Multi-objective variational autoencoder: an application for smart infrastructure maintenance

Ali Anaissi[1] · Seid Miad Zandavi[1] · Basem Suleiman[1] · Mohamad Naji[2] · Ali Braytee[2]

Abstract

Multi-way data analysis has become an essential tool for capturing underlying structures in higher-order data sets where standard two-way analysis techniques often fail to discover the hidden correlations between variables in multi-way data. We propose a multi-objective variational autoencoder (MO-VAE) method for smart infrastructure damage detection and diagnosis in multi-way sensing data based on the reconstruction probability of autoencoder deep neural network (ADNN). Our method fuses data from multiple sensors in one ADNN at which informative features are being extracted and utilized for damage identification. It generates probabilistic anomaly scores to detect damage, asses its severity and further localize it via a new localization layer introduced in the ADNN. We evaluated our method on multi-way laboratory-based and real-life structural datasets in the area of structural health monitoring for damage diagnosis purposes. The data was collected from our deployed data acquisition system on a cable-stayed bridge in Western Sydney, a reinforced concrete cantilever beam which replicates one of the major structural components on the Sydney Harbour Bridge and a laboratory based building structure obtained from Los Alamos National Laboratory (LANL). Experimental results show that the proposed method can accurately detect structural damage. It was also able to estimate the different levels of damage severity, and capture damage locations in an unsupervised aspect. Compared to the state-of-the-art approaches, our proposed method shows better performance in terms of damage detection and localization.

Keywords Autoencoder neural network · Multi-way data · Structural health monitoring · Damage detection · Data fusion

1 Introduction

The concept of smart infrastructure maintenance emerged in the recent years as a continuous automated process known as structural health monitoring (SHM). It aims to build a condition-based inspection system driven by data for early damage identification which results in better life-safety and economic benefits. Most of the current structural maintenance's approaches are considered as a time-based visual inspection which often follows a predefined regular schedule. This kind of time-based inspection for a such structure may results in certain economic and potential life losses if it was too late or too early. Moreover, some

Seid Miad Zandavi, Basem Suleiman, Mohamad Naji and Ali Braytee are contributed equally to this work.

✉ Ali Anaissi
   ali.anaissi@sydney.edu.au

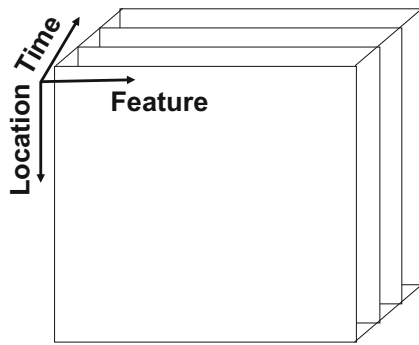Extended author information available on the last page of the article.

structures such as high bridges raise other challenges in terms of accessibility. SHM has earned a lot of interests during the last decade due to the fact that it leads to enhance understanding the behaviour of infrastructure and increases its life span whilst maintaining a high level of life-safety.

In the realm of data science, SHM has attracted many researchers working in the areas of machine learning and data mining to handle the wealth of vibration responses being simultaneously measured over time by many sensors attached to a structure at different locations, and further to identify structural damage. These measured responses lead to high dimensional, multi-way and correlated data which raises many challenges in analyzing and extracting informative features to learn a damage identification model. The SHM sensing data can be arranged as a three-way data (feature × location × time) as described in Fig. 1. Feature is the information extracted from the raw signals in time domain (e.g. features in frequency domain). Location represents sensors, and time is data snapshots at different timestamps. Each cell is a feature value extracted from a particular sensor at a certain time.

Fig. 1 Multi-way data with three modes in SHM applications

Rytter classified damage identification into four different levels of complexity [1]: damage detection (level 1), localization (level 2), severity assessment (level 3) and failure prediction (level 4). The damage detection level can be solved using two-way analysis techniques by constructing a standard anomaly detection model. However damage localization and severity assessment require multi-way data analysis techniques to capture the physical meaning of the structure. On the other hand, level 4 is not considered as a machine learning problem since it requires understanding of the physical characteristics of the damage progression in the structure. These requirements have motivated us to study the deep neural networks (DNN) as a feature learning method to handle the complexities associated with the multi-way SHM data. DNN has become popular and attracted many researchers working in the area of data analytic. It has been successfully applied to solve complex pattern recognition problems such as vision [2] and speech [3]. Sutskever et al. [4] claim that DNN often produces powerful models that achieve high performance in comparison to other state-of-the-art machine learning algorithms.

Generally speaking, data instances from at least two different classes are required for the training stage in DNN. However, in many applications such as SHM [5], only data instances from one state (i.e. undamaged or healthy) are available, and the samples from other states (i.e. damaged), if not impossible, are too difficult or costly to acquire. Thus, the classification process becomes as an anomaly detection problem. Anomaly detection methods build a model based on a given positive training dataset, and for a new arrived data instance, the model estimates the agreement between the new instance and the trained model. Data instances which do not fit into the trained model are classified as anomaly [6].

In the context of anomaly detection, an autoencoder deep neural networks (ADNN) model may be more practical when only data from positive/normal states are available. It was originally proposed for dimensionality reduction

problems. However, it has proved throughout several applications that it was very capable to handle the case of one class learning and solve anomaly detection problems. Furthermore, it can be also utilized as a data fusion structure which can construct an internal representation for input data collected from multiple sources and then extracts anomalous sensitive features. Recently, Anaissi and Zandavi [7] use ADNN to propose a multi-object autoencoder for fault detection and diagnosis in higher-order data based on the reconstruction error of ADNN.

This paper is an extension of the aforementioned work in [7]. In combining with the multi-objective autoencoder in [7], this paper utilizes the variational autoencoder method to propose a multi-objective variational autoencoder (MO-VAE) deep neural network for damage detection, localization and severity assessment. In contrast to [7], MO-VAE performs damage detection based on reconstruction probability but not reconstruction error. It performs data fusion by taking a frontal slice from a multi-way training data. Stochastic gradient descent algorithm is then used to learn reconstructions that are close to its original input slice followed by constructing a sensor identity matrix which will be used for damage localization. For each new incoming data slice we calculate its anomaly score based on reconstruction probability which further used for damage assessment. The sensor identity matrix is finally utilized to locate the identified damage.

This work is part of our broader efforts to apply data-driven SHM approaches to real bridges in operation, including the Sydney Harbour Bridge (SHB). We extensively evaluated our proposed method on laboratory-based and real-life structures datasets. The evaluation shows that MO-VAE model has the capability to perform data fusion and extract damage sensitive features which were able to accurately detecting damage. The reconstruction probability also demonstrates the ability to localize the detected damage. It further reflects the fact that it has the potential to estimate the severity of damage by analyzing the obtained reconstruction probability values. The contributions of this paper are as follows.

1. Sensing multi-way data are fused using ADNN to efficiently extract damage sensitive features and then learn reconstruction of the original input .
2. Damage detection is accomplished using reconstruction probability which has the capability to identify damage without using any preset fixed threshold parameter.
3. Damage localization is accomplished using a new layer introduced in ADNN.
4. Experiments using data obtained from laboratory-based and real-life structures datasets show the effectiveness of our approach in damage identification and localization.

The remainder of this paper is structured as follows. Section 2 reviews some related work. Section 3 describes our novel MO-VAE method for learning reconstruction error and localizing anomalous data, while Section 4 presents our experimental results and evaluations. Finally, Section 5 discusses the contributions, future work and concludes this paper.

2 Related work

Anomaly detection methods have been employed in many application domains such as damage detection in civil structure [8–11], intrusion detection in network [12, 13] and numerous other fields. They are mainly proposed to handle the cases when only normal/positive data are available. For instance, [14] designed a robust one-class support vector machine (OCSVM) to eliminate the influence of outliers to the learned boundary and used it to detect damage in a simulated structure. Mahadevan and Shin et al. [14] and [15], proposed an approach for fault detection and diagnosis using OCSVM and SVM-recursive feature elimination. Further, the authors [14] and [15] used OCSVM to detect damage in a rotating machinery and the results showed that the performance of the proposed method is superior to the state-of-the art methods. However, the work above focused on damage detection using two-way matrix data generated via individual sensor which might help in detecting damage but not in assessing its severity or localize it.

In the recent years, various data fusion methods have been used in SHM applications to deal with the multi-way data [16–18]. Some of these methods performed data fusion in an unsophisticated manner by simply concatenating features obtained from different sensors [16]. However, more advanced methods including principle component analysis (PCA), neural networks and Bayesian methods have been adopted at this level [19]. In this context, khoa et al. [20] used advanced tensor analysis to fuse data from multiple sensors followed by constructing a OCSVM model for damage detection. The authors were able to successfully detect and assess the severity of the damage but failed to localize it.

