Citation
Fan, G. and Li, J. and Hao, H. 2020. Vibration signal denoising for structural health monitoring by residual convolutional neural networks. Measurement: Journal of the International Measurement Confederation. 157: ARTN 107651
http://doi.org/10.1016/j.measurement.2020.107651

Vibration Signal Denoising for Structural Health Monitoring by Residual Convolutional Neural Networks

Gao Fan, Jun Li, Hong Hao *

Centre for Infrastructure Monitoring and Protection, School of Civil and Mechanical Engineering, Curtin University, WA 6102, Australia

Emails: gao.fan@postgrad.curtin.edu.au; junli@curtin.edu.au; hong.hao@curtin.edu.au

Abstract

In vibration based structural health monitoring (SHM), measurement noise inevitably exists in the vibration data, which significantly influences the usability and quality of measured vibration signals for structural identification and condition monitoring. As a result, there is a high demand for developing effective methods to reduce noise effect, especially in harsh and extreme environment. This paper proposes a vibration signal denoising approach for SHM based on a specialized Residual Convolutional Neural Networks (ResNet). Dropout, skip connection and sub-pixel shuffling techniques are used to improve the performance. The effectiveness and robustness of this developed approach are validated with acceleration data measured from Guangzhou New TV Tower. The results show that the proposed approach is effective in improving the quality of the acceleration data with varying levels of noises and different types of noises. Modal identifications based on signals contaminated with intensive noise and de-noised signals are conducted. Modal information of the weakly excited modes masked by noise and the closely spaced modes can be clearly and accurately identified from the de-noised signals, which could not be reliably identified with the original signal, indicating the effectiveness of using this developed approach for SHM. Besides white noise, a group of data contaminated with pink noise, which is not included in the training data, is also tested. Good results are obtained. The developed ResNet extracts high-level features from the vibration signal and learns the modal information of structures automatically, therefore it can well preserve the most important vibration characteristics in vibration signals, and can assist in distinguishing the physical

modes from the spurious modes in structural modal identification.

Keywords: Denoising; Modal identification; Noise; Residual convolutional neural network; Structural health monitoring; Vibration signal

* Corresponding Author, John Curtin Distinguished Professor, Centre for Infrastructural Monitoring and Protection, School of Civil and Mechanical Engineering, Curtin University, Kent Street, Bentley, WA 6102, Australia. Email: hong.hao@curtin.edu.au.

1. Introduction

Ensuring a high degree of availability, reliability and safety of civil infrastructure is critical for the society to well function. With the ageing of structures, the strength of current structures may degrade below the designed safety strength which threatens the public safety and may even cause catastrophic failure of structures. The degradation can be caused by the progressive deterioration and the accumulated damage induced by adverse operating conditions and extreme events. Vibration based Structural Health Monitoring (SHM) technology has attracted significant attention in recent decades. SHM systems permanently or periodically measure the dynamic responses of structures for early warning of abnormal status and assist engineers and asset owners to make better decisions on maintenance and operation. An increasing number of vibration based SHM systems have been installed on infrastructures such as large scale bridges [1, 2] and high rise buildings [3, 4]. Measured dynamic responses containing crucial information of structures are further processed to obtain structural vibration characteristics, extract damage sensitive features for assessing conditions and detecting local damages in structures.

Effective and accurate condition assessment of structures usually demands high-quality data that contain significant vibration information of structures and low-level noises in measurements. However, the in-field testing condition of civil structures is difficult to control. Measured dynamic responses can be easily contaminated by strong noise from various sources such as environmental noise, measurement noise and instrumental noise. Strong noise may mask changes induced by minor damages in structural responses, leading to modal identification or damage detection inaccurate or virtually impossible in such cases. Consequently, noise immunity ability is one of the critical criteria for evaluating the effectiveness of vibration based SHM methods. It has been reported that 10-20% noise injected into vibration signals was considered as a high noise level in numerical studies [5, 6] when validating damage detection methods. Considering substantial uncertainties during in-field tests, measurement noise may exceed 20% noise level easily in situations with some poor testing conditions. Therefore, besides improving noise immunity of SHM methods for modal analysis or damage detection, another important research topic named vibration signal denoising, which aims to remove

noise components or mitigate noise effect from measurements without affecting quality of vibration signals, has gained significant attention. Signal denoising can improve the usability and quality of the measured signals before any specific analysis.

Recent researches have been conducted to perform the vibration signal denoising in the time, frequency and time-frequency domains for different types of data [7]. Conventional time domain averaging methods are more suitable for periodic signals [8]. The denoising effect on other kinds of signals may be limited due to the varying vibration frequency or amplitude of signals. Filtering methods, such as low-pass filtering and band-pass filtering, etc., are typical frequency domain methods that eliminate noise outside the user-defined frequency band of interest. The prior knowledge of structures including modal frequencies and the corresponding variation ranges induced by operational and environmental condition changes needs be known, which limits the application of those methods. Time-frequency domain methods, such as wavelet transform based denoising techniques have been widely investigated [9, 10]. Those techniques consider both time and frequency characteristics, which are suitable for both stationary and nonstationary signals. However, wavelet transform based methods require manually selected and tuned optimal parameters such as the suitable wavelet basis regarding target signals, the proper threshold value of wavelet transform layers and packets to discriminate and remove noise while avoid losing useful components [7, 11]. Other methods based on singular value decomposition [12], empirical mode decomposition [13] and global projection [14] have also been developed to conduct specific denoising tasks. Among all of the abovementioned methods, different parameters are required for optimizing, resulting in the effectiveness of signal denoising closely dependent on engineers' empirical experiences. Recently, machine learning especially deep learning based techniques such as deep Artificial Neural Network (ANN), Denoising Auto-Encoder (DAE) [15] and Convolutional Neural Networks (CNN) [16] have been extended to denoise image [17, 18], audio [19, 20] and medical domain signals [21] in the time, frequency or time-frequency domains with its original waveform or selected features. These networks learn nonlinear relationships between noisy and original signals from training data, and then map the newly obtained noisy signals to denoised ones. Compared with existing methods, deep learning based

techniques for signal denoising automatically extract the noise insensitive features from measured signals to realize the denoising, which effectively avoid manual selecting and fine-tuning of the abovementioned parameters.

Currently, deep learning techniques are widely applied in the SHM community [22] such as damage and crack detection using images [23-25], vibration signals [26-31] and vibrational characteristics [32, 33], abnormal measurement diagnosis [34] and vibration lost data recovery [35]. Even there are applications of deep learning techniques for denoising signals in other fields, however, to the authors' best knowledge, researches on developing vibration signal denoising method for SHM using deep learning techniques have not been reported. Real-time SHM systems continuously measure and record a large amount of vibration data, which fit well with the concept of "big data", that is a crucial requirement of applying deep learning techniques.

This paper presents the development and application of Residual Convolution Neural Networks (ResNet) with a bottleneck structure incorporated with skip connection, dropout and sub-pixel shuffling techniques for denoising measured vibration signals. The proposed ResNet is trained in a supervised manner using training datasets, which consist of paired originally measured and synthetic noisy signals. Through training, the important features of training data are learned and network parameters including weights and biases of convolutional layers are tuned. The trained network is then used to map extracted features from input noisy signals to the denoised ones for realizing the denoising. It should be highlighted that the exact noise-free vibrational signal is not required for training the network. The proposed approach requires less prior knowledge of the structure and human intervention and is very applicable in practice. Validations of the proposed method by using the measured vibration data from Guangzhou New TV Tower are conducted to investigate the accuracy and performance of the proposed approach.

In the following sections of this paper, the architecture, selected hyperparameters of the designed ResNet and the involved techniques to enhance the effectiveness of the developed approach will be introduced in Section 2. Experimental studies on a super-high slender structure, namely Guangzhou New TV Tower, are conducted for evaluating the effectiveness and performance of using the proposed

method for denoising the vibration signals with weakly excited and closely spaced modes. The details of the structure, the procedure to implement the denoising and the evaluation of the denoised results with comparison to a classic wavelet transform based method will also be introduced in Section 3. To further verify the robustness of the proposed method, Section 4 presents an evaluation of the trained network by using noisy signal contaminated with pink noise that is not included in the training data. Finally, the conclusions made from the denoising results and the recommendations for further studies are discussed in Section 5.

2. Methodology

CNN as one of the most remarkable progresses in machine learning is gaining significant research interest recently. Instead of extracting features by fully connection layers as traditional ANN, CNN formed by stacked convolutional layers can automatically extract higher-level features of input data using convolutional kernels. Without fully connection, CNN has much fewer trainable parameters and minor overfitting. The proposed ResNet inspired by a previous study on audio super resolution [36] is a fully convolutional feedforward network. The developed ResNet has a bottleneck structure which originates from Deep Auto-Encoder [37] and is named as Convolutional Encoder-Decoder Network [38]. The bottleneck structure encourages models to gradually extract higher-level features of input data by reducing the length of encoder features. Meanwhile, the reduction of dimension leads layers to only keep robust features and sift noise components. The proposed ResNet utilizes advantages of both bottleneck structure and convolutional layers by replacing all the fully connected encoder and decoder layers by convolutional layers. Therefore, compared to Auto-Encoder, convolutional layers consisted of multiple convolutional kernels can extract more abundant and comprehensive features with minor overfitting effect.

The main contribution of this work is the sophisticatedly designed ResNet which is specialized in performing the denoising task of vibration signals. It is fully convolutional so that the network can process input data with different lengths. All the convolutional layers are modified as one-dimensional which is very adaptable for one-dimensional vibration signals. Furthermore, all these

embedded techniques are carefully designed with the following features:

- Skip connection greatly improves the training efficiency of ResNet.
- Dropout technique mitigates the overfitting of ResNet.
- Sub-pixel shuffling technique upscales feature maps and also effectively realizes the function of deconvolution.

Using deep learning techniques to denoise vibration signals has many advantages, such as less human intervention on parameter selection of filters and automated feature extraction. Previous studies utilized the deep learning technique to denoise photos, speeches and electrocardiograms. However, these techniques have not been developed and applied to vibration signal denoising in SHM. This paper proposes a vibration signal denoising approach based on deep learning techniques, i.e. ResNet, to remove the noise in the vibration measurement data. The feasibility and effectiveness of the proposed technique on vibration signal denoising are demonstrated. This is the first attempt of using a deep learning technique in denoising the vibration signals.

2.1 The Architecture of the proposed ResNet

The architecture of the proposed ResNet consists of an input layer, eight convolutional layers and one output layer, as shown in Figure 1. The input and output layers have the same number of neurons, equal to the sampling points of input signals to be denoised. Among the convolutional layers, there are three compression layers, one bottleneck layer, three reconstruction layers and a final resize layer. The number of bottleneck and resize layers is fixed according to the bottleneck structure. The depth of the network is determined by testing the performance of ResNet with 1 to 6 pairs of compression and reconstruction layers. 3 pairs of layers are selected since the performance stops gaining after using 3 pairs of layers, while the computational burden continues increasing. The compression and bottleneck layers reduce the dimension of feature maps and extract the higher-level representation of input signals gradually by using convolution kernels with a stride of two. Meanwhile, the number of feature maps are doubled to enrich the higher-level features. The output feature maps of the bottleneck layer possess the smallest dimensionality but represent the highest level of features of input data.

Then, ResNet turns to reconstruct output denoised signals only considering the representative information of input signals by the reconstruction layers. The reconstruction layers gradually enlarge the dimension and reduce the number of feature maps.

One special operation named skip connection is implemented in the reconstruction layers that concatenate output features of reconstruction layers with output features of compression layers in mirrored positions as shown in Figure 1. The paired original and noisy signals are quite similar especially when only a low level of noise exists in the noisy signal. Due to the symmetric architecture of ResNet, the layers in the mirrored position are sharing a large number of features such as the waveform and structural frequencies information. It has also been observed that it is easier to learn residual between the output and input, rather than only the input. The skip connections between each pair of layers effectively shuttle the features extracted in bottom layers to top layers to provide supplementary information that is lost during convolution [39]. It also allows gradients to be back-propagated to bottom layers directly, which mitigates the gradient vanishing and provides the ability to train deeper networks efficiently [40]. On the other hand, skip connections or residual connections can be used to skip the training of a few layers. One common issue of deep learning based techniques is that if the depth of networks keeps increasing, the accuracy will start to saturate at one point and even degrade. In general, the optimal number of layers required for a neural network might depend on the complexity of the datasets which is hard to determine. Therefore, instead of treating the number of layers as an important hyperparameter to tune, by adding skip connections to the network, allowing the network to skip training for the layers that are not useful and do not add value in overall accuracy is beneficial and may optimally tune the number of layers during training.

