtering, the model mismatch is a major challenge [6]. In addition, there is no method for perfectly removing temperature effects from data sets [7]. Hence, we cannot remove temperature from the problem.

Researchers have also proposed machine learning based approaches for monitoring structural integrity [8, 9]. Yet, obtaining real-world guided wave datasets for training machine learning models is resource intensive, particularly for damage detection. While the problem can be framed as a binary classification problem, it is difficult to obtain training data that is representative of both damage and no-damage classes.

Instead, in this paper we pose the damage detection problem as an out-of-distribution (OoD) detection problem. Generative models are commonly used for OoD detection. These include strategies based on likelihood models [10], autoregressive networks [11], adversarial networks [12], and variational autoencoders (VAE) [13], among others. A possible strategy that uses a likelihood model would be to learn a model for a related task, such as damage localization, and then threshold the likelihood value for damage detection. Yet, such an approach lacks robustness as it cannot effectively capture the input variability. Indeed, researchers have shown that generative methods are not necessarily robust as they assign spurious likelihood values to OoD inputs [14].

We propose a VAE ensemble approach with two salient features to enhance applicability to realistic guided wave damage detection:

- We train the VAE ensemble on simulation data alone, eliminating the need to set up resource intensive experiments for training data generation.
- We simulate wave physics-based adversarial perturbations in the training data to enable robustness to input variability (temperature variations).

We choose the VAE objective value (lower bound on the data likelihood) [15] as the damage detection statistic since large values indicate that we are in-distribution and therefore match the simulation data. We compare the performance of our approach with other deep generative approaches on experimental data with non-uniform temperature variations. Our

This research is funded by the National Science Foundation under award number EECS-1839704.

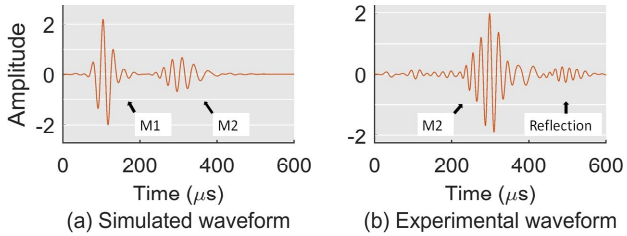

Fig. 1: Simulation and experimental waveform comparison. Both the waves travel a distance of 0.5681m

approach achieves superior detection performance and well-separated statistic values signifying superior robustness

2. SIM-TO-REAL DAMAGE DETECTION FRAMEWORK

We propose a VAE framework for detecting the presence of damage in a structure using guided waves with two salient features: training on simulation data alone with simulated adversarial perturbations for robustness to temperature variations. The problem setup and the framework is explained in the following subsections.

2.1. Guided wave setup

In this paper, we simulate guided wave structural health monitoring data. We assume the structure to be a square plate that is investigated for the presence of damage. A sensor array placed on the structure transmits and receives signals (wideband signals with Q frequencies and M sensor pair measurements). A Lamb wave model is used to describe the wave propagation [16]. The general Lamb wave model is given as,

$$x(\omega, r) = \sum_n \sqrt{\frac{1}{\kappa_n(\omega) r}} s(\omega) e^{-j\kappa_n(\omega) r}, \quad (1)$$

where the guided wave signal, $x(\omega, r)$, for frequency ω and at distance r from source is modeled as a superposition of wave modes. $s(\omega)$ is the transmitted signal and $\kappa_n(\omega)$ is the frequency and mode dependent wavenumber.

We assume the received signal travels two paths. The first path is directly from the transmitter to the receiver (baseline signal: x_b). The second path is from the transmitter to the damage and then to the receiver (damage signal: x_s). This is mathematically expressed as

$$x(\omega, r) = x_b(\omega, r) + \alpha x_s(\omega, r), \quad (2)$$

where α is the reflection coefficient. As we standardize data, the choice of α does not affect the results.

Fig. 1(a-b) shows the simulated and experimental guided wave signals respectively. Note that the wave mode-2 (referred to as A0 mode [16]) coincides in the simulated and experimental signal but mode-1 (referred to as S0 mode [16]) is weak in the experimental signal.

