

Pushing Distributed Vibration Analysis to the Edge with a Low-Resolution Companding Autoencoder: Industrial IoT for PHM

Alexandre Trilla¹, David Miralles², and Verónica Fernández³

^{1,3} *Alstom, Santa Perpètua de la Mogoda, Barcelona, 08130, Spain*
alexandre.trilla@alstomgroup.com
veronica.fernandez@alstomgroup.com

² *Grup de Recerca en Tecnologies Media, La Salle - Universitat Ramon Llull, Barcelona 08022, Catalonia, Spain*
david.miralles@salle.url.edu

ABSTRACT

The Industrial Internet-of-Things (IIoT) has disrupted the way of collecting physical data for predictive maintenance purposes. At present, networks of intelligent wireless sensors are pervasive, finding success in many environments and industries, including the railways. However, when it comes to data-intensive applications like vibration monitoring that require the delivery of large amounts of records, the limitations of these devices arise. The shortfalls are mainly driven by the low-bandwidth transmission capacity of their radio interfaces, and the low-power features of their battery-operated (and/or energy-harvested) electronics. In sight of these limited resources, this article explores a vibration data compression strategy for diagnosis purposes. To maximise the amount of transferred information with the least amount of bytes this method works in three stages: first, it extracts the most useful features for vibration-based analytics. Then, it compresses the raw signal waveform using an Autoencoder neural network with an undercomplete representation, assessing its optimum regularisation approach: the denoising, sparse, and contractive configurations. Finally, it reduces the resolution of the compressed data by quantising all the resulting real values into single-byte unsigned integers. The proposed strategy is evaluated with a dataset of railway axle bearings with different levels of degradation. The results of the analysis show that with compression rates up to 10 the vibration signals are practically unaffected by this procedure, and once the signals are reconstructed with a minimum quality standard, many diagnosis goals like anomaly detection, fault location, and severity appraisal can be performed. This approach yields a wide range of business opportunities for on-board predictive maintenance with IIoT technology.

Alexandre Trilla et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 United States License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1. INTRODUCTION

The Industrial Internet-of-Things (IIoT) has emerged as one of the leading technologies to deploy the remote condition monitoring of machines (Boyes, H., Hallaq, B., Cunningham, J., and Watson, T., 2018), especially when such machines are transportation assets that move around the territory. This work is particularly concerned with the application of Prognostics and Health Management (PHM) to the maintenance of mechanical rolling-stock components (Atamuradov, V., Medjaher, K., Dersin, P., Lamoureux, B., and Zerhouni, N., 2017), specifically those able to be inspected with vibration-monitoring technology. In this regard, Alstom has developed The Motes (Trilla, A., and Gratacòs, P., 2016, 2013), which is a network of intelligent wireless sensors that capture the vibration signature of such mechanical elements and provide feedback about their actual degradation stage, see Figure 1. These sensors have been designed to acquire vibration in different operational regimes, both on the workshop floor (low-speed environment) and in commercial service (up to high-speed rail). Ultimately, the fleet management team can take advantage of their added-value and make better informed decisions on how to schedule the various maintenance actions with the available resources. In this setting, one of their main objective components are the axle bearings, also known as axleboxes.

The axlebox is a heavy-duty safety-critical railway element (Tsui, K. L., Chen, N., Zhou, Q., Hai, Y., and Wang, W., 2015). It bears the weight of the train, minimises the friction with the rotating axle, and its failure in service might cause derailment. Therefore, its maintenance is of utmost importance to guarantee the availability of the fleet. To this end, in a predictive maintenance scenario, the collected vibration signature must be reliable and truly representative of the actual degradation of the asset. However, this often comes at the cost of transmitting a greater amount of data, i.e., its raw

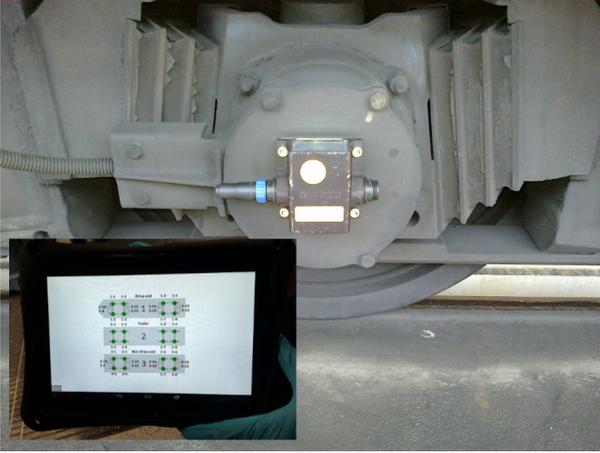


Figure 1. The Motes in use with an axlebox at the Morden depot in London for the Northern Line underground. The small window at the bottom-left corner also shows a tablet, which is used to manage the network of sensors.

signal waveform. Relaying big loads of data works against the business-case for the IIoT, especially for remote battery-powered devices, which are designed with wide-range but low-bandwidth and low-energy radio interfaces, and are expected to operate intermittently to last a long time unattended. In addition, the activity of the sensors must not delay the limited time of the maintenance staff during their inspection actions. Overall, this exposes the need to maximise the throughput of information with the smallest volume of vibration data, and to do so, this article explores the use of signal compression as a key enabler to achieve a cost-effective, robust, and easy-to-implement PHM solution (Tsui, K. L., Chen, N., Zhou, Q., Hai, Y., and Wang, W., 2015).

