School of Civil and Mechanical Engineering

Clustering and Deep Learning Techniques for Structural Health Monitoring

Gao Fan

This thesis is presented to the Degree of

Doctor of Philosophy

of

Curtin University

May 2020

DECLARATION

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgement has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature: …….

Date: ……..

ABSTRACT

Ensuring the safety and reliability of civil engineering structures is an important research topic in society. Structural health monitoring (SHM) as an advanced interdisciplinary research area has attracted significant interest from civil engineering community. SHM systems consist of a variety of sensors for measuring and collecting real-time vibrational responses and environmental conditions of structures. With continuously measured data, structural performance and vibration behaviour can be obtained for monitoring and assessing the structural condition to identify the structural anomaly at an early stage. However, barriers for effective application of SHM techniques include inevitable noise and response loss in the huge amount of continuous measurement data, leading to an urgent demand for high-efficiency and automated data processing techniques. Clustering-based techniques are effective tools for automated modal information extraction. Owning to the strong capability of automated feature extraction, deep learning as an emerging branch of machine learning techniques, has achieved outstanding progress in addressing problems with a large amount of data. The development and application of deep learning techniques for SHM is recently a hot topic that mainly focuses on structural condition and performance monitoring. With many successful applications in other fields such as image processing and audio enhancement, deep learning techniques are also promising for addressing data processing and recovery issues in SHM.

To develop the applicability of long-term SHM, an automated operational modal analysis (OMA) method is first proposed for extracting vibrational characteristics. Three procedures including hard validation criteria removal, an improved statistics-based clustering procedure, and a developed cluster merging procedure are combined for automatically interpreting the stabilization diagram from the stochastic subspace identification (SSI) output, without a priori knowledge on the modal parameters and no manual tuning during interpreting. Numerical validation results on a frame structure demonstrate that the proposed approach has strong noise immunity. Experimental studies on a laboratory steel frame structure and a real footbridge are conducted. Accurate identification results demonstrate the robustness and applicability of using the proposed approach for automated OMA and modal tracking. The extracted

vibrational characteristics can be applied in analysing structural long-term vibration behaviour and used as input to deep learning models for structural damage detection.

Recovery of randomly lost data due to inevitable transmission and environmental issues are investigated based on a bottlenecked convolutional neural network (CNN) featured with skip connection. Skip connection shuttles low-level features extracted by shallow layers directly to top layers, which effectively improves the information flow among networks and mitigate gradient vanish issue when applying gradient backpropagation. Data segmentations measured by the concerned sensor without any response loss are processed as training datasets to facilitate the proposed CNN to learn nonlinear relationships between complete and incomplete data. Two experimental studies for single-channel and multiple-channel data recovery are conducted using vibration data from a real bridge. In both cases, the lost data can be effectively recovered and modal parameters successfully identified from recovered data.

The proposed CNN is then updated for vibration data denoising. Paired synthetic noisy data and the corresponding original data are used for training. Experimental studies are conducted with acceleration data measured from Guangzhou New Television Tower. After denoising, the quality of the acceleration data with varying levels and different types of noises are improved. 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 data, which could not be reliably identified with the noisy data, indicating that the proposed deep learning model can automatically learn robust data features without any human intervention. These learned features can be regarded as modal parameters from an engineering point of view.

To reconstruct continuous response loss, a densely connected network (DenseNet) is carefully designed. DenseNet consisting of skip and dense connections shares all the advantages of skip connection. Moreover, densely connected layers possessing rich information interaction inside dense blocks are more outstanding in feature extraction than successive single layers connection. Inter-channel correlation between available and desired unavailable channels are learned by DenseNet. Experimental studies are conducted using the measured acceleration response from Guangzhou New Television Tower to investigate the

effects of the locations of available response, the numbers of available and unavailable channels, and measurement noise. The reconstructed response matches well with the baseline response in both the time and frequency domains with strong noise immunity. The applicability of the proposed approach for SHM is further demonstrated by the highly consistent modal parameters identified from the reconstructed and true responses.