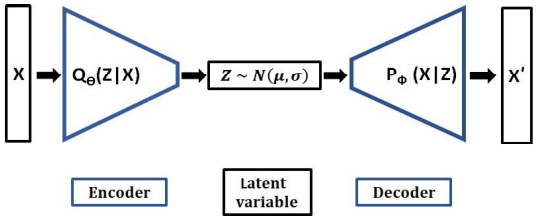

Fig. 2: VAE architecture

Before processing the data, we apply baseline subtraction. That is, the baseline signal (x_b) is subtracted in order to isolate the damage signal (x_s). Ideally, only the damage signal and noise should remain after baseline subtraction. Yet, baseline subtraction is not perfect in presence of distorting effects of temperature variations as we describe in Section 3.2. While methods exist to reduce the effects of temperature [17], no method can perfectly remove them.

2.2. Variational autoencoder: VAE

We pose the damage detection problem as an OoD detection problem for which we use a VAE. We train the VAE on the baseline subtracted signal with the assumption of damage as in (2). Hence signals without damage component are out-of-distribution. VAE consists of an encoder and a decoder as shown in Fig. 2. It reconstructs the input data using a probabilistic latent variable model. Specifically, the lower bound on data likelihood (ELBO: evidence lower bound) [15] is maximized. The ELBO is written as,

$$\log p(x) \geq E_{z\sim Q_\theta}[\log P_\phi(x|z)] - \mathcal{D}[Q_\theta(z|x)||P_\phi(z)], \quad (3)$$

where Q_θ and P_ϕ are the encoder and decoder networks, respectively. The first term in (3) is the cross entropy and the second term is the Kullback-Liebler divergence. The latent random variable z is assumed to follow a Normal distribution.

2.3. Closing the sim-to-real gap

Researchers have built robust deep learning models by training on data with adversarial perturbations [18, 19]. Therefore, we simulate adversarial perturbations in the training data based on empirical modeling of temperature effect on wave propagation. These perturbations are simulated by multiplying the wavenumber in (1) by a random factor defined as,

$$\kappa'_n(\omega) = \gamma \kappa_n(\omega), \quad (4)$$

where γ is the random multiplicative factor sampled uniformly from the interval $[1-\delta, 1+\delta]$. We choose $\delta = 0.02$ to match the range of wavenumber variations ($\pm 2\%$) in the experimental data as caused by temperature. Further, we train an ensemble ($n = 10$) of VAE's each with a different weight initialization, as ensembling has also been shown to increase robustness on acoustic tasks [20].

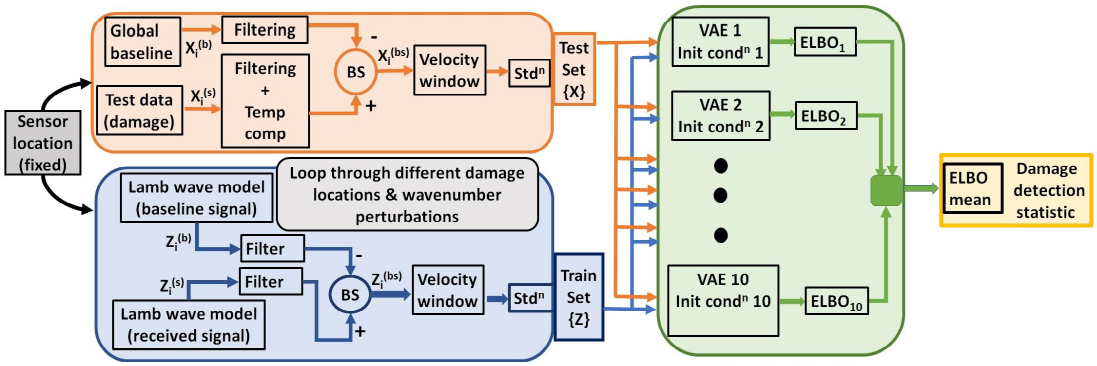

Fig. 3: Framework validation setup

2.4. Damage detection

We define the damage detection statistic (τ) as,

$$\tau(x) = \frac{1}{Q \times M} \frac{1}{n} \sum_{i=1}^{n} ELBO_i(x), \tag{5}$$

where $\tau(x)$ is the mean of ELBO estimates (ELBO$_i$'s) from the VAE ensemble normalized for Q frequencies and M sensor pair measurements. Intuitively, we should have high statistic value when the test sample is "in-distribution" (damage) and conversely low statistic value when test sample is "out-of-distribution" (no damage).