In the context of wireless sensor networks for diagnosing machinery, vibration signal compression has already been attained using different signal processing methods like the Discrete Cosine Transform (Alsalaet J. K., Najem, S. I., and Ali, A. A., 2012), the Empirical Mode Decomposition (Chan, J. C., and Tse, P. W., 2009), and Wavelets (Hao, W., and Jinji, G., 2012). However, the electronics used for some IIoT devices populate low-power processors that aim at the minimisation of energy consumption at the expense of featuring somewhat modest processing capabilities. Thus, implementing such costly complicated time-frequency transforms is oftentimes out of reach. In this regard, this article proposes the use of neural networks as a general-purpose function approximator because of their overall good effectiveness, and also because their industrialisation reduces to making use of linear algebra operations like matrix multiplication and vector addition, which are already widely supported by many embedded platforms. Specifically, the proposed approach focuses on using the Autoencoder neural network.

The Autoencoder is a particular layered neural architecture

that inherently learns to replicate data through a compressed representation. Its previous use in PHM highlights its capacity to detect anomalies (Goldthorpe, P., and Desmet, A., 2018) and to construct health indices (Trilla, A., Janjua, F., and Bermejo, S., 2019), among others. This article uses the compressed layer of the Autoencoder to obtain a condensed description of the raw signal waveform, which is the most critical factor in terms of transmitted data volume. Additionally, a set of vibration health features are also extracted and appended to the compressed signal to refine its eventual expanded reconstruction. The computational cost of this stage is not relevant in this context, but the amount of computed indicators must be kept to a minimum to reduce the amount of transmitted data. Finally, this array of information is quantised into a low-resolution single-byte representation to build a compact frame for the IIoT infrastructure, thus attaining the goal of transmitting a high-quality vibration signal with a fraction of the originally acquired data sample.

The article is organised as follows: Section 2 describes the distributed compression/expansion analysis procedure, including the Autoencoder technique, and the description of the railway axlebox data. Section 3 shows the results of the signal reconstruction evaluation. Section 4 discusses the overall approach, and Section 5 concludes the manuscript, reflects on its impact to the current maintenance actions, and provides avenues of future improvement.

2. METHOD

This section describes the process that has been followed to obtain a reliable vibration compression procedure.

2.1. Distributed Vibration Combanding

In order to reduce the amount of transmitted data while retaining the fundamental characteristics of the vibration signal, the whole process needs to be split into the following functions:

- **Compression** of the time-varying signal waveform and its features on the edge (i.e., the sensing device).
- **Expansion** of the compressed signal and its feature-corrected reconstruction on the user side (i.e., the cloud, or a mobile platform like a tablet).

Figure 2 shows the complete combanding procedure (note that “combanding” is the portmanteau of “compressing” and “expanding”). The specific operations performed by the edge device for the compression stage are described as follows:

1. **Data Acquisition** The sensing device equips an accelerometer that is used to obtain an instance of the vibration signature for the degrading asset (e.g., the axlebox). The dynamic range of the sensor and the sampling frequency of use are adjusted to the test conditions (i.e., at the depot or in commercial service). A sequence of

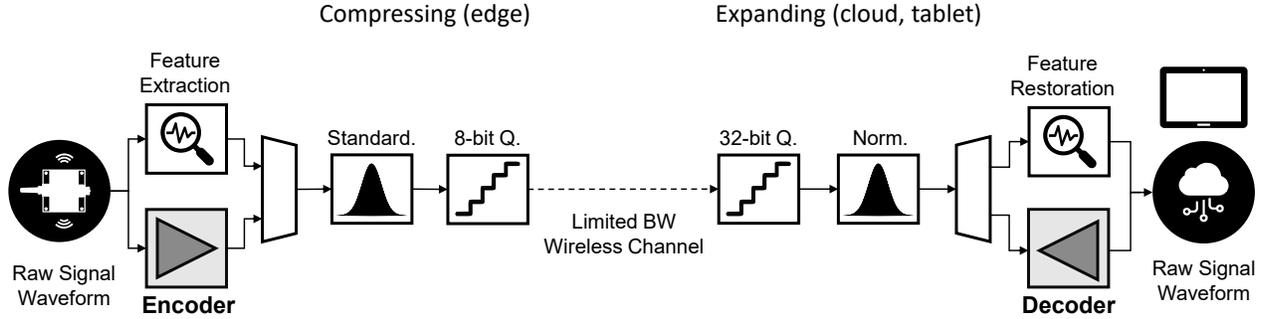


Figure 2. Diagram of distributed vibration compression and expansion processes for transmitting information over a limited bandwidth (BW) wireless channel. The role of the encoder and the decoder (both implemented with the Autoencoder) operating on the raw signal waveform (most critical data volume) is highlighted in boldface.

real-valued samples are collected; thus, a signed 32-bit floating-point arithmetic is used.