With the advent of deep learning methods, ADNN attracted many researchers working in the area of anomaly detection due its promising achievements in many domains [21–23]. Jinwon and Sungzoon [24] propose a variational autoencoder (VAE) for anomaly detection tasks. They used a probability measure to generate the anomaly score instead of reconstruction error. The work in [25] also uses autoencoders for anomaly detection in videos. The authors evaluate their method on real-world datasets and reported better performance over other state-of-the-art methods. The authors in [26] use deep learning methods to hierarchically learn features from the sensor measurements of exhaust gas temperatures. Then they use the learned features as input to an ADNN for performing combustor anomaly detection.

Further, Akcay et al. [27] propose a model composed of generative adversarial networks (GANs) and encoder-decoder-encoder sub-networks which is known as (GANomaly) . The aim of the model is to minimise the distance between real, generator images and their latent representation. While the author in [28] propose a Skip-GANomaly by using an encoder-decoder convolution neural network (CNN) with skip connection . The enhancement added to [27] is in the generator network to cope with higher resolution image. The author in [29] use a CNN based on decision-tree learning to propose an anomaly detection algorithm to detect a threat in X-ray cargo image. The work in [30] propose an end-to-end trainable consist of Convolutional Long Short-Term Memory (Conv-LSTM) networks which is known as (AnoGAN). The model is able to predict the evolution of video sequence from a limited number of input frames. The author in [31] develop an EGBAD model which is based on GAN at the same time to learn the encoder during training which is used for image anomaly detection.

In fact, there are still few works in which researchers try to apply ADNN methods to other data analytic tasks such as data fusion in multi-way datasets. In this study, we propose a MO-VAE deep neural network as a data fusion method to extract damage sensitive features from three-way measured responses and to perform damage detection based on the reconstruction probability. Further, the average distance between the anomaly scores of each corresponding sensor nodes are used as an another measure to localize and assess the severity of structural damage.

3 Background

3.1 Autoencoder deep neural network

Autoencoder deep neural network is an unsupervised learning process which has the ability to learn from one class data. It is an extension to the deep neural network which is basically designed for supervised learning when the class labels are given with the training examples. The rational idea of an autoencoder is to force the network to learn a lower dimensional space Z for the input features X, and then try to reconstruct the original feature space to $\hat{X}$. In other words, it sits the target values to be approximately equal to its original inputs. In this sense, the main objective of autoencoders is to learn reproducing input vectors $\{x_1, x_2, x_3, \ldots, x_m\}$ as outputs $\{\hat{x}_1, \hat{x}_2, \hat{x}_3, \ldots, \hat{x}_m\}$. Figure 1 illustrates the architecture of ADNN composed of L hidden layers ($L = 3$ for simplification). Layer X is the input layer which encoded into the middle layer

Z, and then decoded into the output layer $\hat{X}$. Each layer consists from a set of nodes denoted by circle in Fig. 2. The nodes in the input layer represents the input features which are often aligned with the number of features for a given dataset. However, the number of nodes in the hidden layer(s) are selected by user. In contrast to the traditional neural network, the number of nodes in the output layer are aligned with the same number of the input layers.

The learning process of ADNN successively computes the output of each node in the network. For a node i in layer l we calculate an output value $z_i^{(l)}$ obtained by computing the total weighted W_{ij} sum of the input values plus the bias term b_i using the following equation:

$$z_i^{(l)} = \sum_{j=1}^{n} W_{ij}^{(l-1)} a_j^{(l-1)} + b_i^{(l)} \tag{1}$$

The parameter W is the coefficient weight written as W_{ij} when associated with the connection between node j in layer $l-1$, and node i in layer l. The b_i parameter is the bias term associated with the node i in layer l and $a_j^{(l-1)}$ is the output value of node j in layer $l-1$. The resultant output is then processed through an activation function denoted by $a_i^{(l)}$, and it is defined as follows:

$$a_i^{(l)} = f(z_i^{(l)}) \tag{2}$$

Intuitively, in the input layer $a^1 = x$, and in the output layer, $a^3 = \hat{x}$. The most common activation functions in the hidden layers are the sigmoid and hyperbolic tangent defined in (3) and (4), respectively. However, in autoencoder settings a linear function is used in the output layer since we don't scale the output of the network to a specific interval ($[0, 1]$ or $[-1, 1]$).

$$f(z) = \frac{1}{1 + e^{-z}} \tag{3}$$

$$f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \tag{4}$$

Lets say that an autoencoder is composed of two systems known as encoder $g(\theta)$ and decoder $f(\phi)$. The encoder maps an input vector X to a latent vector Z. Then the decoder maps Z back to the original input feature $\hat{X}$. The autoencoder uses back propagation algorithm to learn the parameters (θ, ϕ). In each iteration of the training process, we perform a feedforward pass which successively computes the output values $a_i^{(l)}$ for all layer's nodes. Once completed, we calculate the cost error $J(\theta, \phi)$ using (5) and

then propagate it backward to the network layer.

$$J(\theta, \phi) = \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{2} \|x^{(i)} - \hat{x}^{(i)}\|^2 \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{n} \|x^{(i)} - f_\theta(g_\phi(\hat{x}^{(i)}))\|^2 \right) \tag{5}$$

In this setting, we perform a stochastic gradient descent step to update the learning parameters (θ, ϕ). This is done by computing the partial derivative of the cost function $J(\theta, \phi)$ (defined in (5)) with respect to θ and ϕ as follows:

$$\theta := \theta - \alpha \frac{\partial}{\partial \theta} J(\theta, \phi) \tag{6}$$

We update ϕ in the same way. The complete steps are summarized in Algorithm 1.

Input: A set of n positive samples $x = \{x^{(i)}\}_{i=1}^{n}$

- Initialize (θ, ϕ) to a small random value $\mathcal{N}(0, \epsilon^2)$

repeat For $i = 1 : n$

- Perform a feedforward pass to compute all nodes activations using (1) and (2).
- Compute the sum of reconstruction cost error $J(\theta, \phi)$ using (5)
- Update (θ, ϕ) using (6)

until convergence of parameters θ, ϕ
Output: encoder $f(\theta)$, decoder $g(\phi)$.

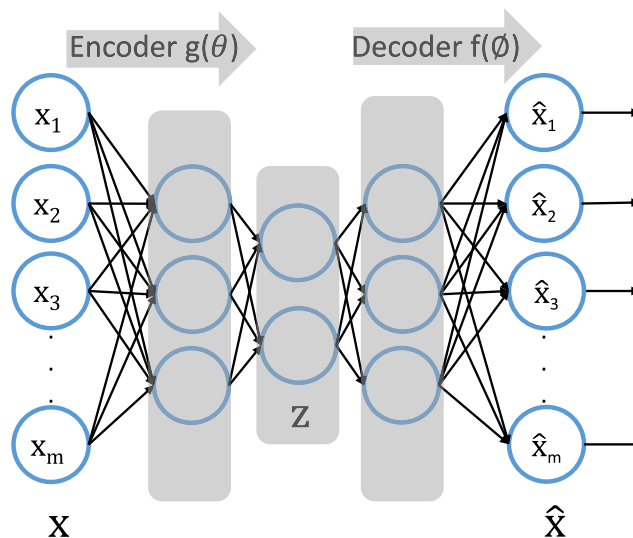---

Algorithm 1 Autoencoder training algorithm.

Fig. 2 Autoencoder neural network architecture

Once the autoencoder get trained, the network will be able to reconstruct an new incoming positive data, while it fails with anomalous data. This will be judged based on the reconstruction error (RE) which is measured by applying the Euclidean norm to the difference between the input and output nodes as shown in (7).

$$RE(x) = \|x^{(i)} - \hat{x}^{(i)}\|^2 \tag{7}$$

The measured value of RE is used as anomaly score for a given new sample. Intuitively, examples from the similar distribution to the training data should have low reconstruction error, whereas anomalies should have high anomaly score. Algorithm 2 shows the process of anomaly detection based on the reconstruction error of autoencoders.

Input: A set of new arrived samples $x = \{x^{(i)}\}_{i=1}^{n}$, and α

- $\theta, \phi \leftarrow$ Algorithm 1

For $i = 1 : n$

- Perform a feedforward pass to compute all nodes' activations using (1) and (2).
- Compute RE using (7)
- **if** $RE > \alpha$
  - $x^{(i)}$ is an anomaly
- **else**
  - $x^{(i)}$ is normal

Algorithm 2 Autoencoder anomaly detection algorithm.