3. FRAMEWORK VALIDATION

The complete damage detection framework is illustrated in Fig. 3. The individual framework components are explained in the subsequent subsections.

3.1. Implementation

For training, we simulate $Q = 1000$ frequencies and $M = 240$ sensor pair measurements (16 element sensor array) to match the experimental setup, described in Section 3.3. We refer to one spatio-temporal observation matrix of dimensions $Q \times M$ as one sample. We simulate $t = 5000$ samples with 4000 samples used as training set and the rest 1000 samples used as the validation set. We convert samples to time-domain and standardize them before inputting to VAE.

The wave signal is pulse compressed to remove unwanted dependence on the transmitted signal phase. We remove the initial 40 μs of signal to remove electromagnetic interference. We pass the signal through a low pass Gaussian filter (center frequency $f_c = 37.5$ kHz and bandwidth $B = 30$ kHz) as the effects of damage are observed at these lower frequencies. We apply an exponentially tapering velocity window ($v_{win} = 1500$ m/s) to remove the unwanted reflections from the plate boundaries.

The VAE architecture is detailed in Table 1. We use 1D convolutional layers in the encoder and transposed 1D convolutional layers (also called deconvolution) in the decoder. We apply batch normalization after every layer for faster training. We apply dropout regularization for all dense layers with dropout probability = 0.1. We use the reparameterization trick for latent space sampling [15]. We train each VAE in the ensemble with simulated data alone for 15 epochs with a batch size of 16. All the above mentioned parameter values are chosen to maximize detection performance.

Table 1: VAE architecture

Layer	Layer description	Activation
Conv1D	Filters = 12 ; kernel size = 3 ; stride length = 2	ReLU
Conv1D	Filters = 24; kernel size = 3 ; stride length = 2	ReLU
Dense	Fully connected ; nodes = 1200	Sigmoid
Dense $\times 2$	latent_dim = 2	-
Dense	Fully connected ; nodes = 1200	Sigmoid
Dense	Fully connected ; nodes = Q$\times$M	Sigmoid
Conv1D Tanspose	Filters = 24; kernel size = 3 ; stride length = 2	ReLU
Conv1D Tanspose	Filters = 12; kernel size = 3 ; stride length = 2	ReLU

3.2. Baseline subtraction

The baseline signal has to be subtracted from the received signal to isolate the damage signal while compensating for the effect of temperature. Popular strategies include choosing from a baseline bank [21] and / or stretching signals using scale transform [17]. We create a calibration signal bank by choosing one signal each from the damaged and undamaged experimental signal set randomly. The calibration bank is considered as an extension of the validation set. For a particular test signal, we choose the calibration signal that minimizes residual energy. We stretch the test signal using the scale transform [17] (removing some of the effects caused by temperature) to match the chosen calibration signal and then subtract a globally chosen baseline signal.

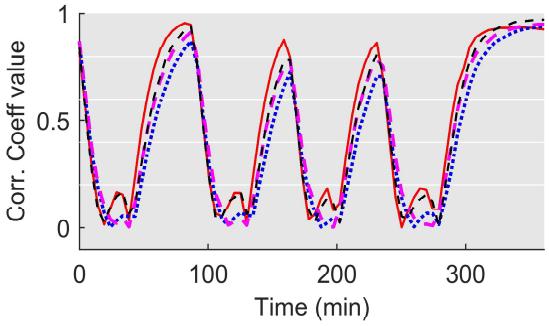

Fig. 4: Correlation between first experimental measurement and successive measurements for 4 sensor pairs

3.3. Experimental setup

We mount 16 sensors at random locations on an aluminum plate of size 1.22 m $\times$ 1.22 m. We set up the data acquisition system to have a sampling rate of 1 MHz. We transmit a 0.1 ms long chirp signal with a frequency sweep of 50 kHz to 500 kHz. We record 4 ms long measurements on the receiving sensors. Spatio-temporal variations are introduced over the plate using a heating fan placed in a corner. The temperature is varied periodically from approximately 24°C to 39°C across the plate. We collect 76 measurements over $\approx$ 6 hours. We physically simulate damage by placing a mass at $(0.53, 0.60)$ m from the 37^{th} measurement onward. Fig. 4 shows the correlation values between first and successive measurements. This illustrates the effect of temperature variations on wave propagation.