2. **Feature Extraction** An array of statistical health indicators for vibration data are extracted, e.g., peak magnitude, variance, skewness, kurtosis, crest factor, etc. (Trilla, A., Janjua, F., and Bermejo, S., 2019; Tsui, K. L., Chen, N., Zhou, Q., Hai, Y., and Wang, W., 2015). These features describe particular aspects of the asset’s degradation (e.g, driven by the failure modes).
3. **Encoding** The stream of raw vibration waveform data is segmented into short-time windows, and each of these frames is then compressed with the Autoencoder, yielding a fraction of the initial acquisition size. The next section provides further details about this operation.
4. **Standardisation** Each of the variables obtained so far (the features and the compressed vibration map) is statistically standardised so that their resulting distribution has zero mean and unit standard deviation (a Gaussian shape is also assumed), i.e., $\mathcal{N}(0, 1)$. This process is also known as Z-score normalisation.
5. **8-bit Quantisation** The resulting real values are finally rescaled so that the ultimate normal distribution is centred on the 0-255 value range. Therefore, each variable now has a $\mathcal{N}(128, 64^2)$ distribution, which is discretised and may be represented with an unsigned 8-bit integer arithmetic after a rounding operation, thus obtaining a low-resolution representation. It is to note that this final step requires the truncation of the standardised distribution to fit into the limited range of the single byte representation. The truncated range is arbitrarily set to cover 95% of the real values (i.e., 2 standard deviations).

Similarly, the specific operations performed by the end user device for the expansion stage (i.e., the cloud or a mobile platform like a tablet) reverse the process described above: first, the low-resolution samples are quantised into a real-valued 32-bit floating-point arithmetic. Then, the original variable distributions are normalised, which recovers the vi-

bration features directly. And finally, the encoded waveform values are decoded into the initial vibration signals with the Autoencoder. It is to note that this is a lossy compression procedure, so one last post-processing step is applied to ensure that the reconstruction preserves the original health features. In this work, the peak magnitude of the vibration is maintained because it is mostly indicative of the severity of the incipient failure.

2.2. Autoencoder Neural Network

The Autoencoder (AE) is a connectionist machine learning technique that replicates “essential information”. It is data-specific, so it only works with instances that are of same nature as the examples it has learnt from. To this end, it uses a self-supervised learning technique that exploits auto-association (Kramer, M. A., 1992; Stone, V. M., 2008), which is a specific mode of supervised learning where the targets are generated from the inputs. As a result, this neural network learns a distributed representation of the data that captures its meaningful attributes as its main factors of variation (Bengio, Y., 2009).

For the end-to-end vibration compression purposes that this work pursues (implemented on the edge device, and on the cloud/tablet), the design of the proposed neural network architecture is feed-forward and shallow, i.e., memoryless with one single hidden layer. This reduces both the memory footprint and the computational burden, and the resulting weights that define the behaviour of the model may be directly industrialised through a set of matrix multiplications (Goldthorpe, P., and Desmet, A., 2018). In addition, the framework of the presented Autoencoder shows a converging layout from its input dimensionality D into H at half of its depth (i.e., the encoding, compression stage), and then a diverging structure back to D toward its output (i.e., the decoding, expansion stage), see Figure 3. This undercomplete configuration forces the Autoencoder to learn the most salient features of the training data, and thus it develops a compressed repre-

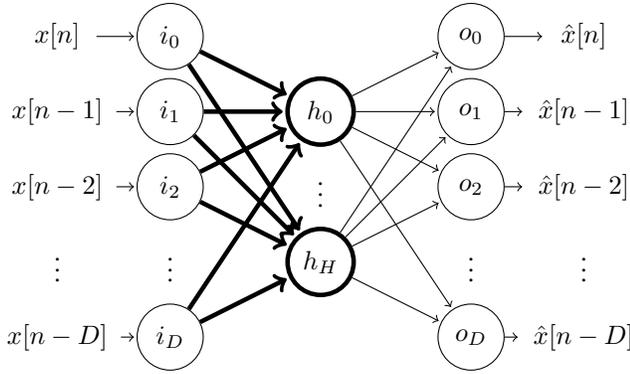


Figure 3. Compadding Autoencoder architecture. D is the input data dimensionality, and H is the size of the hidden/encoding layer, which defines the learning capacity of the neural network. The undercomplete representation is ensured as long as $H < D$. The encoder part is shown with thick arrows (along with thick states for the compressed vector), whereas the decoder is shown with thin arrows.

sensation. The amount of hidden units H in the “bottleneck” layer, which must be smaller than D in this case, defines the expressiveness of this neural network and therefore modulates its learning capacity. Additionally, if these hidden neurons apply a nonlinear activation function like a Rectified Linear Unit, the network gains the ability to capture multi-modal aspects of the input distribution (Japkowicz, N., Hanson, S. J., and Gluck, M. A., 2000), although in this case the compression transformation is essentially linear (from input to hidden layer). Obviously this Autoencoder-based approach is lossy, in the sense that the replica only retains the principal characteristics of the data, but not the details (or the noise). A greater reconstruction quality may be obtained with the identity, the principal component, or the overcomplete representations (making H equal to or greater than D), but these would clearly work against the compression objective.

