

Application of deep autoencoder model for structural condition monitoring

PATHIRAGE Chathurdara Sri Nadith¹, LI Jun², LI Ling^{1,*}, HAO Hong², and LIU Wanquan¹

1. Department of Computing, Curtin University, Perth WA 6102, Australia;

2. Centre for Infrastructural Monitoring and Protection, Curtin University, Perth WA 6102, Australia

Abstract: Damage detection in structures is performed via vibration based structural identification. Modal information, such as frequencies and mode shapes, are widely used for structural damage detection to indicate the health conditions of civil structures. The deep learning algorithm that works on a multiple layer neural network model termed as deep autoencoder is proposed to learn the relationship between the modal information and structural stiffness parameters. This is achieved via dimension reduction of the modal information feature and a non-linear regression against the structural stiffness parameters. Numerical tests on a symmetrical steel frame model are conducted to generate the data for the training and validation, and to demonstrate the efficiency of the proposed approach for vibration based structural damage detection.

Keywords: auto encoder, non-linear regression, deep auto encoder model, damage identification, vibration, structural health monitoring.

DOI: 10.21629/JSEE.2018.04.22

1. Introduction

Structural health monitoring (SHM) aims to assess the structural performance and evaluate the safety conditions of civil infrastructure under operational conditions. Civil structures continuously accumulate damage during their service life due to material deterioration, cyclic loading, and environmental conditions. By analyzing the measurements from various sensors installed on structures, SHM techniques detect and track the possible anomalies that could potentially produce more damage and finally lead to catastrophic structural failures with a huge loss. Measured data from the structures are widely used to detect not only the existence and location of possible damage, but also the severity of the damage.

Vibration measurements are usually used for structural condition monitoring. It is based on the fact that when

structures have damages or condition changes, the structural physical properties (i.e., stiffness, mass and damping matrices) will change and therefore vibration properties (i.e., modal parameters) will also change. The changes in the structural vibration characteristics, such as natural frequencies, mode shapes and other modal information, can be used to identify the location and severity of damages.

Neural network techniques have been applied for vibration based damage identification in civil engineering. Yun et al. [1] presented using the neural networks technique to estimate structural joint damage from modal data. It was found that the noise injection learning with a realistic noise level for each input component is effective to better understand the noise effect. Bakhary et al. [2] proposed a statistical approach to take into account the effect of uncertainties in developing an artificial neural network model for structural damage detection. Artificial neural networks are generally applicable to problems where a significant database of labelled information is available, but difficult to specify an explicit algorithm.

Back propagation (BP) neural network is one of the most popular shallow learning algorithms for training the neural network based on an optimization method such as the gradient descent method. BP based training has been found to be effective, providing: i) initial weights are close enough to a good solution; ii) computers are fast enough; iii) data sets are big enough. However, it is difficult to optimize the weights in the networks with multiple hidden layers. Theoretical results such as those shown in [3–5], and reviewed and discussed by [6] demonstrate that deep learning (DL) architectures are favourable in learning this kind of complicated functions that can represent high level abstractions in computer vision [7–10], natural language processing [11], information retrieval [12] and robotics [13]. DL based neural networks significantly outperform shallow learning based counterparts in many applications, and often match or overwhelm the state-of-the-art ap-

Manuscript received January 03, 2017.

*Corresponding author.

This work was supported by the Australian Research Council.

proaches.

Autoencoders [14] are simple learning models which transform inputs into outputs with the least possible amount of distortion. They have been used in the “deep architecture” approaches [6,14,15], where autoencoders are stacked and trained from bottom up in an unsupervised manner, followed by a supervised learning phase to train the top layer and fine-tune the entire architecture. The unsupervised learning algorithm is used to pre-train each layer, to capture the main variations in its input. This unsupervised pre-training phase sets the stage for a final training phase where the deep architecture is fine-tuned with respect to a supervised training criterion with gradient-based optimization. These deep architectures have been shown to lead to state-of-the-art results in a number of challenging classification tasks [7,8,16] and regression problems [9,10].

Deep autoencoder (DAE) models [15,17] that fall in this category allow effective feature learning through hierarchical non-linear mappings via the multiple hidden layers of the model. Inspired by the work of [9,10] where the image processing has been performed with DAE, we propose a DAE model for vibration based structural condition monitoring and identification.