The configurations of the used ResNet are shown in Table 1. The kernel number and kernel size are initialized similarly as Ref. [36] and determined by final turning. Currently, there are no certain rules for selecting hyperparameters of deep learning models. Most of studies tune hyperparameters by trial and error. It should be noticed that designing an adaptive architecture is much more important than optimizing hyperparameters. In this study, it is found that the effectiveness can be guaranteed by simply initializing a large model that contains a considerable number of neurons. The model with

optimized hyperparameters as shown in Table 1 has a minor improvement than the initial ones. The number of kernels starts from 128 which is a multiple of 2. The kernel number is doubled among the compression layers and halved among the reconstruction layers to keep the symmetry of the network by using sub-pixel shuffling (will be introduced in Section 2.4). Since the feature maps are shrunken and then enlarged among the layers, the kernel sizes of layers also change to fit with the corresponding size of the feature maps. To guarantee the output of bottleneck layer as an integer, the input size should be a multiple of 2^3, that is 8. Owning to the fully convolutional and symmetric architecture, the input length can be any times of 8. However, it is recommended to use input data which is larger than 64 points for utilizing the full capability of the network. In this study, the dimension of feature maps in Figure 1 and the input and output shapes of each layer are listed in Table 1, based on the input size of 1024 points, which is the same as that of the training data samples used in the subsequent experimental studies.

Table 1. The detailed configurations of the used ResNet.

Layer	Kernel number	Kernel size	Stride	Padding	Input shape	Output shape	sub-pixel shuffling
Input	-	-	-	-	(1, 1024)	(1, 1024)	-
C1	128	64	2	Same	(1, 1024)	(128, 512)	N
C2	256	32	2	Same	(128, 512)	(256, 256)	N
C3	512	16	2	Same	(256, 256)	(512, 128)	N
Bottleneck	1024	8	2	Same	(512, 128)	(1024, 64)	N
R3	1024	16	1	Same	(1024, 64)	(1024, 128)	Y
R2	512	32	1	Same	(1024, 128)	(512, 256)	Y
R1	256	64	1	Same	(512, 256)	(256, 256)	Y
Final	2	8	1	Same	(256, 512)	(1, 1024)	Y
Output	-	-	-	-	(1, 1024)	(1, 1024)	-

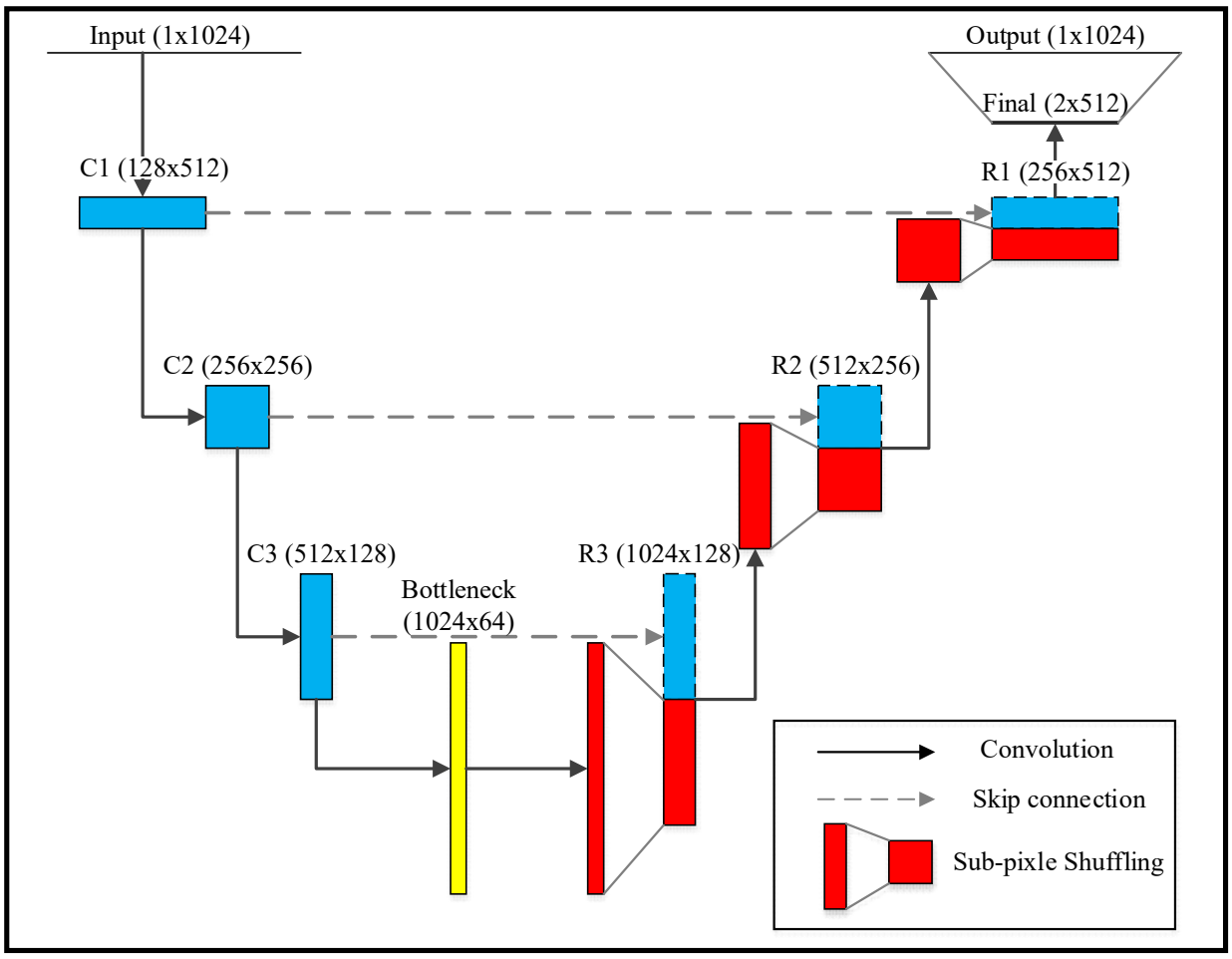


Figure 1. The architecture of the used ResNet with a bottleneck structure.

The compression layers, bottleneck layer, reconstruction layers and the final resize layer have different functions with different operations. The compression layers and bottleneck layer involve two operations: (1) Convolution of the input; and (2) Nonlinear activation of the output using a Leaky rectifier. The Leaky rectifier will be introduced in Section 2.3. The reconstruction layers involve four operations which are: (1) Convolution of the input features; (2) Nonlinear activation of the output using a Leaky rectifier; (3) Resizing of the activated output by the sup-pixel shuffling operation detailed in Section 2.4; and (4) Final concatenation of the activated output with the features from the compression layer in the mirrored position (skip connection). In addition, the last layer contains the convolution and resizing operations.

The ResNet is trained in a supervised manner using paired of noisy and original signals as input and output. Hence, the objective of the used ResNet is to minimize the discrepancy between predicted denoised signal $f_\phi(x)$ and the corresponding original signal y_k. The objective function selected to optimize the parameters of ResNet is defined as the mean of L2-norm error of each sample in a mini batch. The objective function is expressed as

$$Loss = \frac{1}{N}\sum_{k=1}^{N}\left\|y_k - f_\phi(x_k)\right\|_2^2 \quad (1)$$

where N is the number of samples in each mini batch. The ResNet is simply depicted as a nonlinear model $f_\phi(x)$ that is parameterized by ϕ. Consequently, the target of the training process is to determine ϕ by minimizing the *Loss* function based on the entire training datasets.

2.2 Dropout technique

Overfitting is a common issue in training deep neural networks, which means the tuned parameters of networks are too closely fitted with training data. Dropout technique is a recently developed training strategy which is embedded in the training process of this study to address the overfitting issue. Dropout means dropping out neurons and disconnecting those neurons with the adjacent input and output layers [41]. Neurons in networks are dropped randomly when training with a different batch of samples to break up the co-adapted sets of neurons. Those remained neurons are trained more robustly and the generalization capacity of the network is enhanced. Dropout technique in signal denoising tasks also omits the particular relationships concealed in noisy signals by introducing perturbations [42]. Each neuron has an independent probability p to be dropped. p=0.5 is suggested by the inventors and used in this study [41].

2.3 Leaky Rectified Linear Unit (Leaky ReLU)

The relationship between noisy and original signals is undoubtedly complex. Linear functions are insufficient to model this relationship. Therefore, an activation function that transfers the mapping from linear to nonlinear is added after the convolution operation to support ResNet to model such a

sophisticated nonlinear relationship. Leaky ReLU [43] as an improved version of ReLU [44] is an advanced activation function. Leaky ReLU has a nonlinear gradient over its whole domain as follows

$$f(x) = \begin{cases} x, & if\ x > 0 \\ 0.01x, & \text{otherwise} \end{cases} \quad (2)$$

Compared with *'sigmoid'* or *'tanh'* activation functions which squash input into a very narrow output range ([0 -1] for *'sigmoid'*, and [-1 1] for *'tanh'*), Leaky ReLU has a broad range of output which prevents the gradient vanishing. Meanwhile, it allows a non-zero small gradient to overcome the defect of ReLU, that is, a neuron is potentially not be activated once a large gradient passes it [43].

2.4 Sub-pixel Shuffling Operation

To implement the skip connection, the lengths of feature maps in bottom layers to higher layers should be consistent. As shown in Figure 1, the output feature maps of the reconstruction layer have half of the length but double the number of the output of the corresponding compression layer. The sub-pixel shuffling operation is involved to resize the output feature maps from the reconstruction layers. Figure 2 demonstrates an example of the sub-pixel shuffling with four feature maps consisting of 6 features each. It divides the feature maps into two groups and combines two feature maps in the same order of each group as one by interpolating one to another. It is a one-dimensional case of a sub-pixel convolution layer [45]. This is an efficient operation for shuffling, which costs less computation than deconvolution. It has also been attested to have strong workability that introduces less artefact in the output [46].

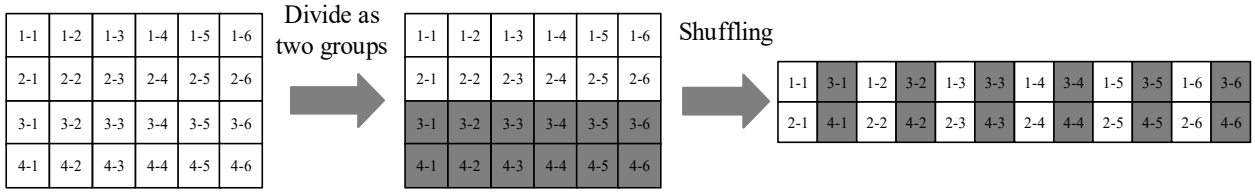


Figure 2. An example of the one-dimensional sub-pixel shuffling with 4 feature maps and 6 elements in each map.

2.5 Evaluation criteria

The direct evaluation of the quality of denoised signals is conducted by using Signal to Noise Ratio (SNR). Given the original signal y and the noisy signal y', noise is estimated as the difference between two signals as y - y'. SNR in decibels (dB) is expressed as,

$$\mathrm{SNR} = 10 \times \log_{10} \frac{P_{signal}}{P_{noise}} = 10 \times \log_{10} \frac{\sum_1^n {y_i}^2}{\sum_1^n (y_i - y_i')^2} \tag{3}$$

where n is the total number of sampling points. SNRs of noisy and denoised signals are computed against the corresponding original signals, respectively. Furthermore, the quality of noisy and denoised signals in the frequency domain is also considered. Original, noisy and denoised signals are transferred to the frequency domain by using Fast Fourier Transform (FFT). The discrepancy of signals in the frequency domain is evaluated by using L_2-norm, and is calculated as

$$\xi = \frac{\|f - f'\|_2^2}{\|f\|_2^2} = \frac{\sqrt{\sum_1^n (f_i - f_i')^2}}{\sqrt{\sum_1^n {f_i}^2}} \tag{4}$$

where f is the original signal in the frequency domain, and f' is the noisy or denoised signals in the frequency domain.

Modal parameters including natural frequencies, damping ratios and mode shapes are widely used in vibration-based SHM for damage detection and condition assessment. This study also evaluates the effectiveness of using the proposed approach for signal denoising by comparing the identification accuracy of modal parameters from noisy and denoised signals. Modal analysis of vibration signals is conducted by using a frequency domain non-parametric method named Frequency Domain Decomposition (FDD) [47], which is a mature operational modal analysis method that has been successfully applied in many studies on civil structures. Modal analysis by using FDD method demands engineering experiences to distinguish physical modes with spurious modes caused by noise and weakly excited local modes. In this case, signals with a low-level noise can produce clearer output diagrams and therefore promote users to select structural modes more accurately with less false identifications. The comparison of identification results is performed from both qualitative and quantitative analyses. The qualitative evaluation of the proposed approach is conducted by observing the output of FDD, which shows decomposed singular values of power spectral matrix. On the other

hand, quantitative evaluation of results is based on the identification accuracy of using noisy, denoised and original signals for modal identification. The accuracies of frequencies and damping ratios are evaluated by relative error and absolute error, respectively. Modal Assurance Criterion (MAC) is used to investigate the accuracy of identified mode shapes, which measures the similarity between two mode shape vectors.