The proposed Autoencoder is trained with an advanced stochastic gradient descent procedure with backpropagation following the Adam algorithm (Kingma, D. P., and Ba, J. L., 2015), which implements the weight updates through the individual estimation of the first and second statistical moments of the gradients (i.e., a momentum on the gradient and its squared value). The specific hyperparameters of use are: a learning rate α of 0.001, a first momentum β_1 of 0.9, and a second momentum β_2 of 0.999. The average root mean square (RMS) error between the reconstruction and the original vibration signal is used as the objective cost function, i.e., $(\hat{x}[n] - x[n])^2$. This conventional optimisation protocol still has room for some improvements through regularisation penalties, yielding different Autoencoder solutions. These refinements are described hereunder.

2.2.1. Ordinary AE

This Autoencoder is directly trained to compress the input into some lower-dimensional representation so that the *exact same input* may thereafter be reconstructed, without any further constraint. This is analogous to a maximum-likelihood estimation of the optimum weights, and therefore it is subject to overfitting. Obviously, some kind of regularisation strategy would be desirable here, but the Ordinary AE does not contemplate it explicitly; this model only relies on the limited representational capacity of the undercomplete architecture. However, this work also exploits the advantage of limiting the number of epochs during training, because gradient descent with early stopping is similar to a squared Euclidean norm regularisation of the weight parameters (Zinkevich, M., 2003). This strategy generalises the performance of the resulting Autoencoder.

2.2.2. Denoising AE

Another strategy for regularising the Autoencoder is by stochastically corrupting the vibration signal input with noise, while the original uncorrupted signal is still used as target for the reconstruction. This method is known as the Denoising AE (Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and Manzagol, P.-A., 2010). This approach learns to preserve the statistical regularities of the input vibration signal, and to undo the random corruption, which can take different forms:

- **Additive White Gaussian Noise (AWGN)** The addition of wideband noise is inspired by many natural processes, and its Gaussian amplitude distribution is driven by the central limit theorem of probability theory when many random processes interact. This is a basic noise model used in information theory, and this work regards its useful convenience for the corruption of the input.
- **Masking** The random setting of some inputs to zero is also a successful regularisation method. This occlusion strategy forces the Autoencoder to deal with data that contains missing values. This is an interesting property because it regards the Autoencoder as a generative model.

2.2.3. Sparse AE

Another strategy for regularising the Autoencoder is via the sparsity in the encoding space. The Sparse AE (Makhzani, A., and Frey, B., 2014) offers an alternative method for constraining the amount of information that may traverse the network and thus require a learned compression of the input data, without reducing the number of hidden units. This Autoencoder adds a sparsity penalty on the activation of the hidden layer so that only a few units may operate at a given time (the correction is increased with the amount of contribution). In this approach, the network gets selective and sensitive to individual hidden units toward specific attributes of the input

vibration data. This sparsity cost is attained by computing the average activations in the hidden layer, and then scoring the Kullback-Leibler divergence between a Bernoulli random variable with this mean value, and another one with a desired small sparse average value.

2.2.4. Contractive AE

There is yet another strategy for regularising the Autoencoder considered in this work that is known as the Contractive AE (Rifai, S., Vincent, P., Muller, X., Glorot, X., and Bengio, Y., 2011). In this approach the Autoencoder is trained so that the derivatives of the hidden layer activations are small with respect to the input. This prevents that small changes in the input may lead to large changes in the encoding space, so in a sense it adds robustness to small perturbations around the data. This effect is attained by introducing a penalty term in the cost function that corresponds to the Frobenius norm of the Jacobian matrix of the encoder activations with respect to the input. It is shown that this results in a localised space contraction, which in turn yields robust features on the activation layer.

2.3. Vibration Data and Stream Processing

In the present PHM environment, real-time data exchange is not necessary because the gradual degradation of mechanical assets like axleboxes does not occur in a short time. Thus, The Motes operate with asynchronous connectivity (Boyes, H., Hallaq, B., Cunningham, J., and Watson, T., 2018). However, the compression feature of the Autoencoder is limited to its input dimensionality D . In order to transmit a whole “long” vibration signal as a stream, the original sequence needs to be buffered and segmented into windows of length D , then compressed into vectors of length H , and finally be transmitted sequentially in the payload of the wireless protocol frames for the available interfaces, e.g., Wi-Fi, ZigBee, Bluetooth LE, or LoRa.