A segment-based generative adversarial network (SegGAN) is developed based on DenseNet to reconstruct nonstationary response. Numerical studies with a frame structure and a nonlinear building model are conducted to evaluate the noise immunity and the capability of nonstationary response features extraction, respectively. Experimental studies with a laboratory frame structure under earthquake excitations, with measurement noise and varying environmental and operational conditions, are conducted to investigate the applicability of the proposed SegGAN. Results are compared with a traditional CNN and DenseNet. SegGAN outperforms the other two networks and produces distinguished reconstruction results in both time and frequency domains. Owning to the adversarial training strategy, SegGAN effectively and automatically extracts linear and nonlinear response features from limited training data with uncertainties. The advantages of less data requirement, complex environmental adaptability and automation in feature extraction make SegGAN very promising for reconstructing SHM data.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my sincere gratitude and respect to my supervisors Prof. Hong Hao and Associate Prof. Jun Li for their expert guidance and consistent patience. They have offered me valuable suggestions throughout my PhD study and raised me to a higher stage of research.

Secondly, I would like to thank the staff in the School of Civil and Mechanical Engineering for their friendly and kind help. I would also like to thank Curtin University for the Curtin International Postgraduate Research Scholarship (CIPRS) for sponsoring me for my research.

I would like to thank all members of my research centre, Centre for Infrastructure Monitoring and Protection (CIMP) for helping me from my research to personal life. Special thanks go to my colleagues Yu Xin, Pinghe Ni and Yizhou Lin, who have provided me valuable supports during my PhD study.

Lastly, I would like to thank my parents and girlfriend who have shared the happiness in my golden time and unconditionally supported me when I face difficulties.

LIST OF PUBLICATIONS

Chapter 2

Fan G., Li J., Hao H. (2019). Improved automated operational modal identification of structures based on clustering. *Structural Control and Health Monitoring*. *26*(12), e2450. https://doi.org/10.1002/stc.2450

Chapter 3

Fan G., Li J., Hao H. (2019). Lost data recovery for structural health monitoring based on convolutional neural networks. *Structural Control and Health Monitoring*. *26*(10): e2433. https://doi.org/10.1002/stc.2433

Chapter 4

Fan G., Li J., Hao H.（2020）. Vibration signal denoising for structural health monitoring by residual convolutional neural networks. *Measurement*. 107651. https://doi.org/10.1016/j.measurement.2020.107651

Chapter 5

Fan G., Li J., Hao H. (in press). Dynamic response reconstruction for structural health monitoring using densely connected convolutional networks. *Structural Health Monitoring*. https://doi.org/10.1177/1475921720916881

Chapter 6

Fan G., Li J., Hao H., Xin Y. (2020). Dynamic response reconstruction under extreme loading conditions using segment based generative adversarial network. (Under review).

STATEMENT OF CONTRIBUTION OF OTHERS

The work presented in this thesis was primarily designed, numerically and experimentally conducted, and the manuscripts and thesis are written by the first author (Gao Fan). Contributions by supervisors and others are described as follows. The signed contribution forms are attached in the appendix.

Chapter 2

Prof. Hong Hao and Associate Prof. Jun Li helped to define the research scope and objective of this study. Associate Prof. Jun Li conducted the experimental test. Prof. Hong Hao and Associate Prof. Jun Li revised and edited the manuscript, provided professional input toward data analysis and results discussion.

Chapter 3 to 5

Prof. Hong Hao and Associate Prof. Jun Li considered the real application of the proposed methods for civil engineering, offered valuable suggestions for improving the reliability and applicability of these methods. They also revised and edited the manuscript, provided intellectual ideas for data analysis and results discussion.

Chapter 6

Prof. Hong Hao and Associate Prof. Jun Li revised and edited the manuscript, provided intellectual ideas for data analysis and results discussion. The same experimental test as used in Chapter 2 adopted here for analysis was conducted by Associate Prof. Jun Li. Yu Xin helped in designing the nonlinear building model.

LIST OF RELEVANT ADDITIONAL PUBLICATIONS

Li J., Fan G., Hao H. Li H. (2018). *An improved SSI approach for structural modal identification*. Paper presented at the 25th Australasian Conference on Mechanics of Structures and Materials (ACMSM25), Brisbane, Australia.

Fan G., Li J., Hao H. (2019). *Using deep learning technique for recovering the lost measurement data*. Paper presented at the 16th East Asia-Pacific Conference on Structural Engineering & Construction (EASEC16), Brisbane, Australia.