4 Multi-objective variational autoencoder

We propose a multi-objective variational autoencoder (MO-VAE) neural network for damage detection and diagnosis based on the reconstruction probability of ADNN. Our MO-VAE method performs multi-way data fusion by taking a frontal slice from the training data (as shown in Fig. 3). Each input slice represents all feature signals across all locations at a particular time. Stochastic gradient descent algorithm is used here to learn reconstructions that are close to its original input slice. Once the network get trained, we create a sensor identity matrix $S \in \Re^{s \times m}$ in which each row captures meaningful information for each sensor location for damage localization purposes. The values in this matrix are obtained by calculating the average total reconstruction probability for each set of m output nodes related to one single sensor.

Our method employed the concept of variational auto encoder (VAE) for computing the anomaly score for each new incoming data slice. It aims to calculate the anomaly score for new arrived data based on its reconstruction probability. Practically, VAE generates multiple reconstructions given a single latent space which allows us to perform a statistical reconstructions with a probabilistic approach for detect anomalous data rather than sitting a fixed threshold for anomalous score. This measure provides more principled and objective decision value than reconstruction errors since it considers the variability of the distribution variables, and does not require presetting fixed threshold parameter for identifying damage. Setting a threshold for reconstruction error is problematic especially in the case of multi-way heterogeneous data. Moreover, the normal and anomaly data might share the same mean value. However, anomalous data will not share the same variance to the normal data and it leads to significant lower reconstruction probability, thus classified as damage. Another advantage of using VAE is its robustness against noise. It is inevitable that there will be a base level of noise in any sensor reading which will not be possible for the decoder to reproduce this signal component exactly. However, the level of noise compared to the signal could be encoded into the covariance of the VAE. The following sections discuss the details of the proposed method.

4.1 Multi-way data fusion

As we observed in this study, a large number of sensors are usually used to collect data in SHM applications which often aim to monitor large civil structures such as bridge or a high-rise building. The sensing data being generated from networked sensors mounted structures are considered as three-way data in the form of $(location \times frequency \times time)$ as previously described in Fig. 1. In this setting, two-way matrix analysis is not able to capture the correlation between sensors [32]. At the same time, unfolding the three-way data and concatenating the frequency features from multiple sensors at a certain time to form a single data instance at that time may result in information loss since it breaks the modular structure inherent in three-way data [32]. Accordingly, data fusion plays a critical role in analyzing structure behaviours and assessing the severity of any damage data.

Basically, ADNN is mainly used for the purpose of dimensionality reduction or as anomaly detection models. In fact, ADNN can be also utilized as data fusion structure which can constructs an internal representation for input data collected from multiple sources i.e. sensors. Therefore, our MO-VAE method utilizes the ADNN as a multi-way data fusion model which automatically learns features via its deep-layered structure.

Fig. 3 Autoencoder deep neural network architecture of MO-VAE
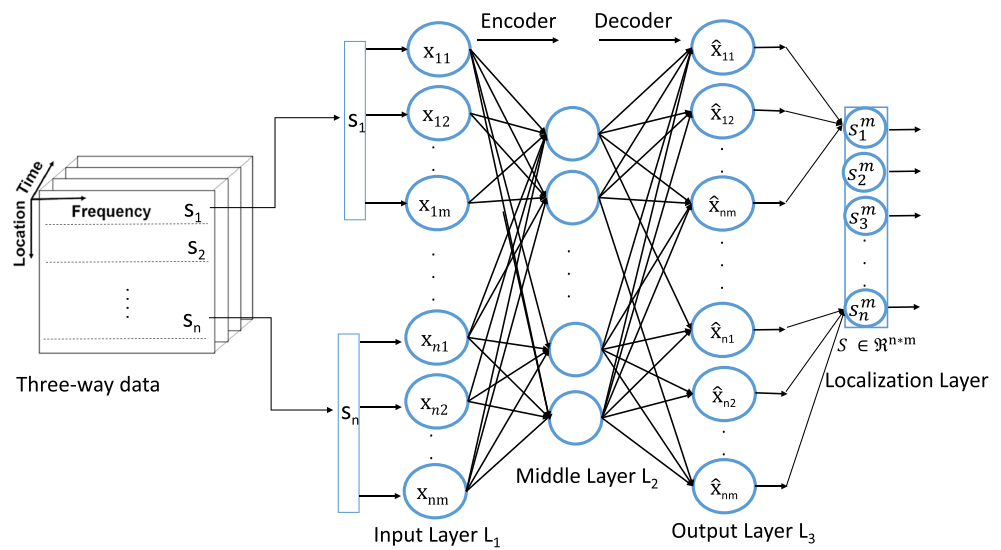
As shown in Fig. 3, ADNN model receives data from multiple sensors at the same time by taking a frontal slice from a training three-way data. Each input slice represents all feature signals across all locations at a particular time. This data from multiple sensors is fed into the input layer to extract damage sensitive features via the encoder layers. The resultant new features in the middle layer (Z) are then used by the decoder layers to determine the damage detection results.

4.2 Probabilistic anomaly detection

The rational idea of anomaly detection in ADNN is to see how well a new data point follows the normal examples. We mentioned before that ADNN aims to learn (encoder) a lower dimensional space Z for input features X, and then try to reconstruct (decode) the original feature space $\hat{X}$. Let's denote the encoder and decoder by $q_\phi(Z \mid X)$ and $p_\theta(X \mid Z)$, respectively. This representation is known as the conditional probability. For example, $p_\theta(X \mid Z)$ is the conditional of X such that Z has happened. Intuitively, the decoder process yields to information loss because the data goes from a low dimensional space Z to a larger dimensional space $\hat{X}$. This loss is known as the reconstruction error which can be measured by calculating the log-likelihood $\log p_\theta(X \mid Z)$ and it will be eventually used as an anomaly score. This measure allows us to see how effectively the decoder has learned to reconstruct an input features X given its latent representation Z.

Our probabilistic anomaly detection method follows the concept of VAE to find a distribution of some latent variable Z which we can sample from $Z \sim q_\phi(Z \mid X)$ to generate new samples $\hat{X}$ from $p_\theta(X \mid Z)$. Each latent variable z_i represents a probability distribution for a given input feature. In the decoding process, we randomly sample from this latent state distribution to generate a vector to be used as an input for the decoder model.

Given X be a set of observed variables and Z is the set of latent variables, the objective function of VAE is considered as an inference problem which aims to compute the conditional distribution of latent variables Z given the observations X i.e. $p(Z \mid X)$. Using Bayesian theorem, we can write it as follows:

$$p_\theta(Z \mid X) = \frac{P_\theta(X \mid Z) \times P(Z)}{P(X)} \tag{8}$$

However, calculating the evidence $p(X)$ is not practical since it requires computing a multidimensional integral in the d unknown variables $z_1, \ldots, z_d$ [33]. Thus, the variational inference (VI) tool is used here to perform approximate Bayesian of the posterior distribution $p_\theta(Z \mid X)$ with a parametric family of distributions $Q_\phi(Z \mid X)$ in a such way that it has tractable solution. The main idea of VI is to pose the inference problem as an optimization problem by modeling $p(Z \mid X)$ using $Q(Z \mid X)$ where $Q(Z \mid X)$ has a simple distribution such as Gaussian.

The $\mathcal{KL}$ divergence method defined in (9) is used here to measure the information loss between the two probability distributions $p(Z \mid X)$ and $Q(Z \mid X)$. In this sense, the optimization problem is to minimize the $\mathcal{KL}$ divergence denoted by $\mathcal{D}_{\mathcal{KL}}$ i.e. ($\min_{\mathcal{KL}} \ p(Z \mid X) \ \| \ Q(Z \mid X))$).

$$
\begin{aligned}
\mathcal{D}_{\mathcal{KL}}(p_\theta(Z \mid X) \ \| \ Q_\phi(Z \mid X)) &= \sum_z Q_\phi(Z \mid X) \log(\frac{Q_\phi(Z \mid X)}{p_\theta(Z \mid X)}) \\
&= E_{Z \sim Q_\phi(Z|X)}\big[\log \frac{Q_\phi(Z \mid X)}{p_\theta(Z \mid X)}\big] \\
&= E_{Z \sim Q_\phi(Z|X)}\big[\log(Q_\phi(Z \mid X)) \\
&\quad - \log(p_\theta(Z \mid X))\big]
\end{aligned} \tag{9}
$$

By substituting (8) in (9), the resultant equation will be as follows:

$$\mathcal{D}_{\mathcal{KL}}(p_\theta(Z \mid X) \;\|\; Q_\phi(Z \mid X)) = E_Z\big[\log(Q_\phi(Z \mid X))$$
$$- \log \frac{P_\theta(X \mid Z) \times P_\theta(Z)}{P_\theta(X)}\big] = E_Z\big[\log(Q_\phi(Z \mid X))$$
$$- \log P_\theta(X \mid Z) - \log P_\theta(Z) + \log P_\theta(X)\big] \quad (10)$$

where $Z = Z \sim Q_\phi(Z \mid X)$. Since the the expectation (E) is based on Z and $P_\theta(X)$ does not involve Z, we can remove $P_\theta(X)$ from (10) and write it as follows:

$$\log P_\theta(X) - \mathcal{D}_{\mathcal{KL}}(p_\theta(Z \mid X)\|Q_\phi(Z \mid X))$$
$$= E_Z\big[\log(p_\theta(X \mid Z))\big] - E_Z\big[\log(Q_\phi(Z \mid X)) - \log P_\theta(Z)\big] \quad (11)$$