To evaluate the effectiveness of the companding method with the Autoencoder, this work uses a dataset of axlebox vibration data acquired for a metro stock, rolling at 5mph, on a straight level test track, in the depot. Each acquisition comprises a waveform of 4 seconds sampled at 3200Hz. The complete dataset includes over 28000 instances of vibration segments (with 500 samples each) divided into different degradation levels (Trilla, A., Janjua, F., and Bermejo, S., 2019), i.e., good, regular, and bad condition.

3. RESULTS

This section compares the different Autoencoder strategies to determine which of them yields the best companding effectiveness for the IIoT, i.e., the maximum compression with the minimum loss. Their performance is estimated with a round of stratified random subsampling with 5% of the instances

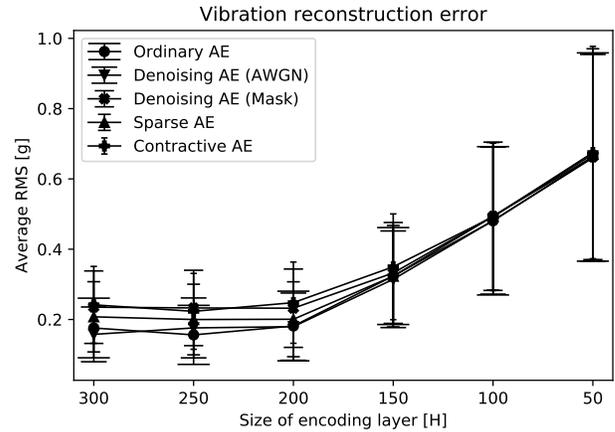


Figure 4. Autoencoder reconstruction error with respect to the size of the hidden/encoding layer (H). The points correspond to the mean value of the RMS error distribution (assuming Gaussian normality), and the whiskers correspond to one standard deviation. Note that the AE strategy of use may be distinguished by the shape of the points and the size of the error caps.

(i.e., around 1400) for testing. Figure 4 shows how the size of the hidden/encoding layer impacts the reconstruction error of the test signals for each AE approach.

In general, it can be seen that regardless of the regularisation strategy of use, all approaches display a flat constant error response down to 200 hidden units (with a greater or lesser offset), and a linear increasing slope beyond that inflection point (also increasing the variability). The interpretation that follows for this effect is that down to 200 hidden units the Autoencoder generalises well, but further compression limits its representational capacity to a point that the neural network underfits the data and so exhibits a steady increase of the reconstruction error. Additionally, it is the Ordinary Autoencoder, which only relies on the undercomplete representation for regularising its performance, the one that attains the lowest reconstruction error. When an additional regularisation strategy is applied, the resulting “over-regularised” Autoencoder diminishes its ability to adapt and converge to a better solution. Taking the inflection point at $H=200$ hidden units as the reference (with input $D=500$), the difference between the least performing strategy (i.e., the Contractive AE, with $\mathcal{N}(0.2479, 0.1156^2)$) and the best (i.e., the Ordinary AE, with $\mathcal{N}(0.1815, 0.0991^2)$) is statistically significant with a confidence interval of 95% using an Independent Samples t-test.

It is to note that this reconstruction performance is averaged over all test instances, which belong to different condition categories. In order to shed some light into this particular aspect, Figure 5 shows the distribution of error values regarding the degradation of the test assets for the best-performing companding strategy, i.e., the Ordinary Autoencoder with 200

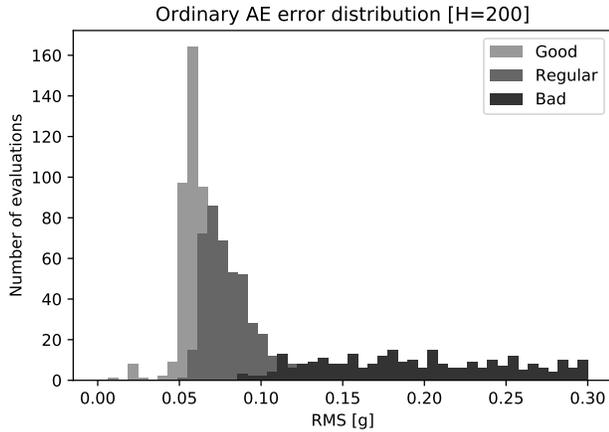


Figure 5. Ordinary AE (with input $D=500$, and compressed output $H=200$) reconstruction error with respect to degraded test asset conditions (good, regular, and bad).

hidden units. This graph makes it clear that as the axleboxes degrade, the reconstruction accuracy of the Autoencoder decreases, and that happens precisely for the most critical situations, when warnings and alarms possibly need to be raised (i.e., for the bad condition). That's why it is of utmost importance to take into account the health features to refine the reconstruction of the waveforms. This loss of reconstruction performance with the progress of the degradation is probably caused by the increased dynamic range and non-stationarity of the signals. In addition, the shape of this distribution questions the previous normality assumption, so the former results must only be taken as indications.