3. Experimental Validations

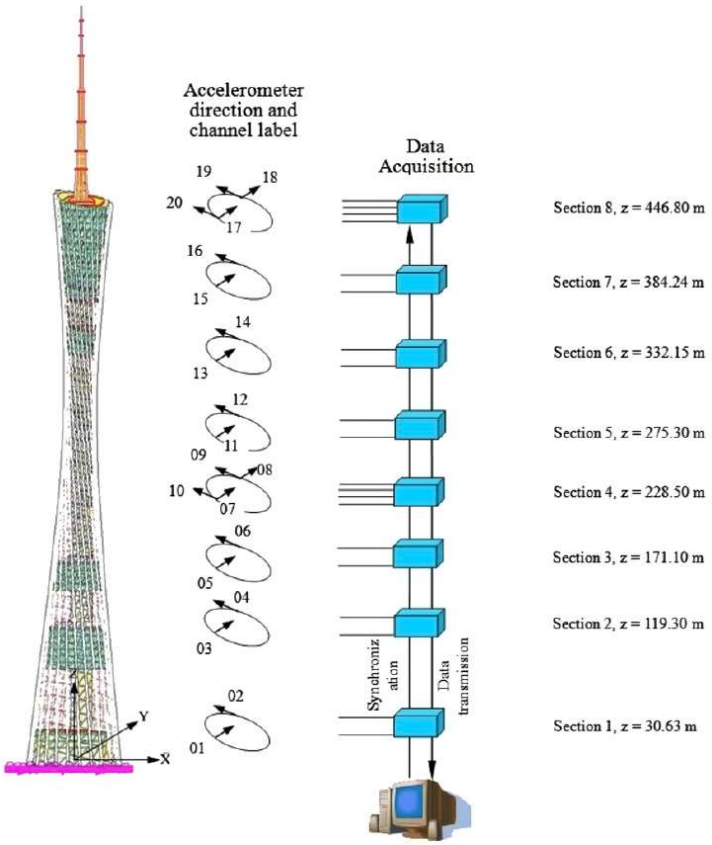
To demonstrate the effectiveness and robustness of using the proposed approach based on ResNet for vibration signal denoising, experimental studies using the in-situ acceleration data measured from a long-term SHM system on Guangzhou New TV Tower are conducted. The vibration signals of Guangzhou New TV Tower contain weakly excited and closely spaced vibration modes. Improving quality of signals and removing noise components without compromising the identification accuracy of these two kinds of modes are challenging for conventional signal denoising methods, especially for selecting bandwidths of filters for the frequency domain methods and selecting wavelet basis and noise eliminating thresholds for the time-frequency domain methods. In this experimental study, instead of training multiple networks for denoising signals of each channel, two ResNets for denoising acceleration responses of Guangzhou New TV Tower in the short and long-axis directions are trained separately for improving efficiency. These two ResNets have the same architecture and hyperparameters but are trained separately with specific datasets to achieve the signal denoising tasks in two directions.

3.1 Guangzhou New TV Tower and its SHM system

Guangzhou New TV Tower is a 604 meters tall slender structure located in Guangzhou, China. It was completed in 2009 and started operation from 2010. As shown in Figure 3(a), Guangzhou New TV Tower is a supertall tube-in-tube structure, which consists of a reinforced concrete inner tube and a steel outer tube with concrete-filled-tube (CFT) columns. Totally 37 floors connecting the inner tube and the outer tube are constructed for different functions such as offices, entertainment, catering

and especially for emission of television signals. A comprehensive SHM system including more than 600 sensors such as accelerometers, strain gauges and other types of sensors, has been installed on the tower for real-time monitoring to acquire the vibration responses and environmental conditions of the structure at both construction and in-service stages. Figure 3(b) shows the deployment of accelerometers and data acquisition system for the in-service stage. The installed accelerometers measure the acceleration responses of the tower in both short-axis direction with sensors numbers 1, 3, 5, 7, 8, 11, 13, 15, 17 and 18, and long-axis direction with sensors numbers 2, 4, 6, 9, 10, 12, 14, 16, 19 and 20. The sensor directions are properly aligned before measurement. The sampling frequency of the data acquisition system is set as 50 Hz and raw measurements are filtered by a high-pass filter with a 0.05 Hz cutoff frequency. More details of the structure and the installed SHM system can be found in Ref. [4]. The previous studies have shown the effectiveness of the SHM system [48, 49] and it has become a benchmark platform for high-rise structures.

(a)



(b)

Figure 3. (a) Guangzhou New TV Tower; (b) Sensor deployment and data acquisition [49].

3.2 Preparation of the datasets for training and testing the network

Quality of datasets significantly affects training performance of networks. An abundant dataset can properly fine-tune trainable parameters and provides a robust neural network. Long-term SHM systems continuously collect a massive amount of vibration responses, such as strain and acceleration data. In this study, totally 24 hours of the long-term monitoring acceleration data from the in-service stage are processed as the datasets to validate the proposed approach. Measured acceleration data from two directions are separated but processed with the same procedure. These 24 hours data are split as 18 hours, 3 hours and 3 hours (75%, 12.5% and 12.5%) for training, validation and testing of the network. The training data are used to train and fine-tune the trainable parameters of the network. The validation data are used to evaluate the network during training process for real-time checking the convergence and overfitting. It should be mentioned that the validation data will not be involved

in tuning the parameters. The testing data are prepared to test the performance of trained networks. When considering the lifetime of the SHM system, it is practical and effective to train a robust ResNet for denoising incoming data by using only 18 hours of measured data

Pre-processing of the original measurements is based on the modal analysis results of the structure in a previous study [49] using field measurements and a calibrated finite element model. Vibration modes of this structure are closely spaced, where 15 modes can be observed from 0 Hz to 2 Hz [50]. In general, the first several modes contribute most of structural vibrations. Hence, the frequency bandwidth to be considered in this study is selected as 0 Hz and 2 Hz. To eliminate the redundant information merged in original measured signals, low-pass filtering with a cut-off frequency of 2.5 Hz is first implemented. After filtering, signals are processed by downsampling the sampling frequency from 50 Hz to 5 Hz. These two pre-processing steps are not compulsory but are important for efficient training of the used ResNet. The filtering process eliminates high frequency noise components and the very high order local vibration modes that beyond the consideration. The filtered signal contains the most important information, which means that a relatively simple nonlinear relationship needs to be constructed to map the input to the output in the training process. Since the architecture of ResNet remains unchanged, the training time mainly depends on the size of training datasets. The downsampling process reduces the size of datasets by ten times and consequently improves the training efficiency significantly. After the data are properly pre-processed, the training and validation datasets are generated in pairs of input and output data corresponding to the noisy and original acceleration signals. Since collecting purely clean acceleration signals is impossible for any practical vibration test, the original signal used as output of the network has been already contaminated with varying levels and different kinds of noises in experimental tests. The noisy signals are generated by further injecting white noise to the corresponding pre-processed original signals as

$$Signal_{noisy} = Signal_{clean} + Noise \times N_l \tag{5}$$

where *Noise* is a normally distributed random vector with zero mean and standard deviation of the original vibration signals [51], and N_l is the noise level. When generating the training and validation

datasets, an original signal will be paired with 9 noisy signals respectively. These 9 noisy signals are contaminated by a certain level of noise from 10% to 90% with a 10% increase step. In addition, the *Noise* vector for each noisy signal and each channel is random and different, which is realistic for in-field measurements. The paired acceleration signals are finally normalized to the range from -1 to 1 to facilitate training stability following the below equation

$$accleration_{normalized} = \frac{accleration}{max\ (|accleration|)} \tag{6}$$

The paired normalized original signal and noisy signals are depicted in Figure 4. The parameters of a network may not be properly tuned and the computational demand will be significant if all the training data are served as input to the network at the same time. Therefore, the training and validation samples (one-hour long each) are formed as small patches, that are segmented from full-time histories of the processed acceleration data and are then grouped as small batches to train the networks in turn. To generate the training and validation samples, a window with a length of 1024 points is implemented for scanning all the training or validation data from the beginning to the end with a 50% overlap. All the generated samples are mixed together with a random order to form the training or validation datasets. Testing datasets are generated using three hours of acceleration data. These three hours data are processed as three groups of testing data named Testing data 1 to 3. Each group contains 9 testing samples that are 9 one-hour-long noisy signals with different random white noises and noise levels. As a result, two sets of training, validation and testing datasets for short-axis and long-axis directions are produced respectively.

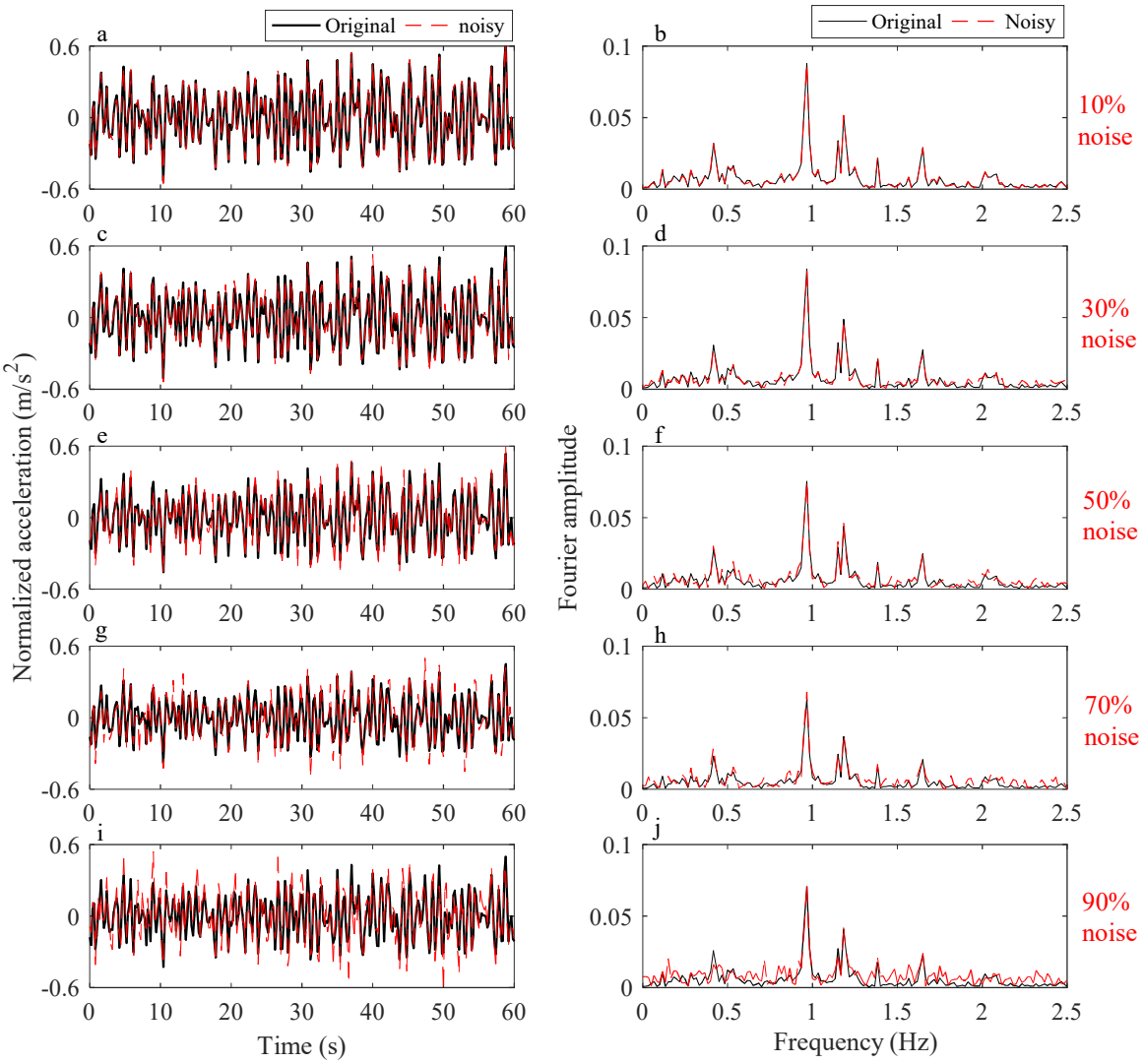


Figure 4. Original and noisy signals with different noise levels. In the time domain: (a) 10%, (c) 30%, (e) 50%, (g) 70% and (i) 90% and in the frequency domain: (b) 10%, (d) 30%, (f) 50%, (h) 70%, (j) 90%.