The final objective function of variational autoencoder is as follows:

$$\log P_\theta(X) - \mathcal{D}_{\mathcal{KL}}(p_\theta(Z \mid X)\|Q_\phi(Z \mid X))$$
$$= E_Z\big[\log(p_\theta(X \mid Z))\big] - \mathcal{D}_{\mathcal{KL}} Q_\phi(Z \mid X)\|p_\theta(Z) \quad (12)$$

The first term i.e. $\log(p_\theta(X \mid Z))$ represents the reconstruction likelihood and the second term i.e $\mathcal{D}_{\mathcal{KL}}$ is the regularization parameter which forces the posterior distribution $Q_\phi(Z \mid X)$ to be similar to the prior distribution $p_\theta(Z)$. The loss function $J(\theta, \phi)$ of our autoencoder is the negative value of the objective function and its defined as:

$$J(\theta, \phi) = -E_Z\big[\log(p_\theta(X \mid Z))\big] + \mathcal{D}_{\mathcal{KL}}\big[Q_\phi(Z \mid X)\|p_\theta(Z)\big] \quad (13)$$

In variational Bayesian method, this loss function is known as the variational lower bound or evidence lower bound (ELBO). This "lower bound" part comes from the fact that $\mathcal{KL}$ divergence is always non-negative. Thus $J(\theta, \phi)$ is the lower bound of $\log P_\theta(X)$, and it is also known that $\mathcal{D}_{\mathcal{KL}}\big[q_\phi(Z \mid X, \lambda)\|p_\theta(Z \mid X)\big] \geq= 0$. As a result $J(\theta, \phi) \leq \log P_\theta(X)$. Therefore by minimizing the loss, we are maximizing the lower bound of the probability generating real data samples.

Now we need to train the variational autoencoder to learn $Q_\phi(Z \mid X)$ using gradient descent algorithm to optimize the loss with respect to the parameters θ, ϕ. This is where the VAE can relate to the autoencoder where the encoder model learns $Q_\phi(Z \mid X)$ by mapping X to Z and the decoder model learns $p_\theta(Z \mid X)$ by mapping Z back to X. For stochastic gradient descent with step size α, the encoder parameters are updated using (6). Once $Q_\phi(Z \mid X)$ is learned, we sample the latent vector Z from $q_\phi(Z \mid X)$ and then feed it into the decoder network $p_\theta(X \mid Z)$ to generate the new data $\hat{X}$. The training steps of MO-VAE are illustrated in Algorithm 3.

Input: A set of n positive samples $x = \{x^{(i)}\}_{i=1}^n$

- Initialize (ϕ, θ) to a small random value $\mathcal{N}(0, \epsilon^2)$

repeat

- **for** $i = 1 : n$

  - Generate L samples from $\epsilon \sim \mathcal{N}(0, 1)$
  - Perform a feedforward pass to compute all nodes' activations.

- **end for**

- Compute the error using (13)

- Update (ϕ, θ) using gradients of E

until convergence of parameters (ϕ, θ)

Algorithm 3 MO-VAE training algorithm.

To get the reconstruction $\hat{X}$, we generate L random samples from $z \sim N(\mu_{z^{(i)}}, \sigma_{z^{(i)}})$ where $\mu_{z^{(i)}}$ and $\sigma z^{(i)}$ are the mean and standard deviation of the middle layer $z|x^i$ in ADNN, respectively. For each random sample in L, we calculate $\mu_{\hat{x}^{(i)}}$ and $\sigma_{\hat{x}^{(i,l)}}$ for the output layer in ADNN. The final reconstruction probability (RB) can be estimated as follows:

$$RP(x_{new}) = \frac{1}{L} \sum_{l=1}^{L} p_\theta(x \mid z^{(i,l)} | \mu_{\hat{x}^{(i,l)}}, \sigma_{\hat{x}^{(i,l)}}) \quad (14)$$

The damage detection steps of MO-VAE are illustrated in Algorithm 4.

4.3 Damage localization

Once a new data slice is identified as anomaly by ADNN, the values from the output nodes are further propagated into another layer called localization layer as illustrated in Fig. 3. It consists from a set of n nodes each representing one sensor data source. The purpose of this layer is to solve the problem of fault localization. The output values to this layer are obtained by calculating the average of the total reconstruction probability for each m output nodes related to one sensor. The resultant outputs are stored in a matrix $S \in \Re^{n \times m}$ where n is the number of sensors and m is the number of features for each sensor. Using S matrix, it is possible to perform a k-nearest neighbouring algorithm on new output scores S_{new} with each row of matrix S to locate the anomalous rows. The average distance difference between S and S_{new} is used as another anomaly score for damage localization.

Input: A set of new arrived samples $X_{i=1}^{n} \in \Re^{n \times m \times s}$

- $(\phi, \theta) \leftarrow$ Algorithm 3

For $i = 1 : n$

- $\mu_{z^{(i)}}, \sigma_{z^{(i)}} = q_\phi(z|x^{(i)})$
- Generate L samples from $z \sim N(\mu_{z^{(i)}}, \sigma_{z^{(i)}})$
- **For** $l = 1 : L$

    - $\mu_{\hat{x}^{(i,l)}}, \sigma_{\hat{x}^{(i,l)}} = p_\theta(x|z^{(i,l)})$

- **end for**

- Compute $RB^{(i)}$ using (14)
- **if** reconstruction probability(i) $\leq \alpha$ **then**

    - $x^{(i)}$ is healthy

- **else**

    - $x^{(i)}$ is damage

end for

Algorithm 4 MO-VAE damage detection algorithm.

5 Experimental results

5.1 Data collection

We conducted experiments on three case studies representing typical types of civil structures. Two case studies are based on real data collected from an Arch Bridge and Cable-Stayed Bridge in Western Sydney, Australia (Fig. 4). The third one is a laboratory based building structure obtained from Los Alamos National Laboratory (LANL) [34].

5.1.1 The cable-stayed bridge

The bridge was instrumented by 24 uniaxial accelerometers and 28 strain gauges. The locations of these sensors were selected using domain-based knowledge from structural engineers, in order to capture the most relevant response signal from the bridge. In this paper we are using only features based on accelerations data collected from sensors Ai with $i \in [1; 24]$. Figure 5 shows the locations of these 24 sensors on the bridge deck. The acceleration data are collected at 600 Hz, with a range of 2G and a sensitivity of 2 V/G.

For the sake of experiments, we emulated two different kind of damage on this bridge by placing a large static load (vehicle) at different location of a structure. Thus, three scenarios have been considered which include: no vehicle is placed on the bridge (healthy state), a light vehicle with approximate mass of 3 t is placed on the bridge close to location A10 ("Car-Damage") and a bus with approximate mass of 12.5 t is located on the bridge at location A14 ("Bus-Damage"). This emulates slight and severe damage cases which were used in our evaluation Section 5.2.1.

5.1.2 A reinforced concrete jack arch from the Sydney Harbor Bridge

The second case study is a major structural component from the iconic Sydney Harbour Bridge (SHB). There are approximately 800 jack arches distributed over a total distance of 1.2 km in Lane 7, see Fig. 6(a). The jack arches are difficult to access and are inspected typically at two yearly intervals according to standard visual inspection practices.

A concrete cantilever beam with an arch section which has a similar geometry to those on the Sydney Harbour Bridge was manufactured and tested, as shown in Fig. 6(b). We instrumented the specimen with ten accelerometers to measure the vibration response resulting from impact hammer excitation. The structure was excited using an impact hammer with steel tip, which was applied on the top surface of the specimen just above the location of sensor A9, as shown in Fig. 6 (b). The acceleration response of the structure was collected over a time period of 2 seconds at a sampling rate of 8 kHz, resulting in 16000 samples for each event (i.e. a single excitation). A total of 190 impact test responses were collected from the healthy condition.

A crack was then introduced into the specimen in the location marked in Fig. 6(b) using a cutting saw. The crack is located between sensor locations A2 and A3 and it is progressively increasing towards sensor location A9. The length of the cut was increased gradually from 75 mm to 270 mm to introduce four different damage cases as shown in Fig. 7(a-d), and the depth of the cut was fixed to 50 mm; a description is provided in Table 1. This experiment generates a total of 760 impact tests related to four damage cases.