2. DAE model for structural health monitoring

This section discusses a DAE model that is proposed to perform the dimensionality reduction and the relationship learning. In the proposed approach, the fact that the natural frequencies and their corresponding mode shapes of a structure relate with the physical properties, such as stiffness, is explored. The modal information, such as frequencies and mode shapes, are used as the input to the proposed DAE model and the output will be the structural elemental stiffness parameters. A new feature is formed with all the frequencies along with their mode shapes to be fed into the DAE. Since a high dimensional feature may contain unnecessary information due to the redundant data and measurement noise effect, learning a relationship from such a high dimensional feature (the measured frequencies and their associated mode shapes) to elementary stiffness parameters will be likely to be less accurate than that of using reduced dimensional features. It is therefore understandable that the problem can be better addressed in two steps. The first attempts to reduce the dimensionality of the feature preserving the required information, while the second is to learn the relationship between the feature with the reduced dimension and the stiffness parameters of the structure as shown in Fig. 1. A DAE model that consists of two main components connected sequentially where each

component is optimized on a specific objective with relevance to the final goal is proposed in this paper. Details of the two DAE components will be presented in the next section.

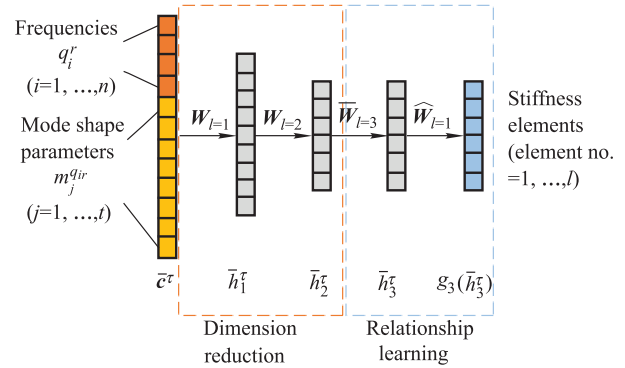


Fig. 1 The proposed DAE model

3. DAE model for pattern recognition

3.1 Dimension reduction

Autoencoders are stacks that learn an under complete representation to form deep architecture. In the proposed autoencoders, the 1st hidden layer performs the feature fusion (both the frequencies and mode shapes from the structure in this study) based on non-linear dimensionality reduction, the 2nd hidden layer performs further compression on the learnt low dimensional feature observed in the 1st hidden layer; and the 3rd and 4th layer perform the corresponding decoding of the low dimensional feature observed in the 2nd hidden layer. Hence the low dimensional feature space observed from the 2nd hidden layer will preserve the information to perform the reconstruction of the original feature since the used error criteria are the mean squares error for the reconstruction of the original feature. Structural natural frequencies and their corresponding mode shapes are fed into the autoencoder as follows:

$$\bar{c}^\tau = [q_1^\tau, \dots, q_n^\tau, m_1^{q_1^\tau}, \dots, m_t^{q_n^\tau}]^T \quad (1)$$

where q_i^τ is the i th ($i = 1, \dots, n$) structural natural frequency included in the τ th sample and $m_j^{q_i^\tau}$ is the j th ($j = 1, \dots, t$) mode shape parameter corresponding to the i th frequency. \bar{c}^τ is the concatenated high dimensional feature that combines n frequencies and nt mode shapes with each frequency having t corresponding mode shape values. This feature \bar{c}^τ is used in the proposed DAE model to perform the pre-training and training. The first layer of the DAE model is trained to optimize the following cost function:

$$[\mathbf{W}_{l=1}^*, \bar{\mathbf{b}}_{l=1}^*, \widehat{\mathbf{W}}_{l=1}^*, \widehat{\mathbf{b}}_{l=1}^*]_j =$$

$$\arg \min_{\mathbf{W}, \bar{\mathbf{b}}, \widehat{\mathbf{W}}, \widehat{\mathbf{b}}} \sum_{\tau=1}^N \|\bar{\mathbf{c}}^\tau - g_1(f_1(\bar{\mathbf{c}}^\tau))\|_2^2 \quad (2)$$

where N is the number of samples involved in the training, $\bar{\mathbf{c}}^\tau$ is the concatenated feature that contains frequencies and mode shape values, $g_1(\cdot)$ and $f_1(\cdot)$ are the decoder and encoder [17] functions respectively.