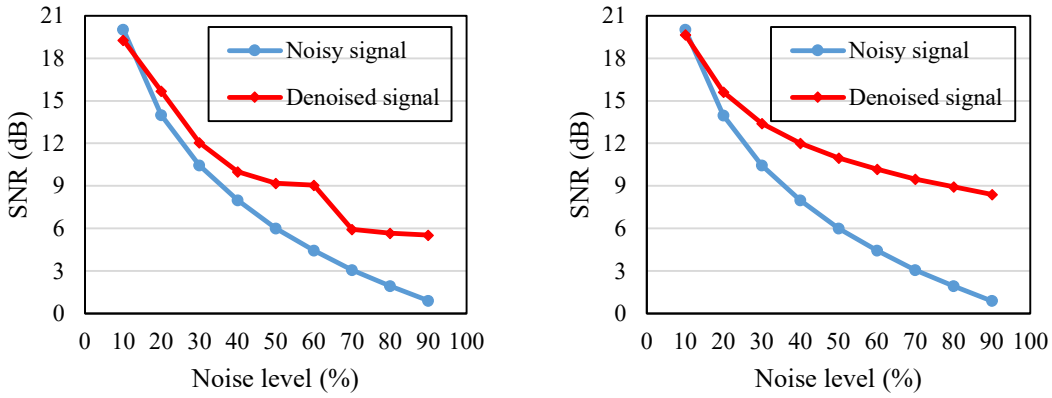
3.3 Results from testing data with white noise

The computer used for training these networks has a GTX1060 GPU, an i7-6700K CPU and 16GB memory. With the abundant datasets, the network converges very fast. The size of the mini batch is 64 and the whole training datasets are repetitively used to turn the trainable parameters of the network 10 times. Two ResNets are trained individually and the training time for each ResNet is around 3 hours. The testing datasets are then used to validate the trained networks. It takes around 4

seconds for each testing sample to produce the denoised signal. It should be mentioned that a testing sample is one-hour long acceleration data from ten channels, however, only a short processing time is needed.

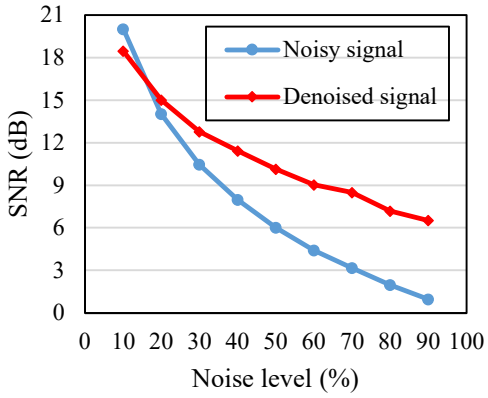
3.3.1 Evaluation of results in time and frequency domains

SNRs of the input noisy and output generated denoised signals are computed based on Equation (3). The comparison of SNRs for the noisy and denoised signals of each group in the short-axis and long-axis directions are shown in Figure 5. With the increasing level of noise, the SNR of the noisy signals decreases, which confirms that the noise has been correctly injected into the original signal. The SNRs of the denoised signals reduce with the increase of noise level monotonously, but are larger than those of the noisy signals for most noise levels of the testing data. The SNR of the denoised signals is up to 8 times better than the corresponding noisy ones. These results indicate that noise components are effectively eliminated from the noisy signals and the denoised signals become closer to the corresponding original signals. The effectiveness of the proposed approach is also validated by evaluating the signals in the frequency domain using Equation (4). The results as shown in Figure 6 demonstrate that the errors significantly decrease by using the proposed approach for the signal denoising, especially when the noise level is high. The errors in the frequency domain reduce more than 30% when the noise level reaches 90% for Testing data 1 in both directions and Testing data 2 in the short-direction as shown in Figures 6(a) – (c).

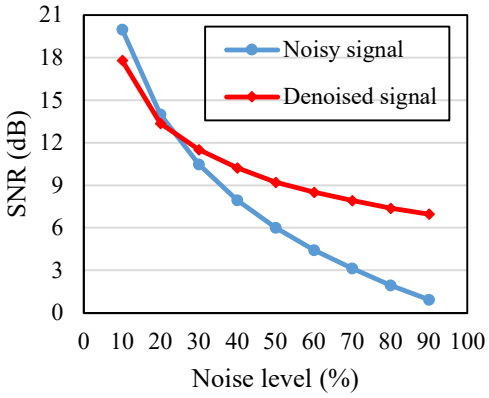


(a)

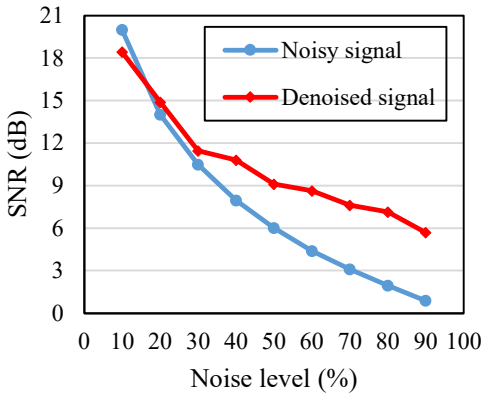
(b)



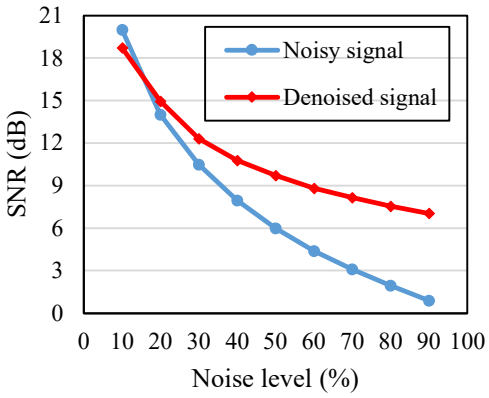
(c)



(d)



(e)



(f)

Figure 5. Comparison of SNRs of noisy and denoised signals. Results for short-axis direction: (a) Testing data 1; (c) Testing data 2 and (e) Testing data 3. Long-axis direction: (b) Testing data 1; (d) Testing data 2 and (f) Testing data 3.

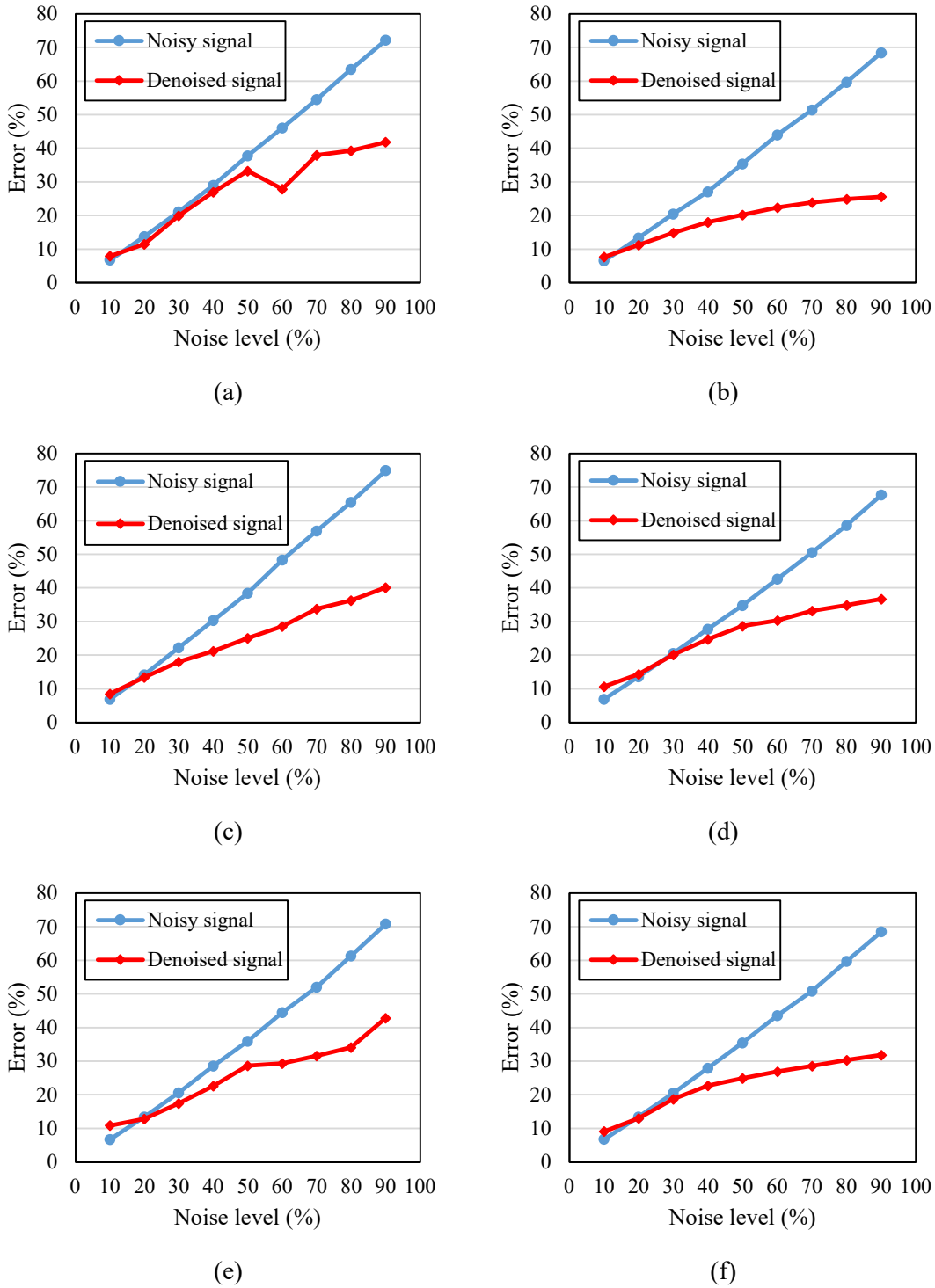


Figure 6. Comparison of errors of noisy and denoised signals in the frequency domain. Results for short-axis direction: (a) Testing data 1; (c) Testing data 2 and (e) Testing data 3. Long-axis direction: (b) Testing data 1; (d) Testing data 2 and (f) Testing data 3.

The trained ResNet remarkably enhances the quality of signals in both the time and frequency

domains. However, denoising of the signals with 10% noise is an exception where the SNRs decrease and the spectral errors increase by the denoising process. The measured acceleration data from channel 18 of Testing data 1 in the long direction is selected to explain this phenomenon. The pre-processed original signal of the selected data is used as input to the trained ResNet. PSD spectra of the input original and the output denoised signals, as well as the PSD errors, are shown in Figure 7. The PSD errors are the discrepancy of amplitudes for each frequency interval. It can be observed that the structural frequency components of the denoised signal remain highly consistent with those of the original signal, while the power of noise components in non-structural frequency bandwidths are decreased. That means the proposed method can not only reduce the manually injected noise but also mitigate the measurement noise.

Therefore, the potential reason for the decrease of SNR and the increase of frequency spectra error when denoising noisy signals with 10% injected noise can be attributed to that the proposed approach reduces the artificially injected noise as well as the measurement noise existed in the original signal at the same time. The reduction of the noise in the original signal, which is used as the bench mark for comparison, enlarges the discrepancy between the original and the denoised signals and causes a decrease of SNR when the injected noise is less significant. It has to be mentioned that the purpose of this study is not to recover but denoise the original signal, reduction of the noise in the original signal in practice improves the quality. These results indicate that the proposed ResNet can accurately extract the important features from the training datasets, which are the vibrational characteristics of the Guangzhou New TV Tower. The finding corresponds well with that in a previous study [26] where the stacked convolutional layers can automatically extract the features of vibration data, which are very similar to the natural frequencies and mode shapes of the structure. Therefore, the proposed approach operates as an auto-adaptive filter to denoise signals regardless of the noise level and source.