Fig. 4 The cable-stayed bridge from our first case study, Western Sydney, Australia (source: Google Earth)
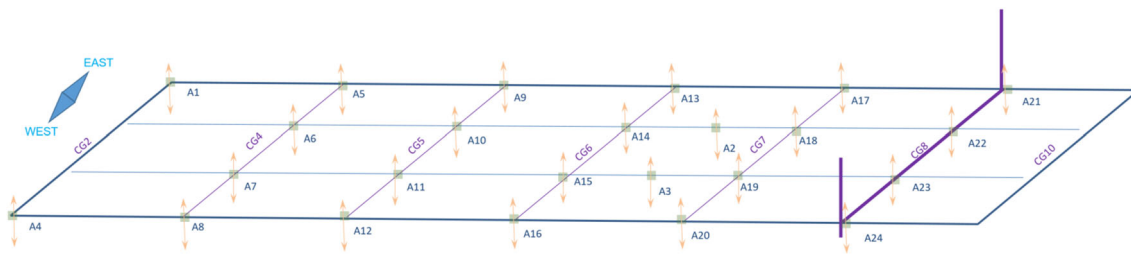
Fig. 5 The locations on the bridge's deck of the 24 *Ai* accelerometers used in this study. The cross girder *j* of the bridge is displayed as *CGj*

5.1.3 Building data

Our second case study was based on the a data collected by [34] from three-story building structure. It is made up of Unistrut columns and aluminum floor plates connected by bolts and brackets as presented in Fig. 8. Eight accelerometers were instrumented on each floor (two on each joint). A shaker was placed at corner D to generate excitation data. It generates 240 samples (a.k.a. events) separated into two main groups, Healthy (150 samples) and Damaged (90 samples). Each event consists of acceleration data for a period of 5.12 seconds sampled at 1600 Hz, resulting in a vector of 8192 frequency values. The Damaged samples were further partitioned into two different damaged cases based on their location: damage in location 3C (60 samples), and the damage in both locations 1A and 3C (30 samples). The damage was introduced by detaching or loosening the bolts at the joints, allowing the aluminum floor plate to move freely relative to the Unistrut column.

Fig. 6 Illustration of the bus lane on the Sydney Harbour Bridge and the manufactured concrete jack arch
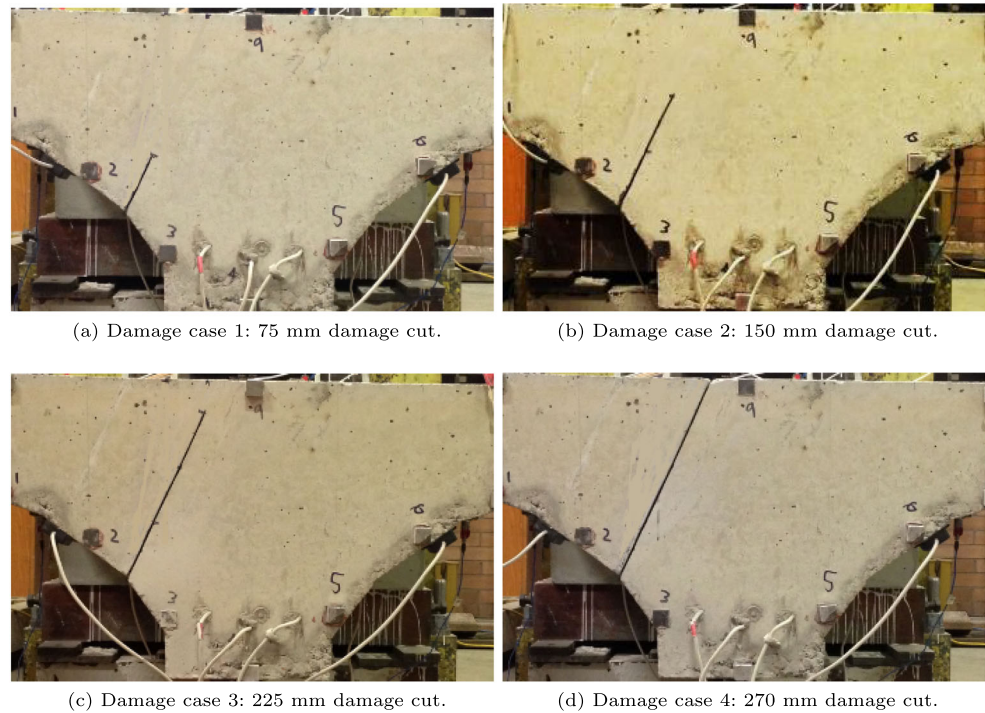

(a) The bus lane on the Sydney Harbour Bridge
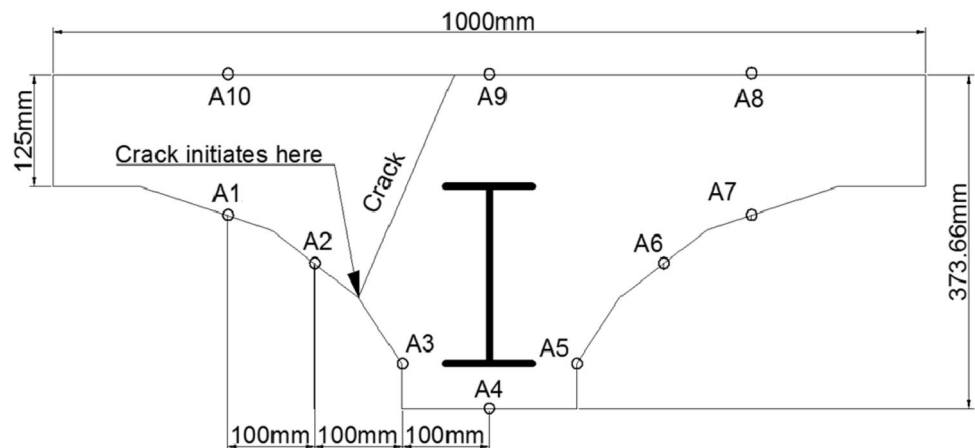
(b) A manufactured concrete jack arches with a similar geometry to those on the Sydney Harbour Bridge and with arrow indicating the cut.

Fig. 7 The crack introduced into the test specimen

(a) Damage case 1: 75 mm damage cut.

(b) Damage case 2: 150 mm damage cut.

(c) Damage case 3: 225 mm damage cut.

(d) Damage case 4: 270 mm damage cut.

5.2 Results and discussions

This section demonstrates how our MO-VAE method can successfully detect and assess the severity of structural damage, and further localize it. It is using the sensor-based data from the above three case studies described in Section 5.1.

For all experiments, six hidden layers were used in MO-VAE and the accuracy values were obtained using the F-Score (FS) measure defined as F-score $= 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$ where Precision $= \frac{\text{TP}}{\text{TP}+\text{FP}}$ and Recall $= \frac{\text{TP}}{\text{TP}+\text{FN}}$ (the number of true positive, false positive and false negative are abbreviated by TP, FP and FN, respectively).

5.2.1 The cable-stayed bridge

Our MO-VAE method was initially validated using vibration data collected from the cable-stayed bridge described in

Table 1 Description of the four damage cases in the test datasets of the reinforced concrete jack arch (specimen)

Case	Damage Description
Damage Case 1	75 mm x 50 mm
Damage Case 2	150 mm x 50 mm
Damage Case 3	225 mm x 50 mm
Damage Case 4	270 mm x 50 mm

Section 5.1.1. We used 24 uni-axial accelerometers to generate 262 samples (a.k.a events) each consists of acceleration data for a period of 2 seconds at a sampling rate of 600 Hz.

For each reading of the uni-axial accelerometer, we normalized its magnitude to have a zero mean and one standard variation. The fast Fourier transform (FFT) is then used to represent the generated data in the frequency domain. Each event now has a feature vector of 600 attributes representing its frequencies. The resultant three-way data has a structure of 24 sensors $\times$ 600 features $\times$ 262 events. We separated the 262 data instances into two groups, 125 samples related to the healthy state and 137 samples for damage state. The 137 damage examples were further divided into two different damaged cases: the "Car-Damage" samples (107) generated when a stationary car was placed on the bridge, and the "Bus-Damage" samples (30) emulated by the stationary bus.