The 2nd hidden layer further reduces the dimension of the low dimensional features obtained from the 1st hidden layer via the following cost function:

$$\begin{aligned} & [\mathbf{W}_{l=2}^*, \bar{\mathbf{b}}_{l=2}^*, \widehat{\mathbf{W}}_{l=2}^*, \widehat{\mathbf{b}}_{l=2}^*]_j = \\ & \arg \min_{\mathbf{W}, \bar{\mathbf{b}}, \widehat{\mathbf{W}}, \widehat{\mathbf{b}}} \sum_{\tau=1}^N \|\bar{\mathbf{h}}_1^\tau - g_2(f_2(\bar{\mathbf{h}}_1^\tau))\|_2^2 \quad (3) \end{aligned}$$

where N is the number of samples, $\bar{\mathbf{h}}_1^\tau$ is the low dimensional representation that is established in the layer 1 for the τ th sample, $g_2(\cdot)$ and $f_2(\cdot)$ are the decoder and encoder functions that are utilized in the 2nd layer respectively. Pre-training is carried out to obtain the parameters using the full batch gradient descent algorithm.

3.2 Relationship learning

This model has three layers, out of which the first two layers are pre-trained as described in the above section. A relationship learning layer after the 2nd hidden layer is introduced. This will facilitate the relationship learning process by being able to utilize the low dimensional feature learnt at the 2nd hidden layer which is indeed a better feature representation than the input itself. Furthermore, the dimensionality reduction process will be supervised with the stiffness information, thus keeping the necessary information on the concatenated feature to predict the elemental stiffness parameters in the output. The 3rd hidden layer (the relationship learning layer) utilizes the output from the 2nd hidden layer to learn the mapping towards the stiffness elements that are presented at the output layer. Hence the observed latent representation in the 2nd hidden layer is utilized to perform a non-linear regression on stiffness elements via the third layer. The following cost function is optimized during this process:

$$\begin{aligned} & [\mathbf{W}_{l=2}^*, \bar{\mathbf{b}}_{l=2}^*, \widehat{\mathbf{W}}_{l=2}^*, \widehat{\mathbf{b}}_{l=2}^*]_j = \\ & \arg \min_{\mathbf{W}, \bar{\mathbf{b}}, \widehat{\mathbf{W}}, \widehat{\mathbf{b}}} \sum_{\tau=1}^N \|\bar{\boldsymbol{\sigma}}^\tau - g_3(f_3(\bar{\mathbf{h}}_2^\tau))\|_2^2 \quad (4) \end{aligned}$$

where N is the number of samples, $\bar{\mathbf{h}}_2^\tau$ is the low dimensional representation that is produced at the 2nd layer for the τ th sample, $\bar{\boldsymbol{\sigma}}^\tau$ is the corresponding expected vector of structural stiffness parameters for the τ th sample. $g_3(\cdot)$ and $f_3(\cdot)$ are the decoder and the encoder functions that are

utilized in the 3rd layer respectively. Pre-training is also carried out with the full batch gradient descent algorithm as mentioned above. Once the optimal parameters are observed, the whole network is fine-tuned again to optimize all layers (L) jointly as below:

$$\begin{aligned} & [\mathbf{W}_{l=1}^*|_{l=1}^L, \bar{\mathbf{b}}_{l=1}^*|_{l=1}^L, \widehat{\mathbf{W}}_L^*, \widehat{\mathbf{b}}_L^*] = \\ & \arg \min_{\mathbf{W}_{l=1}^*|_{l=1}^L, \bar{\mathbf{b}}_{l=1}^*|_{l=1}^L, \widehat{\mathbf{W}}_L, \widehat{\mathbf{b}}_L} \sum_{\tau=1}^N \|\bar{\boldsymbol{\sigma}}^\tau - p(\bar{\mathbf{c}}^\tau)\|_2^2 \quad (5) \end{aligned}$$

where $p(\bar{\mathbf{x}}_i) = g_L(f_L(f_{L-1}(f_{L-2}(\bar{\mathbf{x}}_i))))$ with $L = 3$ and $\mathbf{W}_{l=1}^*|_{l=1}^L$ denotes the encoder's weights while $\widehat{\mathbf{W}}_L$ denotes the only decoder weight as shown in Fig. 1. Joint optimization of the above objective function in (5) towards learning the relationship between the feature $\bar{\mathbf{c}}^\tau$ and the stiffness parameters, the proposed DAE model is expected to perform non-linear dimension reduction preserving the necessary information to facilitate the relationship learning task from input to output.

4. Verification

Numerical verifications on an experimental testing model are conducted to demonstrate the accuracy and efficiency of the proposed approach. A 7-storey steel frame structure is taken as an example. Seventy elements are defined in the structure, hence 70 elemental stiffness parameters are used to describe the health conditions of the structure. The stiffness parameters 0 and 1 denote the complete failure and intact states of the element respectively. A non-zero value would indicate that there is a stiffness reduction in the associated structural element. The data generation is performed assuming that stiffness reduction can occur in each of the 70 elements individually. The experimental model, data generation, pre-processing and the performance evaluation of the proposed DAE model are presented in the following sections.