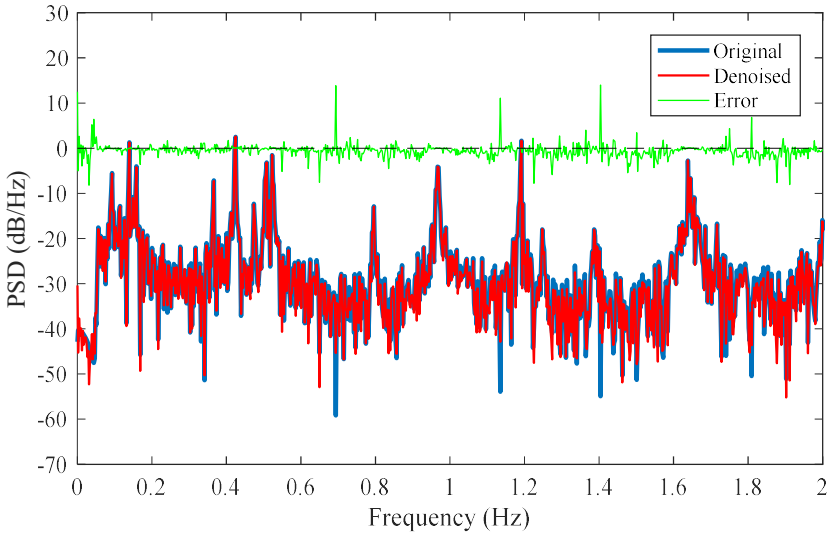


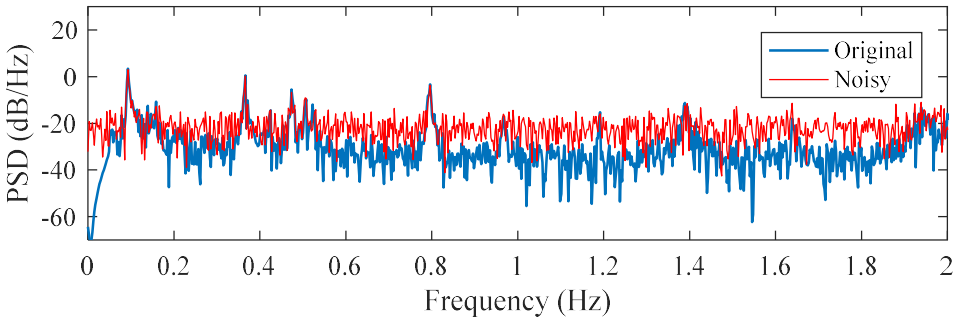
Figure 7. PSD spectra of the pre-processed original and denoised signals, and the errors in PSD.

3.3.2 Comparison of results with wavelet transform based method

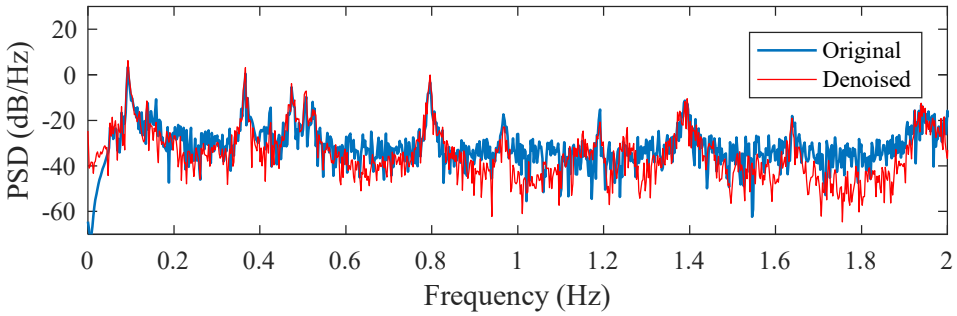
Wavelet transform based method is one of the most popular and most researched methods for vibration signal denoising in the current literature, therefore it is selected for comparing the performance with the proposed method. Wavelet transform decomposes signals into different scales by using a selected mother wavelet function. It concentrates features of signals in a few scale components with large wavelet coefficients. Scale components with small wavelet coefficients that are lower than a defined threshold are categorized as noise and would be removed. With the defined threshold of the coefficients, a signal is then reconstructed as a denoised one using the remained scale components by conducting inverse wavelet transform. Parameters including the mother wavelet function, level of wavelet decomposition and threshold are carefully chosen. In this study, for denoising measured acceleration data of Guangzhou New TV Tower, Symlet wavelet with 8 vanishing moments (Sym8) is used as the mother wavelet function. The level of decomposition is tested from 1 to $\log_2 N$ (14), where N is the total number of sampling points. By comparing SNR value of denoised signals with different levels of decomposition, level 1 decomposition is finally applied. The threshold is defined by using an empirical Bayesian method with a Cauchy prior and a posterior soft threshold rule [52]. PSD spectra of the original signal, noisy signal with 90% noise and denoised signals by using the trained ResNet and wavelet transform based method are also computed and compared in

Figure 8. Figures 8(a) and (d) show that the injected noise seriously contaminates the original signal. More than half of natural frequencies with relatively low power existing in the considered frequency bandwidths are masked by the intensive noise. In this case, natural frequencies are difficult to be identified from the noisy signals without denoising. In contrary, PSDs of the denoised signals by the trained ResNet as shown in Figures 8(b) and (e) have much clearer spectra, where noises existed in the non-structural frequency bandwidths are effectively eliminated, leading to a fact that natural frequencies can be identified accurately. Comparing PSDs of the noisy and denoised signals, the noise level is effectively reduced and there is a very limited influence on using the power spectrum for natural frequency identification based on the denoised signals. The proposed approach automatically learns the vibration features of the structure through training. Recalling the results demonstrated in Figure 6, the errors in the frequency domain between the original and denoised signals is evaluated by considering the entire frequency bandwidth. The performance of the proposed approach is in fact even better if only the most useful information, i.e. the frequency intervals around natural frequencies are considered.

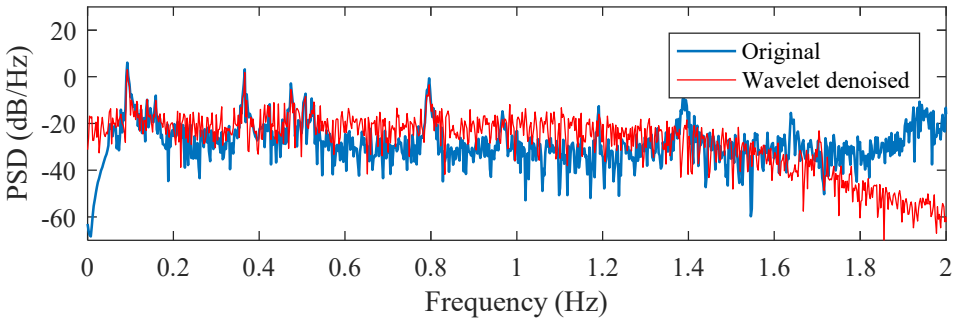
By observing the PSDs of the denoised signals by using the trained ResNet and wavelet transform based method as shown in Figures 8(c), (d), (e) and (f), it is obvious that the proposed approach outperforms the wavelet transform based method on denoising vibration signals. Wavelet transform has a minor improvement in signal quality on denoising the injected and measurement noises. It can be explained that with heavy noise contamination in signals, the wavelet transform based method becomes ineffective to localize the vibration features. Only the information in the selected scales is stored and recovered, therefore it could be difficult to eliminate noise by simply defining a threshold for the wavelet transform based method for signal denoising.


PSD (dB/Hz)
Frequency (Hz)
Original
Noisy

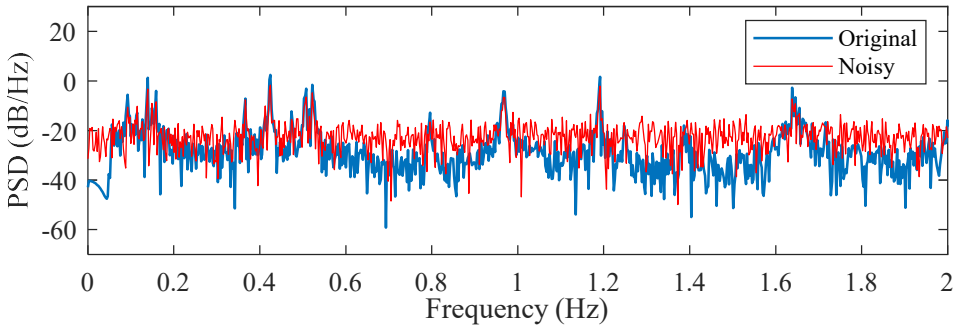
(a)


PSD (dB/Hz)
Frequency (Hz)
Original
Denoised

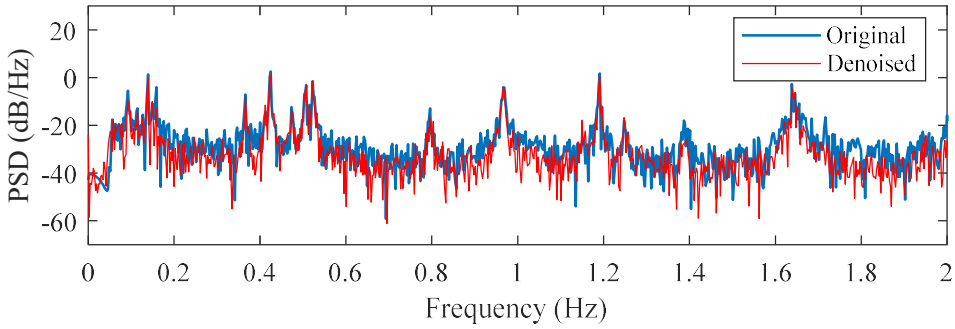
(b)


PSD (dB/Hz)
Frequency (Hz)
Original
Wavelet denoised

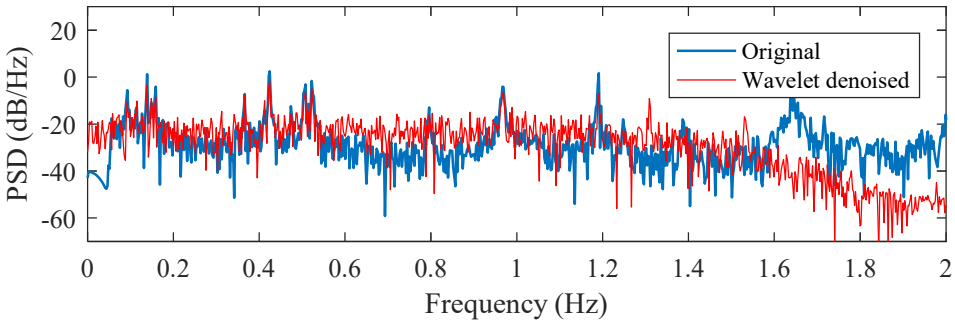
(c)


PSD (dB/Hz)
Frequency (Hz)
Original
Noisy

(d)



(e)



(f)

Figure 8. Comparison of PSD spectra of signals in Testing data 1, Short-axis direction: (a) Original and noisy signals with 90% noise; (b) Original and ResNet denoised signals; (c) Original and wavelet denoised signals. Long-axis direction: (c) Original and noisy signals with 90% noise; (d) Original and ResNet denoised signals. (f) Original and wavelet denoised signals.

3.4 Modal analysis of the testing results using denoised signals

3.4.1 Quantitative evaluation of the denoised signals

FDD is a robust modal analysis method with strong noise immunity especially to white noise. It is noted that the quality of selected testing vibration signals from Guangzhou New TV Tower is high. Only a few modes are buried by 90% noise as shown in the above studies. To further investigate the effectiveness of the approach and the usability of denoised signals, special testing data are generated by using the original signal of Testing data 1. The selected signal is contaminated by a very severe noise with a 500% noise level. Figures 9 and 10 show the identification results of original, noisy and denoised signals in the short and long-axis directions using FDD. Figures 9(a) and 10(a) demonstrate the high quality of the original signals, where all the 15 physical modes can be identified and few

spurious modes are observed. Figures 9(b) and 10(b) show that under such strong noise, the usability of signals is significantly affected with most of the modes submerged by the noise except several strongly excited modes as marked with green dash arrows. In addition, the noise effect at some frequencies is very pronounced as marked with the red arrows in Figures 9(b) and 10(b), which may confuse the engineers' judgement and lead to false identification results. Noisy signals with such a low quality are potentially discarded if no efficient denoising process is conducted. To mitigate the noise effect, the noisy signals are used as input to the corresponding trained ResNets for denoising. Modal analysis of the output denoised signals is conducted. Figures 9(c) and 10(c) present the singular value spectra from FDD analysis with the denoised signals in both directions, respectively. The excellent performance of the proposed approach is obvious on denoising signals contaminated by such severe noises. All those 15 natural frequencies of the structure can be clearly identified from the denoised signals. Both the manually injected and the measurement noise with frequencies outside the natural frequency bandwidths are effectively eliminated. It should be mentioned that these two ResNets are trained with noisy signals containing a maximum of 90% noise. The results demonstrate that the proposed approach can denoise signals with much more severe noise than the training samples. Comparing Figures 9(b) and (c) and Figures 10(b) and (c), the singular value spectra of the denoised signals are much clearer than those of noisy signals, indicating that the quality of the signal is enhanced significantly. It should be highlighted that the noise within five very closely spaced modes between 0.36 Hz and 0.56 Hz is successfully cleaned without affecting the adjacent structural vibration modes. Moreover, these five modes in the denoised signals are accurately separated without any false identification, which is difficult for the traditional methods to define the bandwidths of filters if no prior information of those four modes is provided. The previous findings confirm that ResNets can automatically extract the most important features of the vibrational signals from the noisy signals and mapping the noisy input to a cleaner one. The results demonstrate that the proposed approach is effective and reliable for vibration signal denoising.