We randomly selected eighty percent of the healthy events (100 samples) from each sensor to form training multi-way of $X \in \mathfrak{R}^{24 \times 600 \times 100}$ (i.e. *training* set). The 137 examples related to the two damage cases were added to the remaining 20% of the healthy data to form a *testing* set, which was later used for the model evaluation. Our probabilistic anomaly detection algorithm was able to successfully detect 98% the healthy and damage events in the *testing* data set, and achieved an F-Score of 0.98±0.01. Moreover, this model was able to assess the progress of damage severity in the structure based on the obtained probability decision values. To illustrate that, we plotted

Fig. 8 Three-story building and floor layout [34]
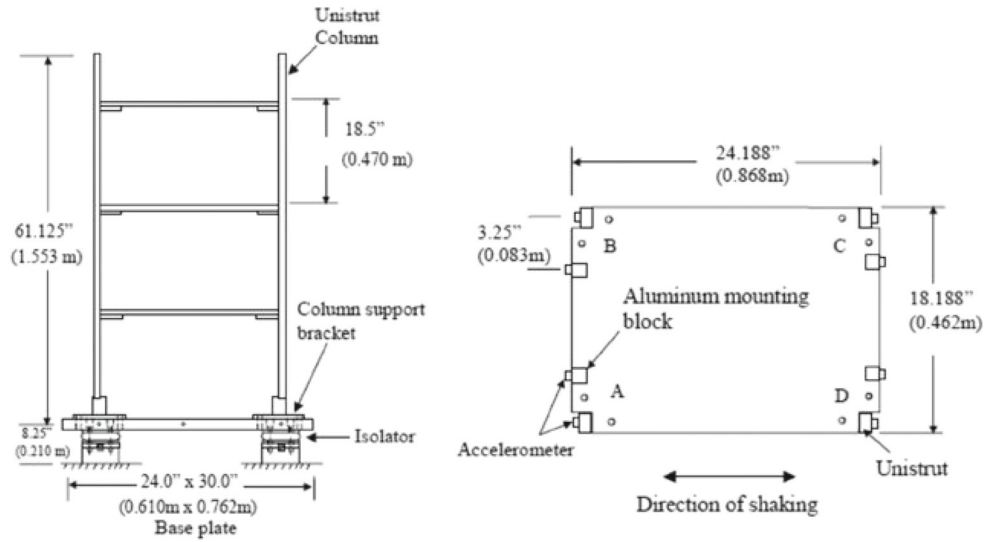
these values for all test samples which are shown in Fig. 9. The horizontal axis indicates the index of the test samples and the vertical axis indicates the magnitude of the probability decision values. A value above the horizontal dashed line indicates a sample classified as healthy, whereas a value below that line indicates an event classified as damage.

As can be seen in Fig. 9(a), the first 25 healthy events denoted by green dot were all correctly classified as healthy samples with a probability decision values above the anomaly threshold value of 3% (97% of confidence interval). 98% the damage samples denoted by yellow and orange dot refer to the "Car-Damage" and "Bus-Damage", respectively, generate high probability decision

Fig. 9 Damage identification results of MO-VAE compared to the state-of-the-art methods applied on the Cable-Stayed Bridge datasets
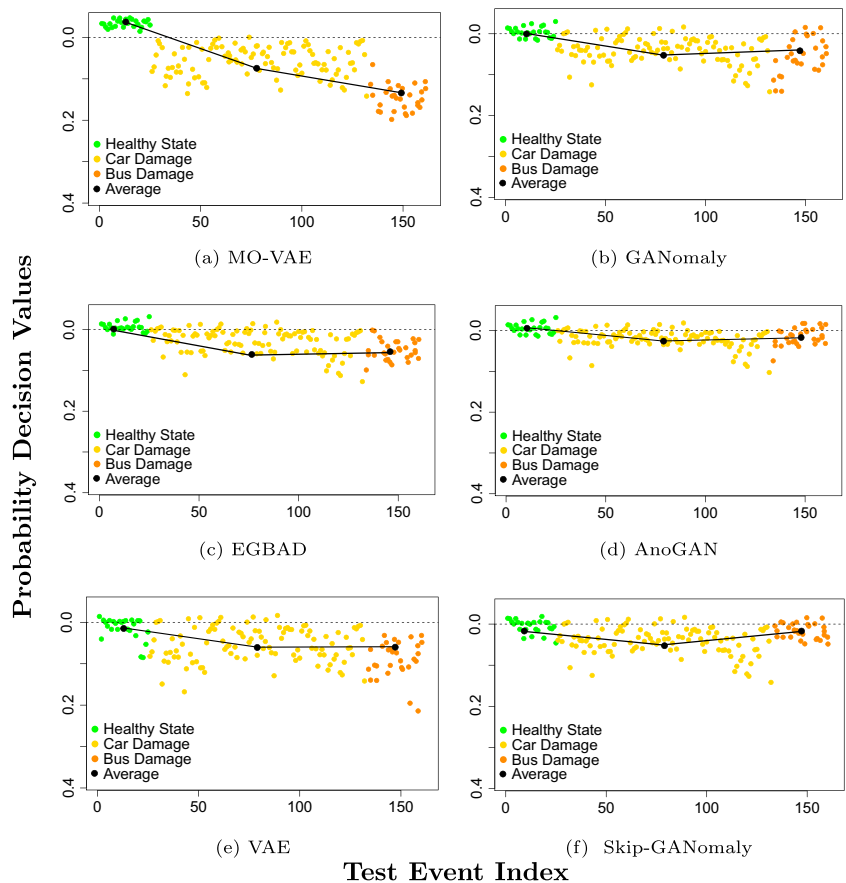
values, thus identified as damage. We further calculated the mean of all the probability decision values for each state to illustrate how the MO-VAE model was also able to asses the severity of the identified damage. Fig. 9(a) shows a solid black line which was drawn to connect the mean values. It can be clearly observed that the MO-VAE model was able to separate the two damage cases ("Car-Damage" and "Bus-Damage") where the probability decision values were further increased for the samples related to the more severe damage cases related to "Bus-Damage". The last step in MO-VAE model was to localize the position of the detected damage by analyzing the identity matrix S_{new} where each row captures meaningful information for each sensor location. We calculated the average distance from each row in matrix S to k-nearest neighbouring to S_{new}. The resultant k-nn score for each sensor is presented in Fig. 10 which clearly shows the capability of MO-VAE for damage localization. As expected, sensors A10 and A14 related to the "Car-Damage" and "Bus-Damage", respectively, behaved significantly different from all the other sensors apart from the position of the emulated damage.

The next experiment was to compare our obtained results with the state-of-the-art methods described in Section 2 i.e GANomaly EGBAD, AnoGAN, Skip-GANomaly and VAE proposed in [24, 27, 28, 30, 31], respectively. The same *training* data set as above was used to construct these models, and the same *testing* data set was used to evaluate their classification performance. The resulted accuracies are shown in Table 2 which demonstrates that our MO-VAE consistently outperforms the other approaches. Moreover, the probability decision values as shown in Fig. 9(b-f) of these state-of-the-art methods are not able to clearly assess the progress of the damage severity in the structure since only one single anomaly score for each event is generated
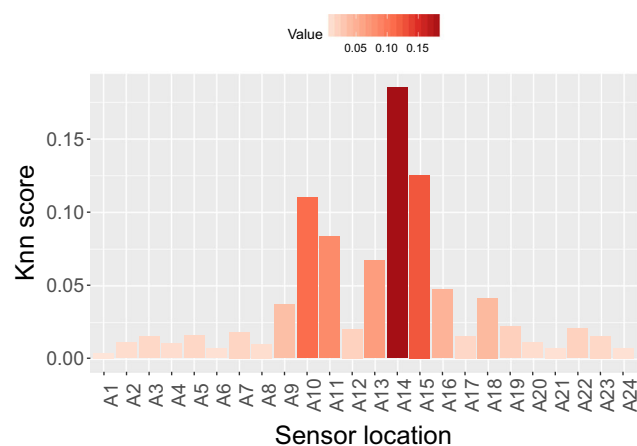

Fig. 10 Location anomaly score in the localization layer applied on the cable-stayed bridge dataset using MO-VAE

by the model using the inputs from sensors $\{A_i\}_{i=1}^{24}$. Consequently, these models are lacking the capability to implement a method for damage localization

5.2.2 A reinforced concrete jack arch from the Sydney Harbor Bridge

Damage identification process was carried out in the same way that was performed in the previous case study. This dataset consists of 950 samples (events) separated into two main groups, healthy state (190 samples) and damaged states (760 samples). Each sample is the measured vibration response of the structure with eight thousand attributes in the frequency domain (8 kHz × 2 sec × 0.5 (considering Nyquist frequency)).

The measured acceleration responses collected from the 10 sensors were utilized to construct the damage sensitive features. Eighty percent of the healthy data were randomly selected for the training stage, while the remaining 20% of the healthy samples and all the damage cases were used for testing. The dimension of the data was reduced into 80 using random projection method. The resultant three-way data has a structure of 10 locations × 80 features × 950 events.