Finally, the transformation of a window of 500 vibration samples into a condensed vector of 200 points yields a compression rate of 2.5, and the 8-bit quantisation that follows applies another rate of 4. Therefore, the final compression rate is of 10, and the resulting system displays a good (almost lossless) companding performance. Figure 6 and Figure 7 show how the Ordinary Autoencoder reconstructs a vibration signal in the worst-case scenario: foreshadowing a failure (the original signal belongs to the "bad" axlebox condition). It can be seen that the time waveform preserves the amplitude that signals the severity of the degradation, and the frequency spectrum retains the location of the source of the failure, so the signal compression process does not modify the result of the analysis that would be obtained with the original raw data. In the healthy case, where the discrepancy between the original waveform and its reconstruction is even smaller, a complete overlap is visually observed, with a signal amplitude an order of magnitude smaller. Consequently, the Ordinary Autoencoder approach enables a fine-grained diagnosis through IIoT monitoring technology.

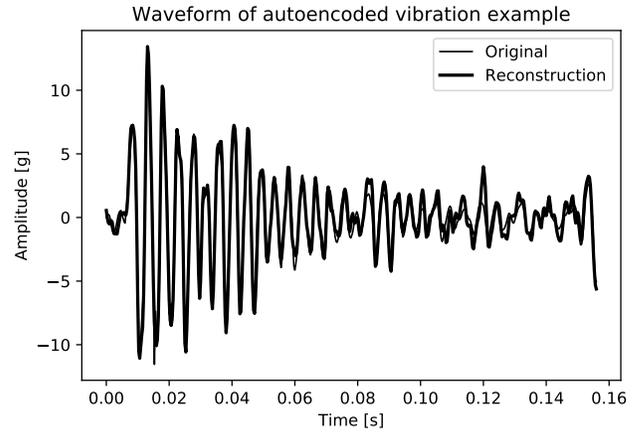


Figure 6. High-peaked non-stationary time waveform of an autoencoded vibration signature showing a bad condition (Ordinary AE with $D=500$ and $H=200$).

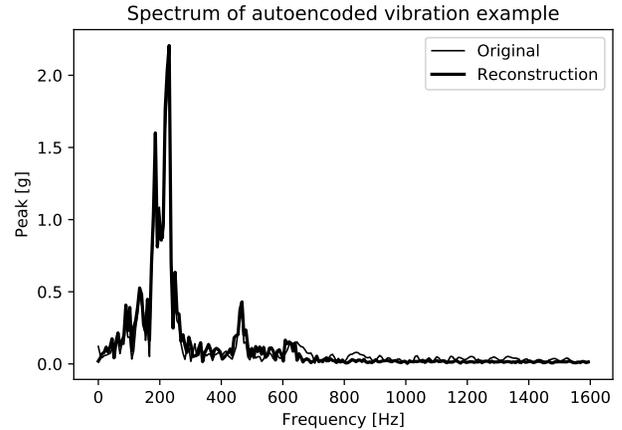


Figure 7. Frequency spectrum of the autoencoded vibration signature shown in Figure 6.

4. DISCUSSION

By trying to approximate the identity function with an under-complete representation, the Autoencoder attains a flexible compression strategy that significantly reduces the amount of data to be transmitted. However, the Autoencoder is not usually considered to be a good compressor in the conventional broad sense, because it lacks the versatility to be applied to data of arbitrary nature. It doesn't operate by exploiting the redundancy in the data to build efficient codewords, so perhaps its performance is limited by this aspect. Nevertheless, it is to note that the compressed layer of the Autoencoders studied in this work corresponds to the linear components of the vibration signals (Duda, R. O., Hart, P. E., and Stork, D. G., 2001), and on that space a clustering technique followed by vector quantisation could still be applied to obtain such an encoded codebook of principal centroids (despite possi-

bly preventing the detection of anomalies in this latent representation). Additionally, the low-resolution quantisation step presented in this work is linear, and a more effective procedure might be obtained with a nonlinear quantiser enhancing the main concentration of data in the feature distribution.

In order to better understand the internal behaviour of the Autoencoder beyond the mapping, other strategies have also been considered, like the use of convolutions and filters. Inspired by the suggestion that the architecture of the neural network is more important than the values of the weights (Gaier, A., and Ha, D., 2019), the use of pairwise correlations has been studied to exploit sparse time dilations like WaveNet (Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., and Kavukcuoglu, K., 2016) and Time-Delay Neural Networks (Peddinti, V., Povey, D., and Khudanpur, S., 2015). In the case that the vibration waveform gets averaged as if by the use of a low-pass filter, the fundamental signal behaviour is retained, but the Autoencoder increases the reconstruction error with an offset. Similarly, the same result is obtained if the input waveform is down-sampled to enhance the details contained in the high-frequency components. In both cases, though, the performance inflection point at 200 hidden units is equally obtained. Therefore, it seems that the densely layered Autoencoder eventually learns the most effective signal transformation, but as the compression rate is incremented, the reconstruction is increasingly smoothed (Trilla, A., Janjua, F., and Bermejo, S., 2019).