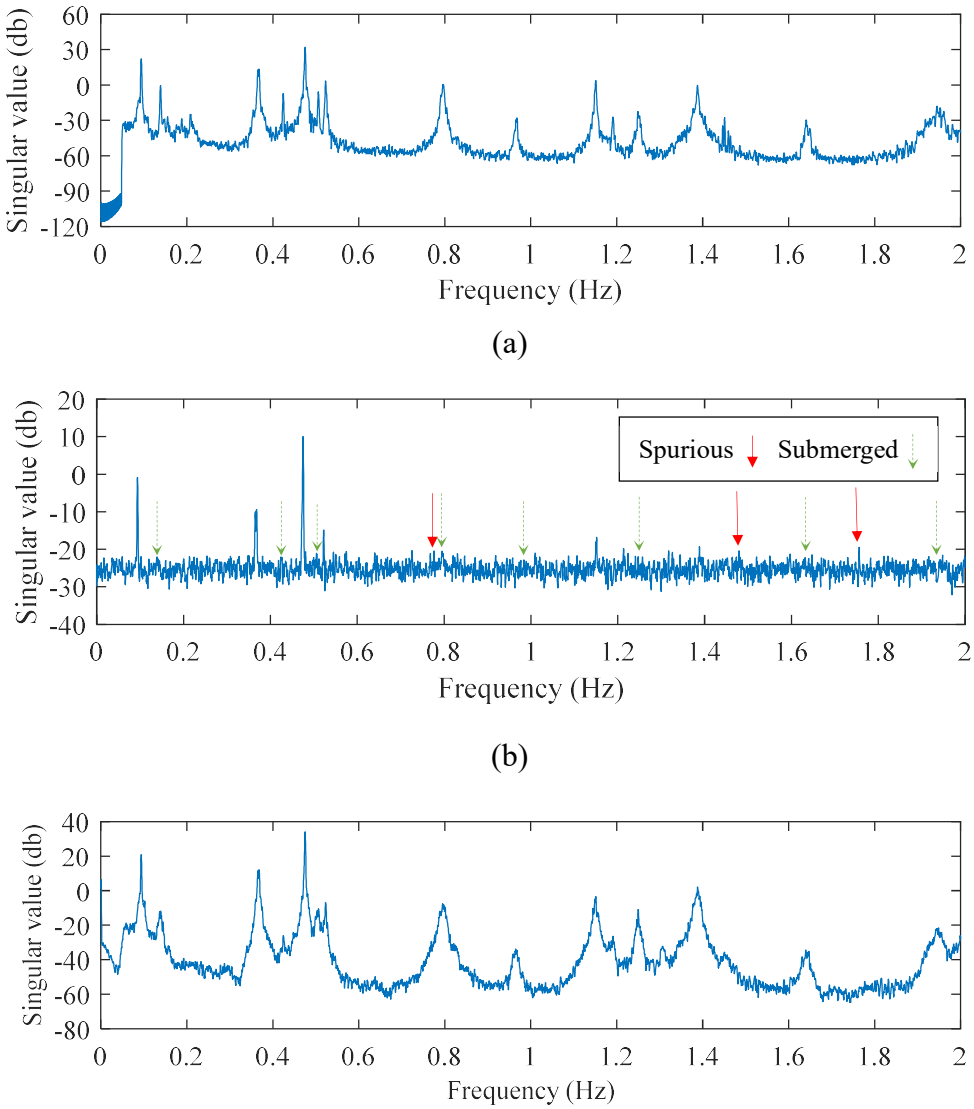
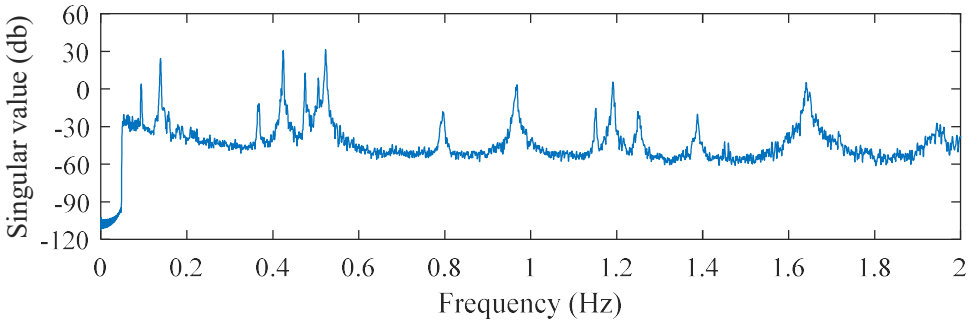
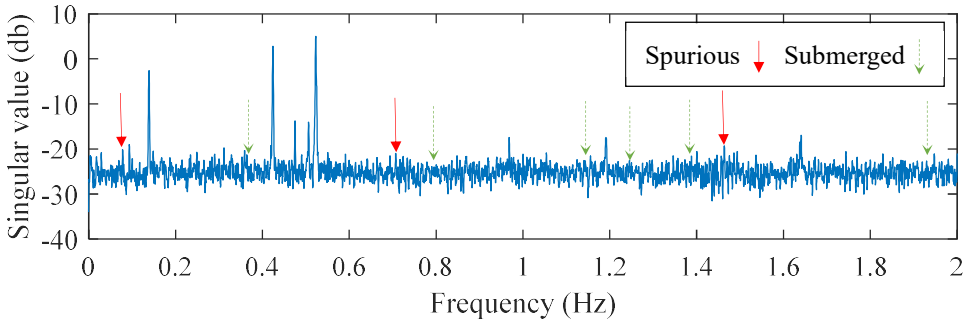


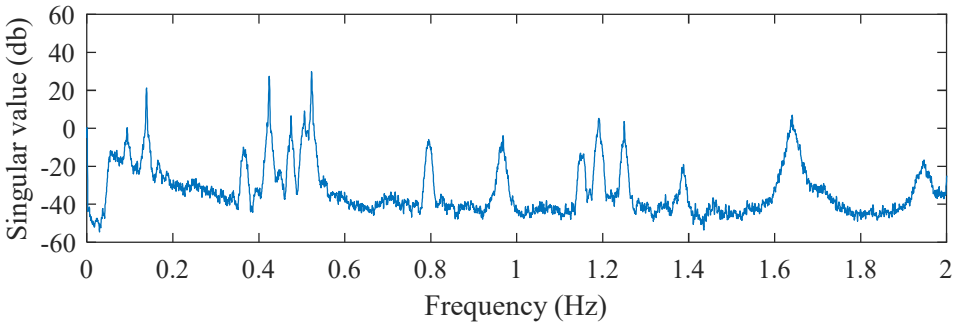
Figure 9. Singular value spectra of: (a) original; (b) noisy; and (c) denoised signals of Testing data 1 in the short-axis direction.



(a)



(b)



(c)

Figure 10. Singular value spectra of: (a) original; (b) noisy; and (c) denoised signals of Testing data 1 in the long-axis direction.

3.4.2 Evaluation of identified modal parameters

To evaluate the denoising accuracy, the identified natural frequencies and damping ratios of vibration modes within 2 Hz from original, noisy and denoised signals of Testing data 1 are listed in Tables 2 and 3 for short and long-axis directions, respectively. Modal identification results using the

original signals show good consistency with those in Ref. [50]. Regardless of the spurious modes and submerged weakly excited modes, most of the identified natural frequencies from the noisy signal are very similar to those from the original signals as presented in Tables 2 and 3. Large frequency errors of the 11th mode in the short-axis direction and the 13th mode in the long-axis direction are observed and classified as false identifications, where the spectrum of the noise is stronger than the adjacent natural frequencies. In contrast, all the identified natural frequencies from the denoised signal are very accurate. Minor differences can be found in the weakly excited modes such as the 2nd, 4th and 9th modes in the short-axis direction. Five closely spaced modes i.e. the 3rd to 7th modes of the structure are well separated and the identification of these five modes is highly accurate in both directions. These results indicate that the used ResNet has mined and memorized different features for these five modes consisted in the vibration signals without any manual intervention. In addition, by comparing the errors in the identified natural frequencies of the noisy and denoised signals, it can be found that the vibration signal denoising process slightly improves the identification accuracy of those strongly excited modes as well.

As shown in Tables 2 and 3, the identification of the damping ratios is not as accurate as the natural frequencies. The injected serious noise leads to estimation of damping ratios very tough. As a result, the identified damping ratios from the noisy and denoised signals have relatively large differences compared with those obtained from the original signals. No clear relationship can be found between the estimation accuracy of damping ratios and the denoising process. However, with the denoised signals, the damping ratios of all those 15 modes can be successfully identified. It should be noted that the damping ratios can only be identified for a limited number of modes if noisy signals are used.

Table 2. Identified natural frequencies and damping ratios from original, noisy and denoised signals in the short-axis.

mode	**Original signal**		**Noisy signal**				**Denoised signal**			
	f (Hz)	ξ (%)	f (Hz)	Error (%)	ξ (%)	Error (%)	f (Hz)	Error (%)	ξ (%)	Error (%)
1	0.094	0.62	0.094	0.00	0.33	0.29	0.094	0.00	0.40	0.22
2	0.139	0.35	-	-	-	-	0.139	0.32	0.78	0.43
3	0.367	0.16	0.365	0.66	0.54	0.38	0.367	0.12	0.20	0.04
4	0.424	0.11	-	-	-	-	0.425	0.29	0.13	0.02
5	0.475	0.22	0.475	0.00	0.23	0.01	0.475	0.00	0.36	0.13
6	0.505	0.12	-	-	-	-	0.507	0.24	0.49	0.37
7	0.522	0.09	0.522	0.00	0.21	0.12	0.522	0.00	0.15	0.06
8	0.796	0.19	-	-	-	-	0.796	0.00	0.51	0.32
9	0.968	0.25	-	-	-	-	0.967	0.13	0.36	0.11
10	1.151	0.11	1.151	0.00	0.15	0.04	1.151	0.00	0.11	0.00
11	1.191	0.07	1.213	1.84	1.67	1.60	1.191	0.00	0.12	0.05
12	1.249	0.16	-	-	-	-	1.250	0.10	0.13	0.03
13	1.388	0.28	1.388	0.00	0.48	0.20	1.388	0.00	0.31	0.03
14	1.641	0.30	-	-	-	-	1.639	0.07	0.23	0.07
15	1.945	0.50	-	-	-	-	1.948	0.19	0.55	0.05

Note: f: frequency; ξ: damping ratio.

Table 3. Identified natural frequencies and damping ratios from original, noisy and denoised signals in the long-axis.

mode	**Original signal**		**Noisy signal**				**Denoised signal**			
	f (Hz)	ξ (%)	f (Hz)	Error (%)	ξ (%)	Error (%)	f (Hz)	Error (%)	ξ (%)	Error (%)
1	0.094	2.12	0.094	0.09	1.55	0.57	0.094	0.00	1.72	0.40
2	0.139	0.66	0.139	0.00	0.76	0.10	0.139	0.00	0.49	0.18
3	0.367	0.23	-	-	-	-	0.368	0.12	0.40	0.16
4	0.424	0.07	0.424	0.00	0.03	0.03	0.424	0.00	0.02	0.04
5	0.475	0.21	0.475	0.00	0.51	0.30	0.475	0.00	0.68	0.47
6	0.505	0.16	0.507	0.24	1.36	1.20	0.506	0.04	0.12	0.03
7	0.522	0.10	0.522	0.00	0.17	0.07	0.522	0.00	0.12	0.02
8	0.796	0.28	-	-	-	-	0.795	0.15	0.32	0.04
9	0.968	0.24	0.968	0.00	0.71	0.47	0.968	0.00	0.22	0.02
10	1.151	0.13	-	-	-	-	1.151	0.00	0.14	0.01
11	1.191	0.05	1.191	0.00	0.05	0.00	1.191	0.00	0.28	0.22
12	1.249	0.18	-	-	-	-	1.250	0.10	0.73	0.55
13	1.388	0.30	1.464	5.45	2.08	1.78	1.388	0.00	0.50	0.20
14	1.641	0.22	1.641	0.00	0.33	0.11	1.641	0.00	0.62	0.40
15	1.945	0.60	-	-	-	-	1.947	0.13	1.20	0.60

Note: f: frequency; ξ: damping ratio.