As shown in Table 2, the MO-VAE model significantly outperformed the other state-of-the-art methods. The average F-score value of MO-VAE was equal to 0.92±0.03. A small number of events (8 events) in Damage Case 1 were miss-classified as healthy. This illustrates that our MO-VAE has capability to identify small defects as well as the progression of the damage as shown in Fig. 11(a). The GANomaly, AnoGAN and Skip-GANomaly methods performed badly on the 10 sensor datasets as shown in Table 2. This is what we anticipated dealing with individual sensors for building GAN models which may lack of capability for capturing the underlying structure of the sensing data. With respect to the VAE method, as we expected, it generates comparable results to our MO-VAE model with an average F-score equal to 0.89±0.02 since it encodes the distribution and regularize it during the training to capture the latent space. The damage progression results using the state-of-the-art methods are presented in Fig. 11. It can be readily realized that by using this approach the performance of the method to monitor the progress of damage is not consistent where the decision values are increasing with the development of the damage as shown in Fig. 11(b) and (c). Based on this, it can be concluded that compared to the MO-VAE, these state-of-the-art methods lack the ability to provide reliable information about the severity of damage in the structure. Damage localization was not carried out in this experiment due to the small size of the specimen.

Table 2 $F - score$ of various methods applied on the three case study datasets

	MO-VAE	GANomaly	EGBAD	AnoGAN	Skip-GANomaly	VAE
Cable	0.98±0.01	0.96±0.02	0.93±0.03	0.94±0.04	0.95±0.04	0.95±0.02
Specimen	0.92±0.02	0.84±0.02	0.88±0.05	0.84±0.06	0.84±0.04	0.89±0.02
Building	0.96±0.02	0.88±0.01	0.91±0.03	0.90±0.04	0.93±0.02	0.92±0.04

5.2.3 Building data

Our last experiments were conducted using the acceleration data acquired from 24 sensors instrumented on the three-story building as described in Section 5.1.3. Similar to the previous experiments, we normalized the accelerometer data to have zero mean and unity variance. Then we applied FFT method to represent the data in frequency domain. For each two adjacent accelerometers at a location, we used the difference between their signals as variables and only the top 150Hz were selected as input features to our MO-VAE model. The resultant three-way data has a structure of 12 locations × 768 features × 240 events. We randomly selected 80% of the healthy events (120 samples) from the 12 locations as a training multi-way data $X \in \Re^{12 \times 768 \times 120}$ (i.e. *training* set). The remaining 20% of the healthy data and the data obtained from the two damage cases were used for testing (i.e. *testing* set).

Our constructed MO-VAE model achieved an F-score of 96%. The false alarm rate was equal to zero where all the healthy samples are correctly detected in the *testing*

data set. Figure 12(a) shows the plot of the probability decision values generated by our MO-VAE. It can be clearly observed from Fig. 12(a) that the more severe damage test data related to locations 1A and 3C were more deviated from the training data with lower probability decision values. Similar to the last case study, we further propagated the probability decision values obtained by the output layer into the localization layer to construct S_{new} matrix. Then we computed the k-nn score for each sensor based on the average distance between each row of matrix S to S_{new}. Figure 13 shows the resultant k-nn score for each sensor. It clearly shows that MO-VAE method correctly captures damage locations. As expected, sensors 1A and 3C produced very high k-nn score due the introduced damage at these two locations. The k-nn score of 3C was higher than 1A because that damage was introduced in both locations 1A and 3C at the same time.

The last experiment was to compare our obtained results with the other state-of-the-art methods . The F-score accuracy of Skip-GANomaly was recorded at 93% with no clear separation between the different levels of

Fig. 11 Damage identification results of MO-VAE compared to the state-of-the-art methods applied on the Reinforced Concrete Jack Arch datasets
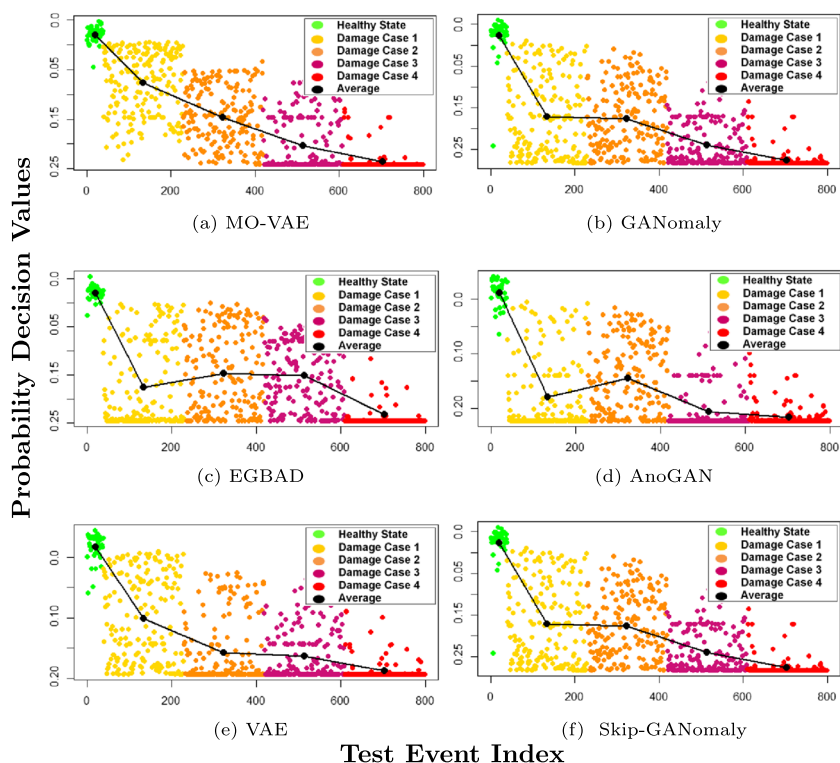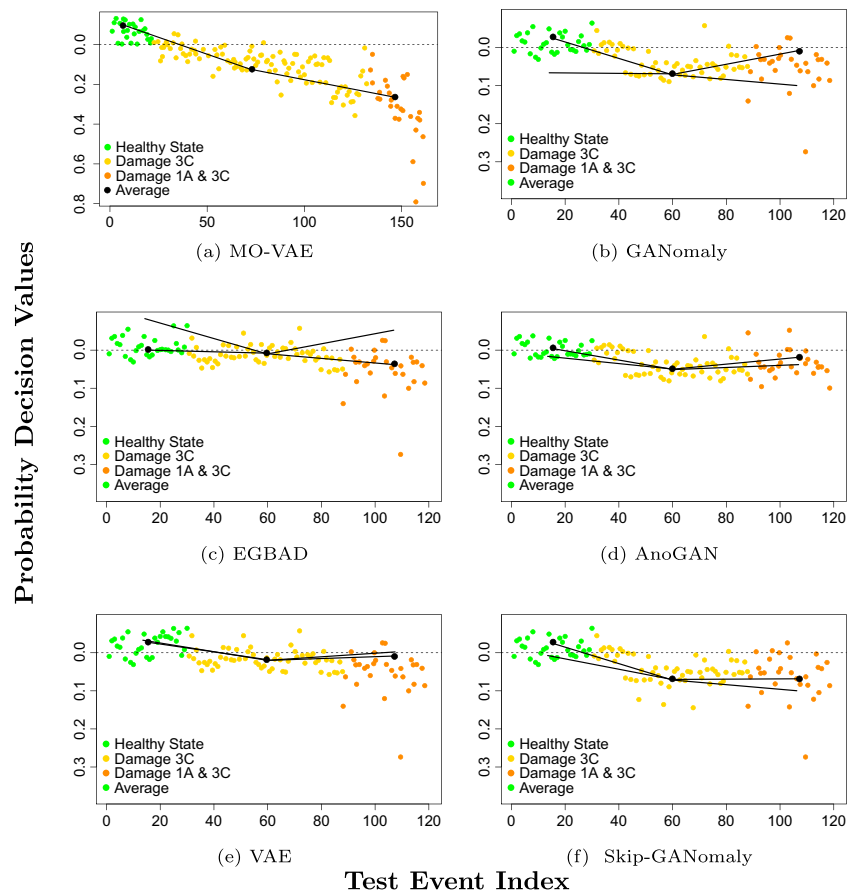
Fig. 12 Damage identification
results of MO-VAE compared to
the state-of-the-art methods
applied on the Building dataset

(a) MO-VAE

(b) GANomaly

(c) EGBAD

(d) AnoGAN

(e) VAE

(f) Skip-GANomaly

Probability Decision Values

Test Event Index

damage as illustrated in Fig. 12(f). GANomaly, on the other hand, generates high false alarm rates with several healthy samples predicted as damage. Moreover, these methods don't have the capability to implement a method for damage

localization since only one single anomaly score for each event is generated by these models using input data from sensors $\{A_i\}_{i=1}^{12}$.