4.1 Experimental model

A seven-storey steel plane frame is fabricated in the laboratory and the dimensions of the frame are described in Table 1.

Table 1 Dimensions of the experimental model

Element	Dimension
Height of each storey/m	0.3
Length of the beam/m	0.5
Cross-section of the column/mm ²	49.98 × 4.85
Cross-section of the beam/mm ²	49.89 × 8.92
Mass density of the column/(kg/m ³)	7 850
Mass density of the beam/(kg/m ³)	7 734.2

The initial Young's modulus is taken as 210 GPa for all members. The connections between column and beam elements are continuously welded at the top and bottom of

the beam section. Two pairs of mass blocks with approximately 4 kg weight each, are fixed at the quarter and three-quarter length of the beam in each storey to simulate the mass from the floor of a building structure. The bottoms of the two columns of the frame are welded onto a thick and solid steel plate fixed to the ground.

Fig. 2 shows the finite element model of the steel frame structure. It consists of 65 nodes and 70 planar frame elements. The weights of steel blocks are added at the corresponding nodes of the finite element model as concentrated masses. Each node has three degrees of freedom (DOFs) (two translational displacements x, y and a rotational displacement θ), and the system has 195 DOFs in total. The translational and rotational restraints at the supports, which are Nodes 1 and 65, are represented initially by a large stiffness of 3×10^9 N/m and 3×10^9 N·m/rad, respectively.

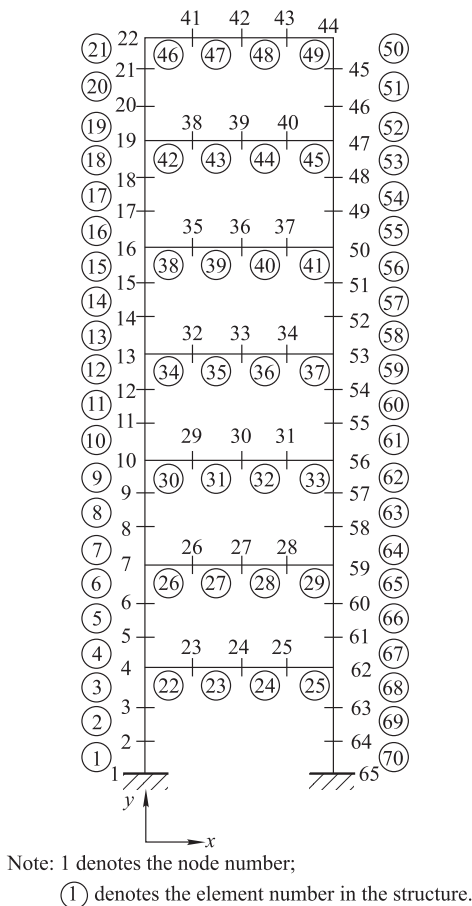


Fig. 2 Finite element model of the frame structure

Finite element model updating of the initial finite element model is conducted to minimize the discrepancies between the analytical finite element model and the experimental model in the laboratory. The difference between the frequencies and mode shapes obtained from the analytical

finite element model and the experimental measurements is minimized.

Measured natural frequencies and mode shapes of the first seven modes are shown in Fig. 3. Only the mode shape values at the 14 beam-column joints are shown. The first-order modal sensitivity-based updating method [18] is used. It should be noted that the first seven measured frequencies and their associated 7×14 mode shape values on the beam-column joints are used in the updating procedure and the 70 elastic modulus values and six support stiffness values are required to be updated. The first seven measured and analytical frequencies, and the modal assurance values (MAC) associated with these seven modes are included in the objective function for model updating. The detailed model updating process can be found in [19,20]. A very good updated model has been achieved to match the analytical and measured modal information. This updated finite element model is taken as the reference model for generating the training and validation data.