Hao H., Li J., Xin Y., Bi K., Fan G. (2018). *Living laboratory project at Curtin University for building structural condition monitoring*. Paper presented at the Australian Earthquake Engineering Society Annual Conference, Perth, Australia.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION

1.1 Background

1.1.1 Current status of SHM

With the ageing of existing structures, development of effective means for monitoring the conditions to ensure the safety and reliability of civil engineering structures is now an important research topic of society. Instead of visual inspection, where minor damage may not be effectively detected, vibration-based SHM as an advanced interdisciplinary research area has attracted significant interest in civil engineering community. Vibration-based SHM systems consist of a variety of sensors for measuring and collecting real-time vibration responses and environmental conditions of structures. Those measured data can assist engineers for real-time assessing the structural condition and early warning of the anomaly to prevent calamity (Farrar & Worden, 2012). Owning to the significant benefits to monitor structural condition and ensure public safety, a rapidly growing number of long-term SHM systems have been deployed in large-scale civil structures (Chen et al., 2011; Cross et al., 2013; Nguyen et al., 2015; Peeters & De Roeck, 2001; Xia et al., 2013; Zhang et al., 2012). However, effectively recording and utilizing the measured data for identifying structural conditions still face difficulties. Barriers of applying SHM for real projects include a great amount of data used for analysis, inevitable noise in measurements and response loss in a huge amount of continuous measurement data during the service system outrage or poor transmission quality, leading to an urgent demand for efficient and automated data processing techniques for modal analysis, signal pre-processing and data recovery.

1.1.2 Automated modal parameters identification by clustering

In SHM, rather than directly use measured raw vibration data, modal parameters representing inherent structural vibration characteristics are more widely used for assessing structural conditions and detecting structural damages. However, modal identification

involving manual tuning and interpretation is time-consuming and labor-intensive for real time SHM with a large amount of continuously measured data. Alternatively, automated identification of structural modal parameters becomes important for extracting long-term structural vibration characteristics and determining the long-term structural vibration behaviours. Operational Modal Analysis (OMA) that identifies modal parameters by using structural vibration responses under operational and environmental excitations has been widely applied for civil engineering structures. Without interrupting structural operations and conducting vibration test for structures, OMA is economic and efficient. In the last decades, several automated procedures have been developed based on OMA techniques including the time domain parametric identification methods such as Stochastic Subspace Identification (SSI) (Van Overschee & De Moor, 2012) and Eigensystem Realization Algorithm (ERA) (Yang et al., 2018; Zhang et al., 2014), and the frequency domain non-parametric method, e.g. Frequency Domain Decomposition (FDD) (Rainieri & Fabbrocino, 2010). Among these methods, SSI has attracted a significant amount of attentions owing to its clear mathematical background and steady performance. Traditionally, modal analysis through SSI commonly involves two steps, namely, 1) Identification of mode candidates in a range of system orders; and 2) Interpretation of the SSI output via the empirical observation from the stabilization diagram and expert engineering knowledge to distinguish the physical modes and spurious modes. In terms of the automated modal identification procedures based on SSI, the automation processes mainly contribute to reducing the user intervention in modal analysis, especially in the interpretation of the SSI output to select the physical modes from the stabilization diagram using clustering techniques (Bakir, 2011; Boroschek & Bilbao, 2019; Cardoso et al., 2017; de Almeida Cardoso et al., 2018; Magalhaes et al., 2009; Neu et al., 2017; Reynders et al., 2012; Sun et al., 2017; Teng et al., 2019; Ubertini et al., 2013). However, those existing clustering-based methods have not realized fully automation, manually selection of one or few parameters is still indispensable for effective identification. Therefore, there is potential and need to further develop these methods for automated identification.

1.1.3 Vibration response reconstruction

Structural response loss is an inevitable issue during wireless transmission that affects the reliability and usability of measured vibration data. It has been reported that the effect of a 0.4% data loss in measured data is equivalent to injecting a 5% noise into the data on the estimated power spectrum density (Nagayama et al., 2007). Measured vibration responses with a high data loss ratio, i.e., the number of lost and fully recorded data, can hardly be used for the modal analysis as it may lead to significant errors in the identification results, which consequently influences the accuracy of structural condition assessment and damage detection. The potential reasons for data loss in a measurement system are complex, including the radio interaction, ambient condition variation and signal attenuation in the long-distance transmission. Furthermore, during the long-term monitoring, data could be lost due to the technical issues such as the malfunction of cable connections, power supply interruption, signal transmission disturbance, sensor malfunction, or regular maintenances such as equipment check and sensor replacement. In practice, the locations of installed sensors on the monitored structures are selected carefully, and the number of sensors is much smaller compared to the total number of degrees of freedom (DOFs) of structures because of the limitations on the budget and available channels of data acquisition systems (Law et al., 2011), and the inaccessibility of some locations for measurement during operations (Kammer, 1997). Common condition assessment and damage detection techniques assume that all the sensors are functioning in perfect condition (Ma et al., 2018). The incomplete acquired measurements with partially or completely missed data of one or more channels will greatly affect the effectiveness of SHM. Therefore, reconstructing the lost data is of great importance.