To evaluate the accuracy of the identified mode shapes, for those true mode shapes of the structure, MAC values between the mode shapes from the original signal and those from the noisy signals with 90% and 500% noises, and MAC values between the mode shapes from the original signals and those from the denoised signals, for short and long-axis directions, are shown in Figures 11 and 12, respectively. It can be observed that the proposed approach can assist in extracting the modes buried in noise effectively. A relatively large discrepancy can be observed between the original and the ones identified from the noisy and denoised signals when the noise level reaches 500%. The identified mode shapes of the noisy and denoised signals are similar for the strongly excited modes. It is noted that the injected significant noise is random for each channel. The inaccuracy in mode shapes can be attributed to that for each channel, the signal components with the same frequencies as the structural natural frequencies are affected by the noise effect in varying extents. When denoising the noisy signal, the proposed approach is mainly concentrated on denoising the noise outside the

natural frequency bandwidths. To the best of the authors' knowledge, to distinguish and eliminate the measurement noise on structural vibration frequencies is a very difficult problem for all the state-of-the-art denoising methods to address. In practice, 500% noise can rarely happen during the field measurements. As can be seen in Figures 11 and 12, mode shapes identified from the noisy signals with 90% noise and the corresponding denoised signals are very accurate, where the MAC values of most of the vibration modes are higher than 0.99. This is because white noise has limited influences on the mode shapes, therefore even 90% white noise is injected into the signal, the mode shape can still be well extracted. These results demonstrate that the vibrational characteristics, such as frequencies and mode shapes, can be preserved after vibration signal denoising by the proposed approach.

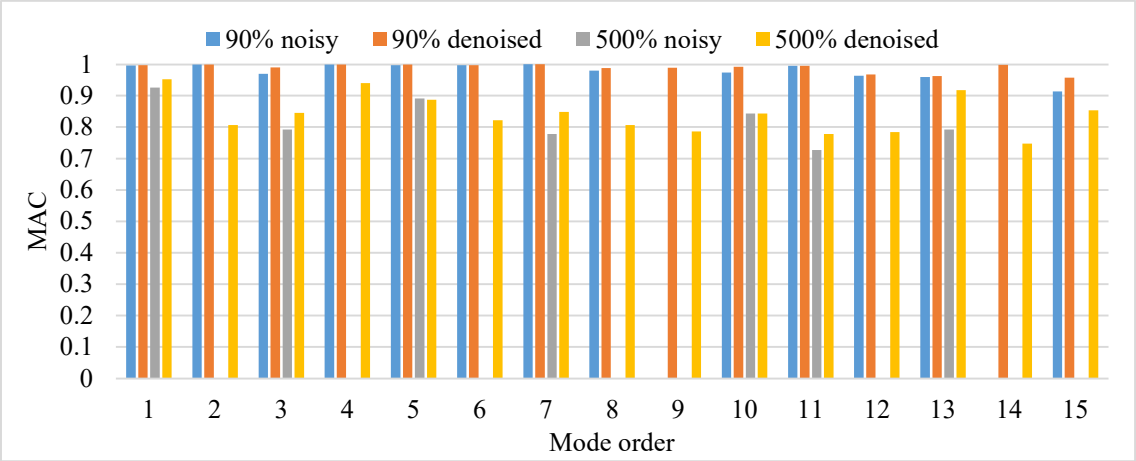


Figure 11. MAC values of the identified mode shapes in the short-axis direction.

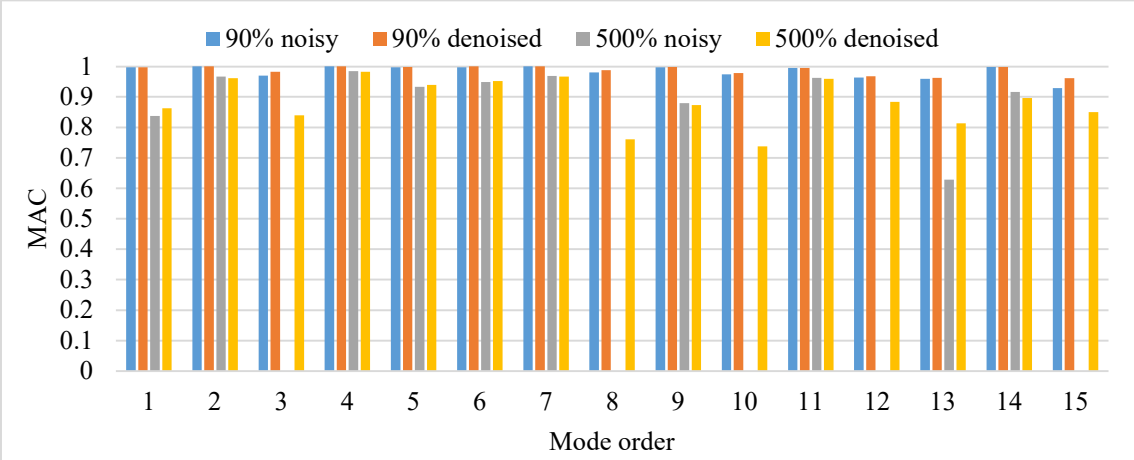


Figure 12. MAC values of the identified mode shapes in the long-axis direction.

4. Discussion on the Effect of Pink Noise

In the above studies, the performance of the proposed method is proved to be effective to denoise signals with serious white noises. However, field measurements could be contaminated by different types of noises. A signal denoising method that is only capable of eliminating one kind of noise is impractical for SHM. To further evaluate the applicability of the proposed method in practice, a group of testing data contained with increasing levels of pink noise is generated using the same original signal as Testing data 1. Different from white noise with energy intensity equally distributed in the whole frequency bandwidths, PSD of pink noise is inversely proportional to the frequency f which can be expressed as,

$$PSD_{pink\ noise} \propto \frac{1}{f} \tag{7}$$

The means to generate a noisy signal with pink noise is similar to that of white noise using Equation (5). The difference is that *Noise* vector becomes a time sequence of pink noise with zero mean and standard deviation of the original signal.

Figure 13 shows the denoised results of both short and long-axis directions in time and frequency domains. The increase of SNR and the decrease in the errors of the frequency spectrum of denoised signals indicate that the injected pink noise is effectively mitigated by the proposed approach. The performance of denoising pink noise is close to that for white noise, which demonstrates the robustness of the proposed method. The error in the frequency domain is higher than that of the white noise cases, because the original signal is high-pass filtered by 0.05Hz while the pink noise has the strongest power in this frequency interval. This notable discrepancy in this non-target frequency range causes a large error, as shown in Figure 14. From this figure, it can also be observed that the quality of PSD spectra of the denoised signals is much higher than that of the noisy signal. The merged natural frequencies by the pink noise can be clearly identified and the spurious modes are much fewer after denoising. The results of denoising pink noise in vibration signals show that the proposed approach is robust for eliminating different types of noise, indicating the feasibility and broad application range

of the proposed method for denoising vibration signals.

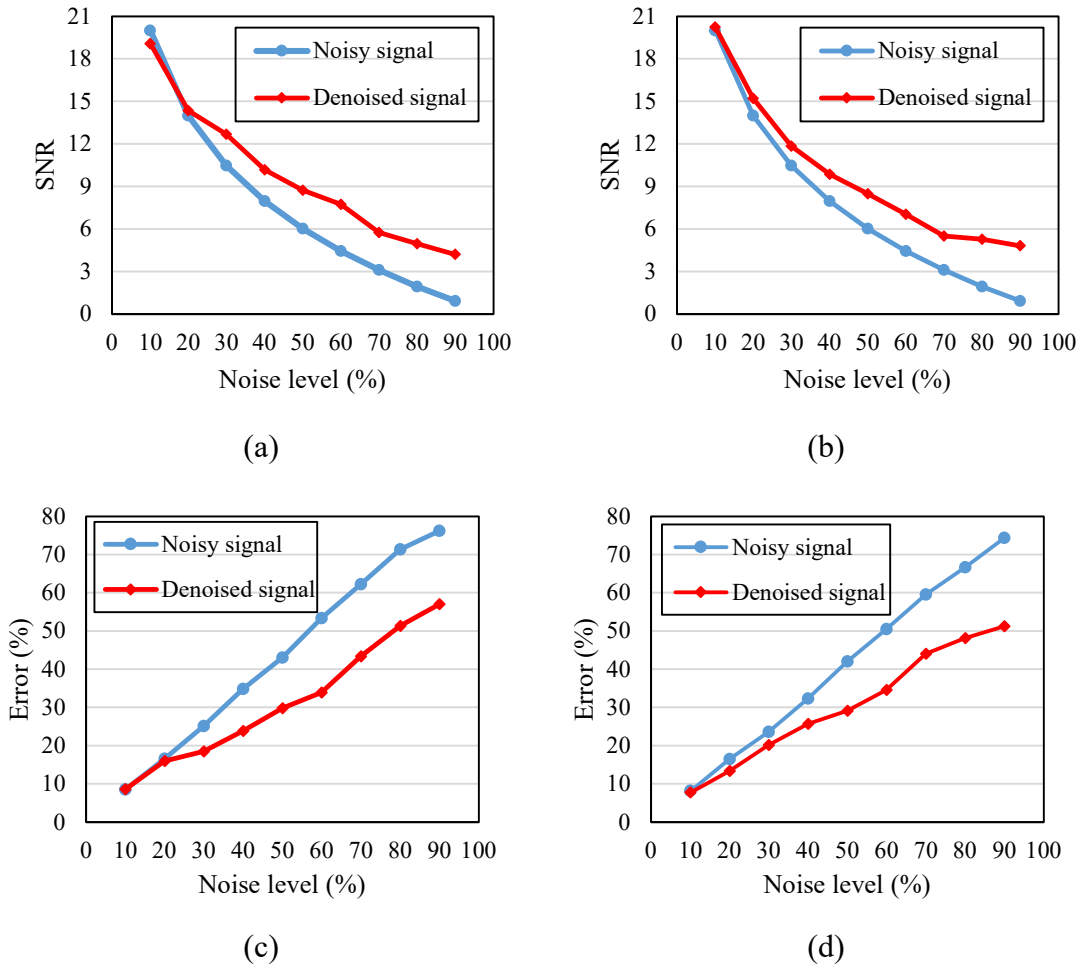
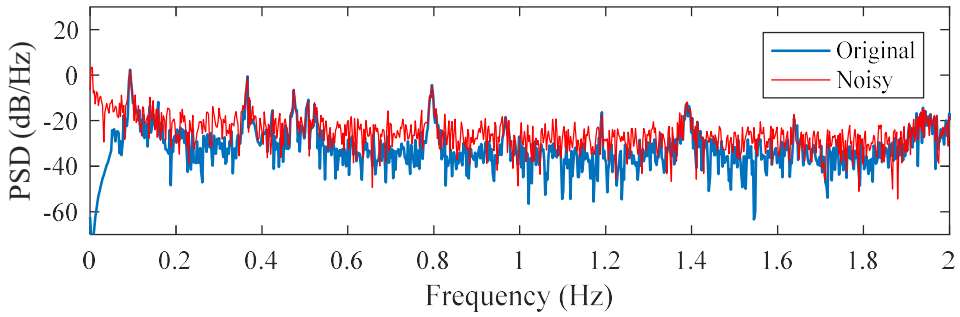
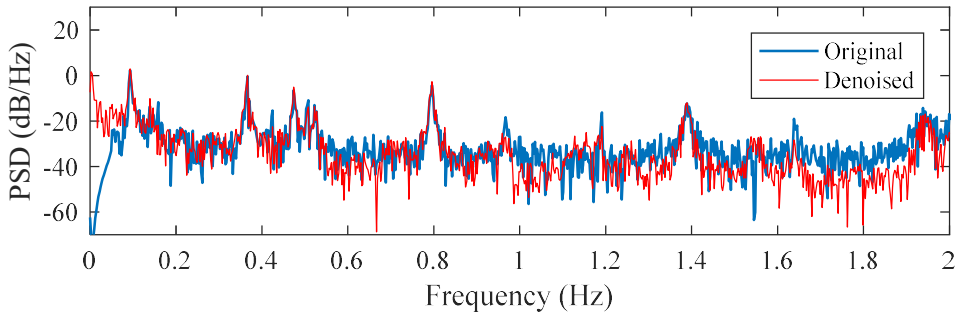


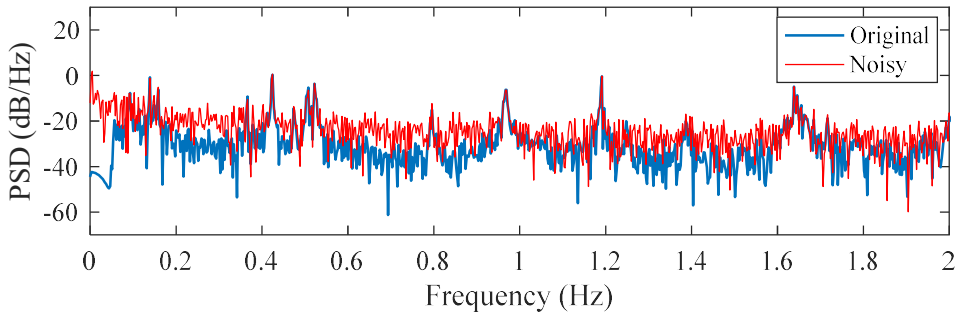
Figure 13. Comparison of errors of the noisy and denoised signals in the time and frequency domains. Results for short-axis direction in: (a) Time domain and (c) Frequency domain; Long-axis direction in: (b) Time domain and (d) Frequency domain.