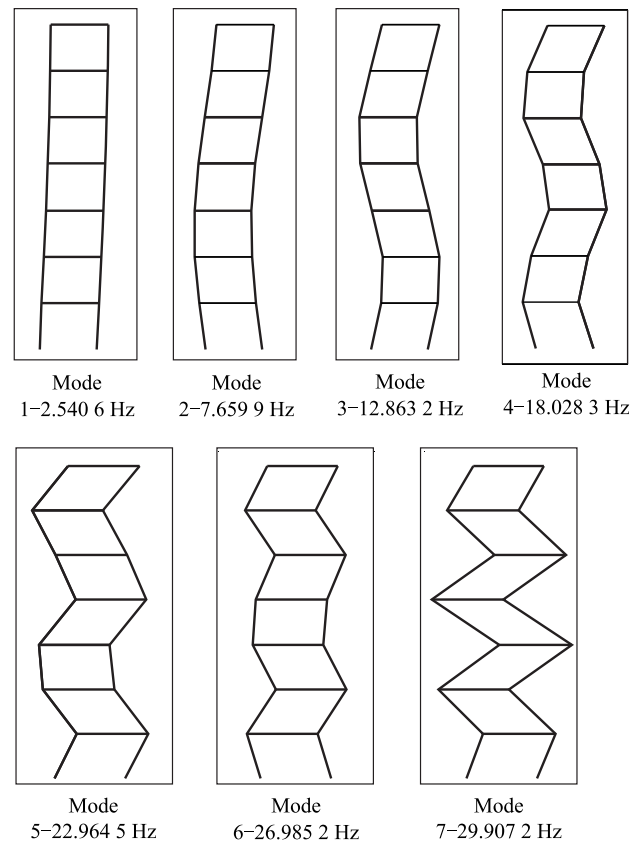


Fig. 3 The first seven measured frequencies and mode shapes of the frame structure

4.2 Data generation

Modal analysis is performed using the reference model to generate input and output data to train the proposed DAE

model. The first seven frequencies and the corresponding mode shapes at 14 beam-column joints are obtained. The elemental stiffness parameters are normalized to the range between 0 and 1, where 1 denotes the intact state and 0 denotes the completely damaged state as mentioned above. For example, if the stiffness parameter of a specific element is equal to 0.9, it means 10% stiffness reduction is introduced in this element. Twenty-one hundred data sets are generated from the baseline model. For each element, the stiffness parameter varies from 1, 0.99, 0.98, \dots , to 0.7 while keeping other elements undamaged. Therefore, 30 data sets are generated for the scenario when a local damage is introduced in a specific element. With 70 elements in the finite element model, totally 2 100 data sets including the first seven frequencies and mode shapes as the input and the elemental stiffness parameters as the output are obtained for the training and validation. These modal information vectors are used as the input to the DAE model and the associated structural elemental stiffness parameters are defined as the output.

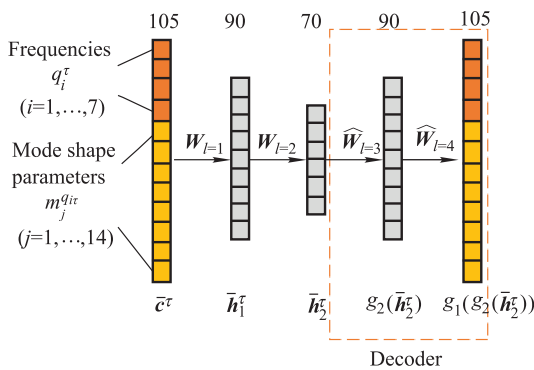


Fig. 4 ADAE model with decoder to reconstruct the original input from encoded feature

4.3 Data pre-processing

Since the concatenated feature \bar{c}^τ contains both the frequencies and mode shapes that are measured in different scales, each feature is normalized separately so that all features lie on a common ground in the learning process. After the normalization process, both the frequencies and the mode shapes will be in the range from -0.5 to $+0.5$. This range is chosen since the operating range of the $\tanh(x)$ activation function falls to -1 to $+1$ and the corresponding inputs should be $+\infty$ and $-\infty$ respectively. The effectiveness of the proposed DAE model is evaluated based on the generated datasets, as described in the following sections.

4.4 Evaluation of the proposed model

The main objective of the proposed DAE model is to learn the relationship between the concatenated feature vector that is fed to the input and the output stiffness parameters.

This is achieved via non-linear dimension reduction followed by a relationship learning phase as described above. Hence the evaluation of the model is performed separately on two components. First the effectiveness of the features learnt via dimensionality reduction component is assessed. Then the quality of learning the mapping as a whole is evaluated. The following two sections describe the evaluation process in detail.

4.4.1 Effective dimension reduction

The hidden feature obtained at the 2nd layer is decoded by introducing the decoder part to the proposed model as shown below.