In the past decades, several types of methods for the reconstruction of structural responses have been proposed utilizing the finite element model (FEM) based techniques. Transmissibility concept (Ribeiro et al., 2000) based structural response reconstruction methods were studied in both the frequency domain (Law et al., 2011) and wavelet domain (Li & Law, 2011). Deriving the transformation matrix utilizing the FEM of the structure to obtain the response at unavailable locations through the available responses is another typical

method (Kammer, 1997). This research was further combined with empirical mode decomposition (EMD), which decomposed the measurable time domain data into several Intrinsic Mode Function (IMFs), and reconstructed the responses at unavailable DOFs using the independent transfer equation of each IMF (He et al., 2012; Wan et al., 2014). An alternative structural response reconstruction method was developed based on optimal multi-type sensors placement and Kalman filter with unknown excitations (Xu et al., 2016; Zhang & Xu, 2016). These FEM based methods demand an accurate FEM of the target structure to calculate the transmissibility matrix, identify the mode shapes or determine the optimal sensor locations. However, developing an accurate FEM for large and complex civil engineering structures with a large number of elements and DOFs is time-consuming and challenging. Furthermore, system and geometry properties, and boundary conditions of the structures are changing with the variations of the environmental and operational conditions, which leads to the establishment of an accurate FEM representing the in-service structures much more difficult. On the other hand, with the rapid development and application of data analysis techniques, realizing the response reconstruction by the data-driven methods are attracting significant attention. The majority of these data-driven methods focus on reconstructing the strain data (Huang et al., 2011; Zhang & Luo, 2017) or the strain distribution (Chen et al., 2018) using the inter-channel correlation. Compared to the strain data, the acceleration response is more sensitive to the structural vibration and accumulated damage in structures. However, nonlinear inter-channel relationships are more complicated. Studies on the reconstruction of vibration acceleration data are mainly conducted to reconstruct the randomly missing data or short-term continuous lost data in a specific channel. For example, Bao et al. (2013) recovered the randomly lost acceleration data by leveraging the compressive sensing and further improved the developed method with group spare optimization (Bao et al., 2018) and machine learning techniques (Bao et al., 2020). Yang and Nagarajaiah (2016) also addressed the random data missing issue with sparse representation for inter-channel data reconstruction of low-rank structures. Wan and Ni (2019) proposed a Bayesian multi-task learning method with a multi-dimensional Gaussian process to predict the short-term continuous lost data using the measurements from the same sensor ahead of the forecasting.

Those methods either need expert knowledge to select the sparse domain of the measured data or can only recover short-term data loss. When a long-term successive acceleration response loss occurs or acceleration responses at critical but inaccessible locations are desirable, these methods may not be capable of accurately reconstructing the required responses.

1.1.4 Vibration response denoising

Measured responses can also be contaminated by inevitable noise from various sources such as environmental noise and instrumental noise, which significantly influences the usability and quality of data. 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. It has been reported that 10-20% noise injected into vibration data was considered as a high noise level in numerical studies (Li & Law, 2012; Li et al., 2012) when validating damage detection methods. Considering substantial uncertainties during in-field tests, measurement noise may exceed 20% level easily in situations with some poor testing conditions. Therefore, besides improving the noise immunity of SHM methods for modal analysis or damage detection, another important research topic named vibration data denoising, which aims to remove noise components or mitigate noise effect from measurements without affecting the quality of vibration data, has gained significant attention. Data denoising can improve the usability and quality of the measured data before any specific analysis. Recent researches have been conducted to perform the vibration data denoising in the time, frequency and time-frequency domains for different types of data (He et al., 2014). Conventional time domain averaging methods are more suitable for periodic signals (Braun, 2011). The denoising effect on other kinds of signals may be limited due to the varying vibration frequency or amplitude. 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 to be known, which limits the application of those methods. Time-frequency domain methods, such as wavelet transform based denoising techniques have been

widely investigated (Bi et al., 2019; Lin & Qu, 2000). 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 for the number of wavelet transform layers and packets to discriminate and remove noise but avoid losing useful components (Chegini et al., 2019; He et al., 2014). Other methods based on singular value decomposition (Zhao & Jia, 2017), empirical mode decomposition (Lei et al., 2013) and global projection (Hou et al., 2011) 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 vibration data denoising closely dependent on engineers' empirical experiences.