4. RESULTS

We compare our VAE based scheme with a likelihood model-based detection scheme, utilizing a feedforward neural network. We train a network with guided wave data generated using (1) to identify corresponding damage locations. This network uses a Gaussian likelihood as the objective function, which is used as a damage detection metric.

The detection threshold(s), (τ_0), for the method(s) are computed individually as follows: The detection statistic values are calculated for the same randomly chosen calibration signals used for baseline subtraction. The mid-point of these values is chosen as the detection threshold. This is done to maximize separation between damage and no damage case as well as for fair comparison of all methods. We define the probability of damage detection (p_d) from the detector as

$$p_d = p(\ \tau \geq \tau_0 \mid H_A), \tag{6}$$

and the probability of false alarm (p_{fa}) as

$$p_{fa} = p(\ \tau \geq \tau_0 \mid H_0), \tag{7}$$

where τ, τ_0, H_0, H_A represent the detection statistic, detection threshold, null hypothesis (no damage), and alternative hypothesis (damage) respectively.

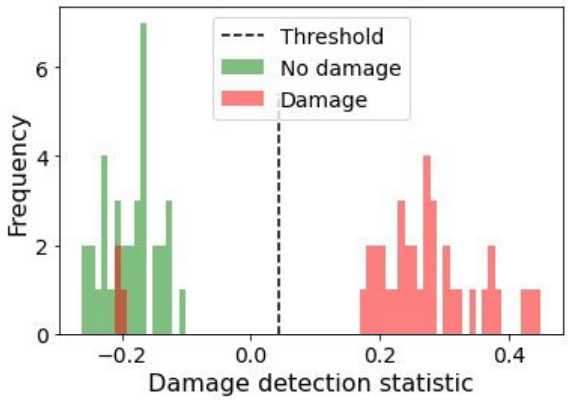

Fig. 5: VAE ensemble damage detection histogram

Table 2: Damage detection performance

Method	p_d	p_{fa}
VAE-adv training $(*)$	0.923	0.000
VAE-ideal training	0.461	0.000
Likelihood model-adv training	0.923	0.600
Likelihood model-ideal training	0.487	0.712
Energy detection	0.897	0.228
Matched filter	0.897	0.028

Table 2 shows the performance comparison of the VAE based approach and the likelihood model-based approach. We also compare the performance of models trained on ideal data and on data with adversarial perturbations (denoted as -ideal training and -adv training respectively). We first observe that models trained on data with adversarial perturbations have superior performance compared to those trained on ideal data. This underscores the importance of adversarial training for enhancing robustness.

Next, we compare the performance of our approach and the likelihood model-based detection scheme. The probability of detection (p_d) is equal for both but our VAE scheme has a much better false alarm rate (p_{fa}: 0.00 compared to 0.60). This is in line with the observation that likelihood models assign spurious likelihood values to OoD inputs. Fig 5 shows the histogram of the damage detection statistic for our scheme (VAE trained on adversarial data). The detection statistic is well-separated, signifying robustness.

5. CONCLUSIONS

Here, we pose the problem of guided wave-based damage detection problem as an OoD detection problem. We propose a VAE ensemble network based approach for this task which is trained on simulation data alone. Ensemble approach together with adversarial training enables robustness. To validate the framework, we set up an experiment to collect guided wave data in presence of non-uniform temperature variations. Results illustrate that the detection performance of the proposed

framework is robust to temperature variations and superior to other deep generative methods.

6. REFERENCES

[1] Joseph L Rose, *Ultrasonic guided waves in solid media*, Cambridge University Press, 2014.