Decoder facilitates the progressive dimensionality gain as opposed to the gradual dimension reduction via encoding layers. We choose the nodes for each hidden layer as shown in Fig. 4. The pre-training for layer 1 and layer 2 is performed as described in Section 3.1. Once pre-trained weights are obtained (layer 1, layer 2 encoding weights), those encoding weights ($\mathbf{W}_{l=1}^*$, $\mathbf{W}_{l=2}^*$) are used to tie the weights [21] for the corresponding decoders as shown below:

$$\widehat{\mathbf{W}}_{l=3} = (\mathbf{W}_{l=2}^*)^\top, \quad \widehat{\mathbf{W}}_{l=4} = (\mathbf{W}_{l=1}^*)^\top. \quad (6)$$

Then the whole network is fine-tuned again to optimize all layers (L) jointly as below:

$$\begin{aligned} & [\mathbf{W}_{l=1}^*|_{l=1}^L, \bar{\mathbf{b}}_{l=1}^*|_{l=1}^L, \widehat{\mathbf{W}}_{l=1}^*|_{l=L}^1, \widehat{\mathbf{b}}_{l=L}^*|_{l=L}^1] = \\ & \arg \min_{\mathbf{w}_i|_{i=1}^L, \bar{\mathbf{b}}_i|_{i=1}^L, \widehat{\mathbf{w}}_i|_{i=L}^1, \widehat{\mathbf{b}}_i|_{i=L}^1} \sum_{\tau=1}^N \|\bar{\mathbf{c}}^\tau - p(\bar{\mathbf{c}}^\tau)\|_2^2 \quad (7) \end{aligned}$$

where $p(\bar{\mathbf{x}}_i) = g_{L-1}(g_L(f_L(f_{L-1}(\bar{\mathbf{x}}_i))))$ with $L = 2$ and $\mathbf{W}_{l=1}^*|_{l=1}^L$ denotes the encoder's weights while $\widehat{\mathbf{W}}_{l=1}^*|_{l=L}^1$ denotes the decoder's weights as shown in Fig. 4. By jointly optimizing the objective function as shown in (6) towards reconstructing the original feature \bar{c}^τ , the decoder weights will be further optimized to perform non-linear dimensionality gain to produce the original feature at the output. Once the model is fine-tuned, the test samples are fed forward to generate the corresponding reconstructed features at the output thus used to assess the quality of the reconstructed feature against the original feature. It is important to note that the frequencies of the original feature are compared with the frequencies of the reconstructed mode shape while original mode shape values to reconstructed mode shape values with the mean squared error (MSE) criterion. The reconstruction errors that are observed in different phases are summarized in Table 2.

As shown in Table 2, it is conceivable that the fine-tuning stage indeed has a significant improvement to the global non-linearity involved in performing the effective dimension reduction. It is conceivable that the fine-tuning

stage indeed has a significant improvement to the global non-linearity involved in performing the effective dimension reduction. The mean squares error is reduced to 0.09% in the final training stage after fine tuning. The L2 regularization is used on the weights of the DAE model to ensure that it has less space to over-fit the training data. The testing error is as low as 0.77% indicating the accuracy of the proposed DAE model in performing the dimensionality reduction to represent the original input features.

Table 2 Evaluation results of the proposed model with the decoder for reconstruction of original feature

Phase	Dataset size	Validation error (MSE)/%	Error (MSE)/%
Pre-training layer 1	1 200	23.31	5.46
Pre-training layer 2	1 200	39.50	9.26
Fine-tuning	1 200	0.39	0.09
Testing	600	—	0.77

4.4.2 Effective relationship learning

Since the quality of the reduced dimensional feature is satisfactory for reconstructing the original feature as demonstrated above, the quality of the proposed DAE model for learning the relationship mapping is assessed in this section. The cost functions denoted in Section 3.2 in pre-training the layers and fine-tuning the whole network at the end are optimized. The effectiveness of the relationship learning process is evaluated with the MSE between the simulated stiffness parameters and the predicted outputs. The test samples are fed into the fine-tuned DAE model to generate the elemental stiffness parameters in the output. One hundred and five modal parameters including seven frequencies and their corresponding mode shapes on 14 beam-column joints are included in the input vector, and 70 stiffness parameters in the output vector. Ninety, seventy and seventy neural nodes are chosen for the first, the second and third hidden layers of the DAE model as shown in Fig. 3. The predicted output stiffness values are compared against the expected stiffness parameters to observe how close the predicted outputs are to the true values. The errors that are observed in the different training phases of the DAE model are summarized in Table 3.