Finally, it is to note that the current compression is obtained with a linear combination of the input vibration samples, which is similar to the data-driven measurement matrix that may be developed in compressed sensing (Wu, S., Dimakis, A. G., Sanghavi, S., Yu, F. X., Holtmann-Rice, D., Storcheus, D., Rostamizadeh, A., and Kumar, S., 2019). The recent state of the art applied to vibration signals (which also involves frequency considerations) obtains compression rates up to 5 (Premanand, B., and Sheeba, V. S., 2020), whereas the approach described in this contribution reaches rates of 10 with the same error. However, the inclusion of an additional hidden layer before (and after) the current encoding would lead to an intricate nonlinear representation, potentially smaller than 200 units, and therefore increase the current compression rate. The universal approximation theorem suggests that this is possible (Cybenko, G., 1989), but it has not been explored in this work to minimise the processing especially on the edge device. In a similar vein, the space complexity is also to be considered in an embedded Machine Learning environment given the limited memory of some microcontrollers (Warden, P., and Situnayake, D., 2020). The largeness of the encoding matrix, thus, may be a limiting factor of the industrial deployment of this solution. Nonetheless, this size may be conveniently reduced by shortening the length of the input buffer while keeping the same compres-

sion rate at the expense of increasing the running time, e.g., compressing 250 vibration samples into 100 (instead of 500 into 200) maintains the same representational capacity with a quarter of the original matrix size (in number of weights), and it takes twice as much to complete the processing.

5. CONCLUSIONS

The use of the activation in the hidden/encoding layer of an Ordinary Autoencoder with an undercomplete representation along with a low-resolution quantisation step, significantly reduces the amount of vibration data to be transmitted through an IIoT monitoring network. With compression rates up to 10, the high quality of the reconstructed signal waveforms permits implementing a fine-grained diagnosis. The proposed approach reduces the needed bandwidth for the transmission, and/or shortens the download time for each acquisition. Also, its impact speeds up the maintenance cycle on the workshop floor, and/or increases the inspection frequency on remote locations.

The future work that is currently envisaged opens up two main fronts. On the one hand, exploring the use of complex numbers to obtain a richer representational capacity of the underlying neural network (Trabelsi, C., Bilaniuk, O., Zhang, Y., Serdyuk, D., Subramanian, S., Santos, J. F., Mehri, S., Rostamzadeh, N., Bengio, Y., and Pal, C. J., 2018). And on the other hand, developing a deep network pruning strategy to facilitate its implementation on embedded systems with limited hardware resources (Han, S., Mao, H., and Dally, W. J., 2016).

ACKNOWLEDGMENT

We would like to show our gratitude to Professor Xavier Vilasís-Cardona for his valuable support and comments that greatly improved the manuscript.

REFERENCES

- Alsalaet J. K., Najem, S. I., and Ali, A. A. (2012). Vibration data compression in wireless sensors network. *Proc. of the International Conference on Signal Processing, Communication and Computing*, 717–722.
- Atamuradov, V., Medjaher, K., Dersin, P., Lamoureux, B., and Zerhouni, N. (2017). Prognostics and Health Management for Maintenance Practitioners - Review, Implementation and Tools Evaluation. *International Journal of Prognostics and Health Management*, 8(60), 1–31.
- Bengio, Y. (2009). Learning Deep Architectures for AI. *Foundations and Trends in Machine Learning*, 2(1), 1–71.
- Boyes, H., Hallaq, B., Cunningham, J., and Watson, T. (2018). The industrial internet of things (IIoT): An