[2] Frank C Robey, Daniel R Fuhrmann, Edward J Kelly, and Ramon Nitzberg, "A cfar adaptive matched filter detector," *IEEE Transactions on Aerospace and Electronic systems*, vol. 28, no. 1, pp. 208–216, 1992.

[3] Harry Urkowitz, "Energy detection of unknown deterministic signals," *Proceedings of the IEEE*, vol. 55, no. 4, pp. 523–531, 1967.

[4] Ajay Raghavan and Carlos ES Cesnik, "Effects of elevated temperature on guided-wave structural health monitoring," *Journal of Intelligent Material Systems and Structures*, vol. 19, no. 12, pp. 1383–1398, 2008.

[5] George Konstantinidis, Bruce W Drinkwater, and Paul D Wilcox, "The temperature stability of guided wave structural health monitoring systems," *Smart Materials and Structures*, vol. 15, no. 4, pp. 967, 2006.

[6] William L Melvin, "Space-time adaptive radar performance in heterogeneous clutter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, no. 2, pp. 621–633, 2000.

[7] Alexander CS Douglass and Joel B Harley, "Dynamic time warping temperature compensation for guided wave structural health monitoring," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 65, no. 5, pp. 851–861, 2018.

[8] Joseph Melville, K Supreet Alguri, Chris Deemer, and Joel B Harley, "Structural damage detection using deep learning of ultrasonic guided waves," in *AIP Conference Proceedings*. AIP Publishing LLC, 2018, vol. 1949, p. 230004.

[9] Ishan Khurjekar and Joel Harley, "Deep neural network-based guided wave damage localization," *Review of progress in quantitative nondestructive evaluation*, 2019.

[10] Christopher M Bishop, "Novelty detection and neural network validation," *IEE Proceedings-Vision, Image and Signal processing*, vol. 141, no. 4, pp. 217–222, 1994.

[11] Ellen Rushe and Brian Mac Namee, "Anomaly detection in raw audio using deep autoregressive networks," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 3597–3601.

[12] Yubo Hou, Zhenghua Chen, Min Wu, Chuan-Sheng Foo, Xiaoli Li, and Raed M Shubair, "Mahalanobis distance based adversarial network for anomaly detection," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 3192–3196.

[13] Vijaya Kumar Sundar, Shreyas Ramakrishna, Zahra Rahiminasab, Arvind Easwaran, and Abhishek Dubey, "Out-of-distribution detection in multi-label datasets using latent space of β-vae," in *2020 IEEE Security and Privacy Workshops (SPW)*. IEEE, 2020, pp. 250–255.

[14] Eric Nalisnick, Akihiro Matsukawa, Yee Whye Teh, Dilan Gorur, and Balaji Lakshminarayanan, "Do deep generative models know what they don't know?," *arXiv preprint arXiv:1810.09136*, 2018.

[15] Diederik P Kingma and Max Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[16] Horace Lamb, "On waves in an elastic plate," *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, vol. 93, no. 648, pp. 114–128, 1917.

[17] Joel B Harley and José MF Moura, "Scale transform signal processing for optimal ultrasonic temperature compensation," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 59, no. 10, pp. 2226–2236, 2012.

[18] Qing Wang, Wei Rao, Sining Sun, Leib Xie, Eng Siong Chng, and Haizhou Li, "Unsupervised domain adaptation via domain adversarial training for speaker recognition," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 4889–4893.

[19] Tejas Gokhale, Rushil Anirudh, Bhavya Kailkhura, Jayaraman J Thiagarajan, Chitta Baral, and Yezhou Yang, "Attribute-guided adversarial training for robustness to natural perturbations," *arXiv preprint arXiv:2012.01806*, 2020.

[20] Fuad Noman, Chee-Ming Ting, Sh-Hussain Salleh, and Hernando Ombao, "Short-segment heart sound classification using an ensemble of deep convolutional neural networks," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 1318–1322.

[21] Yinghui Lu and Jennifer E Michaels, "A methodology for structural health monitoring with diffuse ultrasonic waves in the presence of temperature variations," *Ultrasonics*, vol. 43, no. 9, pp. 717–731, 2005.