Table 3 Evaluation results of the proposed model

Phase	Dataset size	Validation error (MSE)/%	Error (MSE)/%
Pre-training layer 1	1 200	3.84	0.90
Pre-training layer 2	1 200	2.87	0.67
Pre-training layer 3	1 200	2.86	0.69
Fine-tuning	1 200	2.20	0.54
Testing	600	—	2.9

Different layers of DAE training have an effect on the global non-linearity involved in performing the effective dimensionality reduction while the error is further reduced

in the latter training stages especially after fine-tuning as shown in Table 3. The validation error gradually decreases along with the depth of the network. It is also indicated that the error between the identified stiffness parameter values and the true values is 2.9% on the test dataset, demonstrating that the accuracy and efficiency of the proposed DAE model for structural damage identification are highly satisfactory.

4.4.3 Single damage identification

The model is firstly evaluated with single structural damage cases and an example from the test data set is shown in Fig. 5. The identified stiffness reduction is very close to the actual value where the values of positive and negative false identifications are very small and close to zero while the identified locations of the damage are also clearly distinguishable.

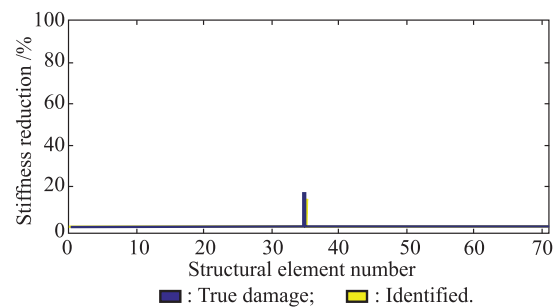


Fig. 5 An example of single damage identification

4.4.4 Multiple damage identification

Multiple structural damage identification is challenging, and needs more precision on the identification of accurate stiffness reductions at the exact stiffness elements compared to the single damage cases. The proposed model is applied for such cases and an example is shown in Fig. 6. The identified stiffness reductions are very close to the actual values with very small false identifications. Hence the identified locations of the damage also show a very good level of accuracy.

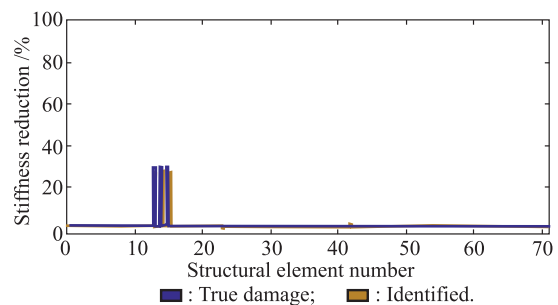


Fig. 6 An example of multiple damage identification

5. Conclusions

This paper proposes a method of applying DAE for structural condition monitoring and damage identification. The relationship that exists in the modal information, such as frequencies and mode shapes, to structural stiffness parameters is trained for SHM. The proposed DAE model is developed not only for dimensional reduction that preserves the necessary information but also for identifying the relationship between the input modal information and the output stiffness parameters. Effective dimension reduction and accurate relationship training are achieved with the proposed DAE model based on the deep learning architecture and the pre-training algorithm. The proposed model has produced excellent stiffness prediction and damage identification with some testing data on a steel frame structural model. The work can be extended to combine more vibration properties and consider noise effect towards a robust training thus improving the performance for structural health monitoring and damage identification.