6 Conclusion

Multiway data analysis has gained a lot of interest in many fields where standard two way analysis don't have the capabilities to learn underlying structure of the multi-way data. We proposed a multi-objective variational autoencoder method for damage detection, localization and severity assessment in multi-way structural data based on the reconstruction probability of the autoencoder deep neural network. The proposed method performs data fusion by taking input features from a networked sensors attached to a structure. Stochastic gradient descent algorithm is then used to learn reconstructions that are close to its original input slice followed by constructing a sensor identity matrix which used for damage localization. For each new incoming data slice we calculate its anomaly score based on reconstruction probability and we use the obtained
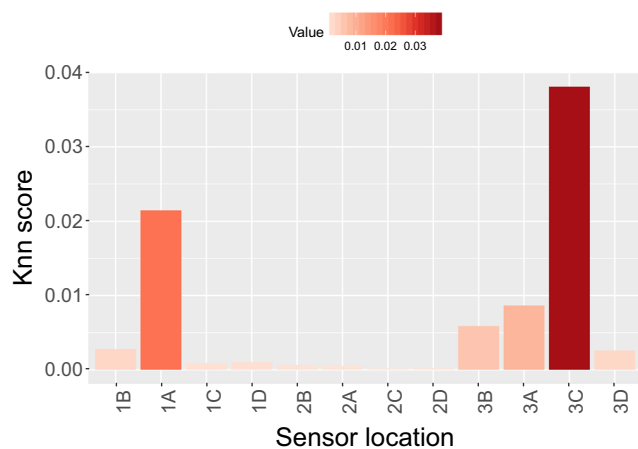

Fig. 13 Location anomaly score in the localization layer on the Building dataset using MO-VAE

reconstruction probability values for damage assessment. The sensor identity matrix is finally utilized to locate the identified damage.

We evaluated our method on multi-way datasets in the area of structural health monitoring for damage detection purposes. Experimental results showed that our approach succeeded at detecting the damage events with an average F-score of 0.95% and higher for all datasets. Moreover, Our model demonstrated the capability to work very well in localizing damage and estimating different levels of damage severity in an unsupervised aspect. Compared to the state-of-the-art approaches, our proposed method shows better performance in terms of damage detection and localization.

References

1. Rytter A (1993) Vibrational based inspection of civil engineering structures. Dept. of building technology and structural engineering Aalborg University
2. Krizhevsky A, Hinton GE (2011) Using very deep autoencoders for content-based image retrieval. In: ESANN
3. Hinton G, Deng L, Yu D, Dahl GE, Mohamed AR, Jaitly N et al (2012) Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Proc Mag 29(6):82–97
4. Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. In: Advances in neural information processing systems; pp 3104–3112
5. Farrar CR, Worden K (2006) An introduction to structural health monitoring. Philos Trans R Soc A Math Phys Eng Sci 365(1851):303–315
6. Anaissi A, Khoa NLD, Rakotoarivelo T, Alamdari MM, Wang Y (2017) Self-advised incremental one-class support vector machines: an application in structural health monitoring. In: International conference on neural information processing. Springer, pp 484–496
7. Anaissi A, Zandavi SM (2019) Multi-objective autoencoder for fault detection and diagnosis in higher-order data. In: 2019 international joint conference on neural networks (IJCNN). IEEE, pp 1–8
8. Anaissi A, Khoa NLD, Rakotoarivelo T, Alamdari MM, Wang Y (2018) Adaptive online One-Class support vector machines with

9. applications in structural health monitoring. ACM Trans Intell Syst Technol (TIST) 9(6):64
9. Anaissi A, Khoa NLD, Alamdari MM, Braytee A, Wang Y, Mustapha S et al (2017) Adaptive one-class support vector machine for damage detection in structural health monitoring. In: Pacific-asia conference on knowledge discovery and data mining. Springer, pp 459–471
10. Anaissi A, Khoa NLD, Wang Y (2018) Automated parameter tuning in one-class support vector machine: an application for damage detection. Int J Data Sci Analytics 6:311–325
11. Khoa NLD, Alamdari MM, Rakotoarivelo T, Anaissi A, Wang Y (2018) Structural health monitoring using machine learning techniques and domain knowledge based features. In: Human and machine learning. Springer, pp 409–435
12. Leung K, Leckie C (2005) Unsupervised anomaly detection in network intrusion detection using clusters. In: Proceedings of the twenty-eighth Australasian conference on computer science, vol 38. Australian Computer Society, Inc., Australia, pp 333–342
13. Mukkamala S, Janoski G, Sung A (2002) Intrusion detection using neural networks and support vector machines. In: Neural networks, 2002. IJCNN'02. proceedings of the 2002 international joint conference on. vol 2. IEEE, pp 1702–1707
14. Yin S, Zhu X, Jing C (2014) Fault detection based on a robust one class support vector machine. Neurocomputing 145:263–268
15. Mahadevan S, Shah SL (2009) Fault detection and diagnosis in process data using one-class support vector machines. J Process Control 19(10):1627–1639
16. Sophian A, Tian GY, Taylor D, Rudlin J (2003) A feature extraction technique based on principal component analysis for pulsed eddy current NDT. NDT E Int 36(1):37–41
17. Lu Y, Michaels JE (2009) Feature extraction and sensor fusion for ultrasonic structural health monitoring under changing environmental conditions. IEEE Sensors J 9(11):1462–1471
18. Anaissi A, Makki Alamdari M, Rakotoarivelo T, Khoa NLD (2018) A Tensor-Based structural damage identification and severity assessment. Sensors 18(1):111
19. Jiang SF, Zhang CM, Koh C (2006) Structural damage detection by integrating data fusion and probabilistic neural network. Adv Struct Eng 9(4):445–458
20. Khoa NLD, Anaissi A, Wang Y (2017) Smart infrastructure maintenance using incremental tensor analysis. In: Proceedings of the 2017 ACM on conference on information and knowledge management. ACM, pp 959–967
21. Hong C, Yu J, Wan J, Tao D, Wang M (2015) Multimodal deep autoencoder for human pose recovery. IEEE Trans Image Process 24(12):5659–5670
22. AP SC, Lauly S, Larochelle H, Khapra M, Ravindran B, Raykar VC et al (2014) An autoencoder approach to learning bilingual word representations. In: Advances in neural information processing systems, pp 1853–1861
23. Germain M, Gregor K, Murray I, Larochelle H (2015) Made: masked autoencoder for distribution estimation. In: International conference on machine learning, pp 881–889
24. An J, Cho S (2015) Variational autoencoder based anomaly detection using reconstruction probability. Spec Lect IE 2:1–18
25. Chong YS, Tay YH (2017) Abnormal event detection in videos using spatiotemporal autoencoder. In: International symposium on neural networks. Springer, pp 189–196
26. Yan W, Yu L (2015) On accurate and reliable anomaly detection for gas turbine combustors: a deep learning approach. In: Proceedings of the annual conference of the prognostics and health management society
27. Akcay S, Atapour-Abarghouei A, Breckon TP (2018) Ganomaly: semi-supervised anomaly detection via adversarial training. In: Asian conference on computer vision. Springer, pp 622–637

28. Akçay S, Atapour-Abarghouei A, Breckon TP (2019) Skip-ganomaly: skip connected and adversarially trained encoder-decoder anomaly detection. In: 2019 international joint conference on neural networks (IJCNN). IEEE, pp 1–8

29. Andrews JT, Jaccard N, Rogers TW, Griffin LD (2017) Representation-learning for anomaly detection in complex x-ray cargo imagery. In: Anomaly detection and imaging with x-rays (ADIX) II. vol 10187. International Society for Optics and Photonics, pp 101870E

30. Medel JR, Savakis A (2016) Anomaly detection in video using predictive convolutional long short-term memory networks. arXiv:161200390

31. Zenati H, Foo CS, Lecouat B, Manek G, Chandrasekhar VR (2018) Efficient gan-based anomaly detection. arXiv:180206222

32. Acar E, Yener B (2009) Unsupervised multiway data analysis: a literature survey. IEEE Trans Knowl Data Eng 21(1):6–20

33. Kingma DP, Welling M (2013) Auto-encoding variational bayes. arXiv:13126114

34. Larson AC, Von Dreele R (1987) Los alamos national laboratory report no LA-UR-86-748

Affiliations

Ali Anaissi[1] · **Seid Miad Zandavi[1]** · **Basem Suleiman[1]** · **Mohamad Naji[2]** · **Ali Braytee[2]**

Seid Miad Zandavi
miad.zandavi@sydney.edu.au

Basem Suleiman
basem.suleiman@sydney.edu.au

Mohamad Naji
Mohamad.Naji@student.uts.edu.au

Ali Braytee
Ali.Braytee@uts.edu.au

[1] School of Computer Science, The University of Sydney, Street, Camperdown, 2006, NSW, Australia

[2] School of Computer Science, University of Technology Sydney, Street, Ultimo, 2007, NSW, Australia