- analysis framework. *Computers in Industry*, 101, 1–12.
- Chan, J. C., and Tse, P. W. (2009). A Novel, Fast, Reliable Data Transmission Algorithm for Wireless Machine Health Monitoring. *IEEE Transactions On Reliability*, 58(2), 295–304.
- Cybenko, G. (1989). Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals and Systems*, 2, 303–314.
- Duda, R. O., Hart, P. E., and Stork, D. G. (Ed.). (2001). *Pattern Classification*. Wiley-Interscience.
- Gaier, A., and Ha, D. (2019). Weight Agnostic Neural Networks. *Proc. of the 33rd Conference on Neural Information Processing Systems*, 1–19.
- Goldthorpe, P., and Desmet, A. (2018). Denoising autoencoder anomaly detection for correlated data. *Proc. of the Fourth European Conference of the Prognostics and Health Management Society*, 1–6.
- Han, S., Mao, H., and Dally, W. J. (2016). Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding. *Proc. of the International Conference on Learning Representations*, 1–14.
- Hao, W., and Jinji, G. (2012). The research of optimal selection method for wavelet packet basis in compressing the vibration signal of a rolling bearing in fans and pumps. *Proc. of the 25th International Congress on Condition Monitoring and Diagnostic Engineering, Journal of Physics: Conference Series*, 364(012033), 1–13.
- Japkowicz, N., Hanson, S. J., and Gluck, M. A. (2000). Non-linear autoassociation is not equivalent to PCA. *Neural Computation*, 12(3), 531–545.
- Kingma, D. P., and Ba, J. L. (2015). Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs.LG]*, 1–15.
- Kramer, M. A. (1992). Autoassociative Neural Networks. *Computers and Chemical Engineering*, 16(4), 313–328.
- Makhzani, A., and Frey, B. (2014). k-Sparse Autoencoders. *Proc. of the International Conference on Learning Representations*, 1–9.
- Oord, A., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., and Kavukcuoglu, K. (2016). WaveNet: A Generative Model for Raw Audio. *arXiv:1609.03499 [cs.SD]*, 1–15.
- Peddinti, V., Povey, D., and Khudanpur, S. (2015). A Time Delay Neural Network Architecture for Efficient Modeling of Long Temporal Contexts. In *Interspeech* (pp. 3214–3218).
- Premanand, B., and Sheeba, V. S. (2020). Compressed encoding of vibration signals using extremum sampling. *Springer Nature Applied Sciences*, 2(1261), 1–10.
- Rifai, S., Vincent, P., Muller, X., Glorot, X., and Bengio, Y. (2011). Contractive Auto-Encoders: Explicit Invariance During Feature Extraction. *Proc. of the 28th International Conference on Machine Learning*, 833–840.
- Stone, V. M. (2008). The auto-associative neural network - a network architecture worth considering. *Proc. of the 2008 World Automation Congress*, 1–4.
- Trabelsi, C., Bilaniuk, O., Zhang, Y., Serdyuk, D., Subramanian, S., Santos, J. F., Mehri, S., Rostamzadeh, N., Bengio, Y., and Pal, C. J. (2018). Deep Complex Networks. *Proc. of the International Conference on Learning Representations*, 1–19.
- Trilla, A., and Gratacòs, P. (2013). Condition based maintenance on board. *Chemical Engineering Transactions Journal*(33), 733–738.
- Trilla, A., and Gratacòs, P. (2016). Maintenance of bogie components through vibration inspection with intelligent wireless sensors: a case-study of axle-boxes and wheel-sets using the Empirical Mode Decomposition technique. *Proc. of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*(230), 1408–1414.
- Trilla, A., Janjua, F., and Bermejo, S. (2019). Developing a Hybrid Expert/Data-Driven Health Index for Railway Axleboxes Using Auto-encoder Neural Networks. *Proc. of the Prognostics and System Health Management Conference*, 1–6.
- Tsui, K. L., Chen, N., Zhou, Q., Hai, Y., and Wang, W. (2015). Prognostics and Health Management: A Review on Data Driven Approaches. *Mathematical Problems in Engineering*, 2015(793161), 1–18.
- Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., and Manzagol, P.-A. (2010). Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 11, 3371–3408.
- Warden, P., and Situnayake, D. (2020). *TinyML: Machine Learning with Tensorflow Lite on Arduino and Ultra-Low-Power Microcontrollers*. UK: O’Reilly UK Ltd.
- Wu, S., Dimakis, A. G., Sanghavi, S., Yu, F. X., Holtmann-Rice, D., Storcheus, D., Rostamzadeh, A., and Kumar, S. (2019). Learning a Compressed Sensing Measurement Matrix via Gradient Unrolling. *Proc. of the 36th International Conference on Machine Learning*, 1–17.
- Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. *Proc. of the 20th International Conference on Machine Learning*, 928–936.

BIOGRAPHIES

Alexandre Trilla graduated from La Salle University of Barcelona with a M.Sc. in Electronics and Telecommunications Engineering in 2008, and a M.Sc. in IT Manage-

ment in 2010. He has an academic research background in spoken language processing, and an industrial research background in PHM. He has authored several publications in scientific conferences and journals (IEEE Transactions on Audio, Speech, and Language Processing, Chemical Engineering Transactions, and the Journal of Rail and Rapid Transit). At present, he is a Senior Data Scientist and R&D Program Manager at Alstom, working on the deployment of PHM to the railway environment. He leads the development of predictive maintenance based on Machine Learning, and he is especially interested in the solutions with artificial neural networks.

Dr. David Miralles holds a degree on Theoretical Physics of the University of Barcelona (1995). From 1996 to 2001 he worked at the Fundamental Physics Department at the same university. In 2001 he obtained a PhD on Mathematical Physics. From 2001 to 2007 he worked at the Depart-

ment of Communication and Signal Theory of Ramon Llull University (URL). He has made several stays in international centres: Instituto de Matemática, Estatística e Computação Científica, Campinas, (Brazil, 1998); International Center of Theoretical Physics, Trieste (Italia, September 2004); Observatoire de Paris (France, March 2005-06) and MIT Media Lab, Cambridge (USA, May 2014). He is member of Grup de Recerca en Tecnologies Media where he leads the Interaction area; a research team focused on designing new interactions between people and objects.

Verónica Fernández graduated from Universitat Autònoma of Barcelona with a B.Sc. in Chemical Engineering in 2016, and a M.Sc. in Automatic Systems and Industrial Electronics Engineering from Universitat Politècnica of Catalonia in 2019. At present, she is a Data Scientist working on PHM in the railway environment.