References

- [1] YUNC B, YU H, BAHNGE Y. Joint damage assessment of framed structures using a neural networks technique. *Engineering Structures*, 2001, 23(5): 425–435.
- [2] BAKHARY N, HAO H, DEEKSA J. Damage detection using artificial neural network with consideration of uncertainties. *Engineering Structures*, 2007, 29(11): 2806–2815.
- [3] YAOA C C. Separating the polynomial-time hierarchy by oracles. *Proc. of the 26th Annual Symposium on Foundations of Computer Science*, 1985: 1–10.
- [4] HÅSTAD J, GOLDMANN M. On the power of small-depth threshold circuits. *Computational Complexity*, 1991, 1(2): 113–129.
- [5] BENGIO Y, DELALLEAU O, ROUXN L. The curse of highly variable functions for local kernel machines. *Proc. of the International Conference on Neural Information Processing Systems*, 2005: 107–114.
- [6] BENGIO Y, LECUN Y. Scaling learning algorithms towards AI. *Large-Scale Kernel Machines*, 2007, 34(5): 1–41.
- [7] ZHU Z, LUO P, WANG X, et al. Recover canonical-view faces in the wild with deep neural networks. <https://arxiv.org/ftp/arxiv/papers/1404/1404.3543.pdf>.
- [8] ZHU Z, LUO P, WANG X, et al. Deep learning identity-preserving face space. *Proc. of the IEEE International Conference on Computer Vision*, 2013: 113–120.
- [9] PATHIRAGE S N, LI L, LIU W, et al. Stacked face denoising auto encoders for expression-robust face recognition. *Proc. of the International Conference on Digital Image Computing: Techniques and Applications*, 2016: 1–8.
- [10] KAN M, SHAN S, CHANG H, et al. Stacked progressive auto-encoders (SPAe) for face recognition across poses. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014: 1883–1890.
- [11] COLLOBERT R, WESTON J. A unified architecture for natural language processing. *Proc. of the 25th International Conference on Machine Learning*, 2008: 160–167.
- [12] SALAKHUTDINOV R, HINTON G. Semantic hashing. *International Journal of Approximate Reasoning*, 2009, 50(7): 969–978.
- [13] HADSELL R, ERKAN A, SERMANET P, et al. Deep belief net learning in a long-range vision system for autonomous off-road driving. *Proc. of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2008: 628–633.
- [14] BENGIO Y. Learning deep architectures for AI. *Foundations & Trends® in Machine Learning*, 2009, 2(1): 1–127.
- [15] HINTONG E, SALAKHUTDINOV R. Reducing the dimensionality of data with neural networks. *Science*, 2006, 313(5786): 504–507.
- [16] PARKHIO M, VEDALDI A, ZISSERMAN A. Deep face recognition. *Proc. of the British Machine Vision Conference*, 2015: 41.1–41.12.
- [17] VINCENT P, LAROCHELLE H, LAJOIE I, et al. Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *Journal of Machine Learning Research*, 2010, 11(12): 3371–3408.
- [18] FRISWELL M, MOTTERSHEADJ E. *Finite element model updating in structural dynamics*. Dordrecht: Kluwer Academic Publishers, 2013.
- [19] LI J, LAWS S, DING Y. Substructure damage identification based on response reconstruction in frequency domain and model updating. *Engineering Structures*, 2012, 41(3): 270–284.
- [20] LI J, HAO H, LO J V. Structural damage identification with power spectral density transmissibility: numerical and experimental studies. *Smart Structures & Systems*, 2015, 15(1): 15–40.
- [21] HINTONG E, OSINDERO S, TEHY W. A fast learning algorithm for deep belief nets. *Neural Computing*, 2006, 18(7): 1527–1554.

Biographies



PATHIRAGE Chathurdara Sri Nadith was born in 1987. He received his B.S. degree in computer science in 2009 and is currently studying for his Ph.D. degree in the Department of Computing, Curtin University, Australia. He worked as a specialist software engineer at London Stock Exchange group. His research interests lie in the field of artificial intelligence in computer vision, concepts of human visual perception via combined intellect of image processing, sparse sensing and deep learning.

E-mail: c.nadithpa@postgrad.curtin.edu.au



LI Jun was born in 1984. He received his Ph.D. degree from Hong Kong Polytechnic University in 2012. He is currently a senior lecturer in Department of Civil Engineering/Centre for Infrastructural Monitoring and Protection at Curtin University. His research interests include structural health monitoring and signal processing techniques.

E-mail: junli@curtin.edu.au



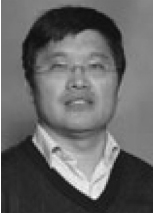
LI Ling was born in 1965. She obtained her B.S. degree in computer science from Sichuan University, China, and Ph.D. degree in computer engineering from Nanyang Technological University, Singapore. She is now an associate professor in the Department of Computing at Curtin University in Australia. Her research interests are mainly in computer vision and graphics, and artificially intelligent beings.

E-mail: L.Li@curtin.edu.au



HAO Hong was born in 1962. He received his B.E. degree from Tianjin University, China, and Ph.D. degree from the University of California at Berkeley. He is a John Curtin Distinguished Professor and the director of Centre for Infrastructural Monitoring and Protection at Curtin University, Australia. His research interests include earthquake engineering, blast engineering and structural condition monitoring.

E-mail: hong.hao@curtin.edu.au



LIU Wanquan was born in 1965. He received his B.S. degree in applied mathematics from Qufu Normal University, China, in 1985, M.S. degree in control theory and operation research from Chinese Academy of Sciences in 1988, and the Ph.D. degree in electrical engineering from Shanghai Jiaotong University, in 1993. He once held the ARC fellowship, U2000 fellowship and JSPS fellowship and

attracted research funds from different resources over 2.4 million Australian dollars. He is currently an associate professor in the Department of Computing at Curtin University and is in editorial board for nine international journals. His current research interests include large-scale pattern recognition, signal processing, machine learning, and control systems.
E-mail: W.Liu@curtin.edu.au