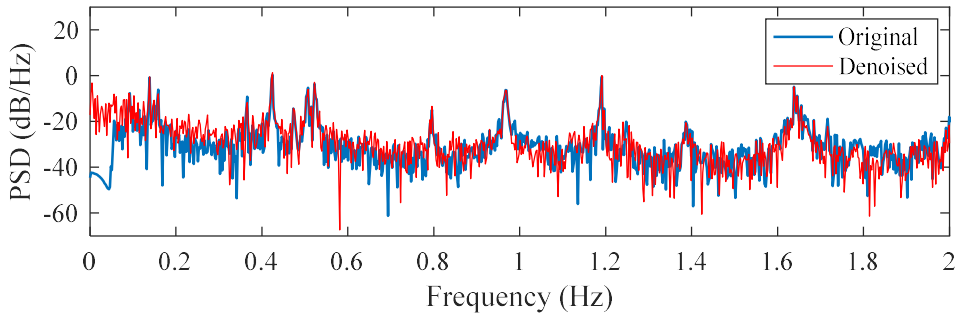
(a)



(b)



(c)



(d)

Figure 14. Comparison of PSD spectra for original, noisy and denoised signals. Short-axis direction: (a) Original signal and noisy signal with 90% noise; (b) Original and denoised signals; Long-axis direction: (c) Original signal and noisy signal with 90% noise; (d) Original and denoised signals.

5. Conclusions

This paper proposes a ResNet based vibration signal denoising approach to remove noise in vibration measurement data. The details of the sophisticatedly designed ResNet and involved techniques to improve the capability of the network are elaborated. The effectiveness and robustness of the developed approach are demonstrated by using recorded data on Guangzhou New TV Tower. Testing results demonstrate that the proposed approach effectively improves the quality of noisy signals injected with varying levels and types of noises. Modal identification of noisy and denoised signals shows that natural frequencies of the weakly excited modes submerged by the noise effect and the closely spaced modes can be clearly identified after denoising. These results indicate that the used networks extract the most important features from the training datasets, namely vibrational characteristics of the structure. The results also demonstrate that the proposed approach has a strong robustness to eliminate another type of noise, i.e. pink noise, although it is not included in training datasets. On the other hand, the networks learn the features such as natural frequencies of the structure automatically and can help to distinguish physical modes from spurious modes to a certain degree. The improvement in the identification of damping ratios and mode shapes is limited. Future studies can be conducted to address the limitation by using improved deep learning networks and objective functions.

Acknowledgement

The work described in this paper was supported by Australian Research Council Laureate Fellowships FL180100196.

References

[1] E. Cross, K. Koo, J. Brownjohn, K. Worden, Long-term monitoring and data analysis of the Tamar Bridge, Mechanical Systems and Signal Processing, 35 (2013) 16-34.
[2] Y. Xia, B. Chen, X. Zhou, Y. Xu, Field monitoring and numerical analysis of Tsing Ma Suspension Bridge temperature behavior, Structural Control and Health Monitoring, 20 (2013) 560-575.
[3] T. Nguyen, T.H. Chan, D.P. Thambiratnam, L. King, Development of a cost-effective and flexible vibration DAQ system for long-term continuous structural health monitoring, Mechanical Systems and Signal Processing, 64 (2015) 313-324.
[4] Y.Q. Ni, Y. Xia, W.Y. Liao, J.M. Ko, Technology innovation in developing the structural health monitoring system for Guangzhou New TV Tower, Structural Control and Health Monitoring, 16 (2009) 73-98.
[5] J. Li, S. Law, Y. Ding, Substructure damage identification based on response reconstruction in frequency domain and model updating, Engineering Structures, 41 (2012) 270-284.
[6] J. Li, S. Law, Damage identification of a target substructure with moving load excitation, Mechanical Systems and Signal Processing, 30 (2012) 78-90.
[7] Q. He, X. Wang, Q. Zhou, Vibration sensor data denoising using a time-frequency manifold for machinery fault diagnosis, Sensors, 14 (2013) 382-402.
[8] S. Braun, The synchronous (time domain) average revisited, Mechanical Systems and Signal Processing, 25 (2011) 1087-1102.
[9] J. Lin, L. Qu, Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis, Journal of sound and vibration, 234 (2000) 135-148.
[10] F. Bi, T. Ma, X. Wang, Development of a novel knock characteristic detection method for gasoline engines based on wavelet-denoising and EMD decomposition, Mechanical Systems and Signal Processing, 117 (2019) 517-536.
[11] S.N. Chegini, A. Bagheri, F. Najafi, Application of a New EWT-Based Denoising Technique in Bearing Fault Diagnosis, Measurement, 144 (2019) 275-297.
[12] M. Zhao, X. Jia, A novel strategy for signal denoising using reweighted SVD and its applications to weak fault feature enhancement of rotating machinery, Mechanical Systems and Signal Processing, 94 (2017) 129-147.
[13] Y. Lei, J. Lin, Z. He, M.J. Zuo, A review on empirical mode decomposition in fault diagnosis of rotating machinery, Mechanical Systems and Signal Processing, 35 (2013) 108-126.
[14] S. Hou, M. Liang, Y. Li, An optimal global projection denoising algorithm and its application to shaft orbit purification, Structural Health Monitoring, 10 (2011) 603-616.
[15] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion, Journal of Machine Learning Research, 11 (2010) 3371-3408.
[16] A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems, Lake Tahoe, USA, 2012, pp. 1097-1105.
[17] V. Jain, S. Seung, Natural image denoising with convolutional networks, Advances in Neural Information Processing Systems, Vancouver, Canada, 2009, pp. 769-776.

[18] J. Xie, L. Xu, E. Chen, Image denoising and inpainting with deep neural networks, Advances in Neural Information Processing Systems, Lake Tahoe, USA, 2012, pp. 341-349.
[19] B. Xia, C. Bao, Speech enhancement with weighted denoising auto-encoder, Interspeech, 2013, pp. 3444-3448.
[20] S.R. Park, J. Lee, A fully convolutional neural network for speech enhancement, https://arxiv.org/abs/1609.07132 (2016).
[21] S. Poungponsri, X.-H. Yu, An adaptive filtering approach for electrocardiogram (ECG) signal noise reduction using neural networks, Neurocomputing, 117 (2013) 206-213.
[22] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, Deep learning and its applications to machine health monitoring, Mechanical Systems and Signal Processing, 115 (2019) 213-237.
[23] Y.J. Cha, W. Choi, O. Büyüköztürk, Deep learning-based crack damage detection using convolutional neural networks, Computer-Aided Civil and Infrastructure Engineering, 32 (2017) 361-378.
[24] K. Jang, N. Kim, Y.-K. An, Deep learning–based autonomous concrete crack evaluation through hybrid image scanning, Structural Health Monitoring, 18 (2019) 1722-1737.
[25] Y. Xu, S. Wei, Y. Bao, H. Li, Automatic seismic damage identification of reinforced concrete columns from images by a region-based deep convolutional neural network, Structural Control and Health Monitoring, 26 (2019) e2313.
[26] Y.Z. Lin, Z.H. Nie, H.W. Ma, Structural damage detection with automatic feature-extraction through deep learning, Computer-Aided Civil and Infrastructure Engineering, 32 (2017) 1025-1046.
[27] O. Abdeljaber, O. Avci, S. Kiranyaz, M. Gabbouj, D.J. Inman, Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks, Journal of Sound and Vibration, 388 (2017) 154-170.
[28] G. San Martin, E. López Droguett, V. Meruane, M. das Chagas Moura, Deep variational auto-encoders: a promising tool for dimensionality reduction and ball bearing elements fault diagnosis, Structural Health Monitoring, 18 (2018) 1092-1128.
[29] X. Guo, L. Chen, C. Shen, Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis, Measurement, 93 (2016) 490-502.
[30] L. Jing, M. Zhao, P. Li, X. Xu, A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox, Measurement, 111 (2017) 1-10.
[31] Y. Zhang, Y. Miyamori, S. Mikami, T. Saito, Vibration-based structural state identification by a 1-dimensional convolutional neural network, Computer-Aided Civil and Infrastructure Engineering, 34 (2019) 822-839.
[32] C.S.N. Pathirage, J. Li, L. Li, H. Hao, W. Liu, R. Wang, Development and application of a deep learning–based sparse autoencoder framework for structural damage identification, Structural Health Monitoring, 18 (2018) 103-122.
[33] C.S.N. Pathirage, J. Li, L. Li, H. Hao, W. Liu, P. Ni, Structural damage identification based on autoencoder neural networks and deep learning, Engineering Structures, 172 (2018) 13-28.
[34] Y. Bao, Z. Tang, H. Li, Y. Zhang, Computer vision and deep learning–based data anomaly detection method for structural health monitoring, Structural Health Monitoring, 18 (2019) 401-421.
[35] G. Fan, J. Li, H. Hao, Lost data recovery for structural health monitoring based on convolutional

neural networks, Structural Control and Health Monitoring, 26 (2019) e2433.
[36] V. Kuleshov, S.Z. Enam, S. Ermon, Audio super resolution using neural networks, https://arxiv.org/abs/1708.00853 (2017).
[37] G.E. Hinton, R.R. Salakhutdinov, Reducing the dimensionality of data with neural networks, science, 313 (2006) 504-507.
[38] V. Badrinarayanan, A. Kendall, R. Cipolla, Segnet: a deep convolutional encoder-decoder architecture for image segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence, 39 (2017) 2481-2495.
[39] P. Isola, J.-Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, Proceedings of the IEEE conference on computer vision and pattern recognition, IEEE, Honolulu, USA, 2017, pp. 1125-1134.
[40] X. Mao, C. Shen, Y.B. Yang, Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections, Advances in neural information processing systems, Barcelona, Spain, 2016, pp. 2802-2810.
[41] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, The Journal of Machine Learning Research, 15 (2014) 1929-1958.
[42] Y. Xu, J. Du, L.-R. Dai, C.-H. Lee, A regression approach to speech enhancement based on deep neural networks, IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 23 (2015) 7-19.
[43] A.L. Maas, A.Y. Hannun, A.Y. Ng, Rectifier nonlinearities improve neural network acoustic models, International conference on machine learning, Atlanta, USA, 2013, pp. 3.
[44] V. Nair, G.E. Hinton, Rectified linear units improve restricted boltzmann machines, Proceedings of the 27th international conference on machine learning, Omnipress, Haifa, Israel, 2010, pp. 807-814.
[45] W. Shi, J. Caballero, F. Huszár, J. Totz, A.P. Aitken, R. Bishop, D. Rueckert, Z. Wang, Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, Proceedings of the IEEE conference on computer vision and pattern recognition, IEEE, Las Vegas, USA, 2016, pp. 1874-1883.
[46] A. Odena, V. Dumoulin, C. Olah, Deconvolution and checkerboard artifacts, Distill, 1 (2016) e3.
[47] R. Brincker, L. Zhang, P. Andersen, Modal identification of output-only systems using frequency domain decomposition, Smart Materials and Structures, 10 (2001) 441.
[48] Y. Xia, Y.Q. Ni, P. Zhang, W.Y. Liao, J.M. Ko, Stress development of a supertall structure during construction: field monitoring and numerical analysis, Computer-Aided Civil and Infrastructure Engineering, 26 (2011) 542-559.
[49] W. Chen, Z. Lu, W. Lin, S. Chen, Y. Ni, Y. Xia, W. Liao, Theoretical and experimental modal analysis of the Guangzhou New TV Tower, Engineering Structures, 33 (2011) 3628-3646.
[50] F.L. Zhang, Y.Q. Ni, Y.C. Ni, Y.W. Wang, Operational modal analysis of Canton Tower by a fast frequency domain Bayesian method, Smart Structures and Systems, 17 (2016) 209-230.
[51] X.Y. Li, S.S. Law, Identification of structural damping in time domain, Journal of Sound and Vibration, 328 (2009) 71-84.

[52] I.M. Johnstone, B.W. Silverman, Empirical Bayes selection of wavelet thresholds, Annals of Statistics, (2005) 1700-1752.