1.1.5 The potential of deep learning techniques for SHM

Deep learning as an emerging branch of machine learning, has achieved outstanding progress in multiple disciplines such as image processing, medical diagnosis, autopilot and so on. It is capable of automatically identifying the pattern and mining the hidden highly abstract features from a large volume of data. Owing to the outstanding feature extraction capacity, deep learning is selected to realize the automated response reconstruction and denoising for SHM. The existing development and application of deep learning techniques in SHM mainly focus on the visual inspection and damage detection of structures by using the crack images (Cha et al., 2018; Gao & Mosalam, 2018; Xu et al., 2019), vibration responses (Abdeljaber et al., 2017; Lin et al., 2017; Zhang et al., 2019) and the pre-extracted vibration characteristics (Pathirage et al., 2018; Pathirage et al., 2019) as the input to the networks. For response reconstruction, Zhang et al. (2019) proposed a long-short term memory network that uses the measured excitation as input to predict the structural seismic response. Dozens of records under earthquakes must be provided for training which is inappropriate for those structures with short service time or located in few earthquake areas. A comprehensive review of deep learning-based SHM techniques is provided in Ref. (Ye et al., 2019).

However, the potential of using deep learning techniques for SHM is more than those applications. Applications in computer vision and speech enhancement have been reported, i.e. the pixel-to-pixel recovery of images and audio super resolution (Dong et al., 2015; Kuleshov et al., 2017) and image inpainting (Mao et al., 2016; Pathak et al., 2016; Yeh et al., 2017). Denoising Auto-Encoder (DAE) (Vincent et al., 2010) and CNN (Krizhevsky et al., 2012) have been extended to denoise image (Jain & Seung, 2009; Xie et al., 2012), audio (Park & Lee, 2016; Xia & Bao, 2013) and medical domain signals (Poungponsri & Yu, 2013). Considering that vibration response contains vibration features of structures, deep learning technique has high potential to be leveraged to conduct the response reconstruction and denoising.

1.2 Research objectives

The research aim of this thesis is to improve the applicability of SHM in civil engineering based on clustering and deep learning techniques. The specific research tasks in this study presented in the thesis include:

1. Developing an automated modal identification method based on clustering technique, and evaluating the effectiveness and robustness of the proposed method using measurement data from a numerical model, an experimental structure and a real bridge;
2. Designing a convolutional neural network for recovering randomly lost vibration data, and investigating its capability based on measured vibration data of a real bridge;
3. Improving the convolutional neural network that is specialised for denoising vibration signal, and validating the effectiveness and robustness of the proposed network using measured data from Guangzhou New TV Tower;
4. Developing a densely connected neural network for structural response reconstruction based on inter-channel correlation, and validating the effectiveness and robustness of the proposed network using measured data from Guangzhou New TV Tower; and
5. Designing a segment-based generative adversarial network for reconstructing nonstationary structural response, and investigating the effectiveness and usability of the

proposed network through a numerical study with a frame model under, a numerical study of a nonlinear building model and an experimental study of a frame structure under earthquake excitation.

1.3 Research outline

This thesis comprises seven chapters. The contents of six chapters following the introduction chapter are described as follows:

Chapter 2 proposes an improved automated operational modal analysis approach based on data-driven Stochastic Subspace Identification (SSI) and clustering techniques with novel criteria. The performance of the proposed method is validated and compared with an existing method by numerical studies on a frame model, experimental studies on a frame structure and a real bridge.

Chapter 3 proposes a novel approach based on CNN for recovering the lost vibration data for SHM. The effectiveness of recovering single channel, multiple channel and the effect of sample rate are investigated based on measured vibration data from a real bridge. Modal identification is conducted to validate the applicability of using recovered data.

Chapter 4 proposes a vibration signal denoising approach for SHM based on a specialised Residual Convolutional Neural Networks. Dropout, skip connection and sub-pixel shuffling techniques are featured to improve the performance. The capability of this developed method on denoising various levels and different kinds of noises are validated with vibration data measured from Guangzhou New TV Tower. Modal identification of denoised data is conducted to evaluate the improvement of the usability of denoised data.

Chapter 5 proposes a novel dynamic response reconstruction approach for structural health monitoring by using Densely Connected Convolutional Networks. Experimental studies are conducted using the measured acceleration responses from Guangzhou New TV Tower to investigate the effects of the locations of available responses, the numbers of available and unavailable channels, and measurement noise. Reconstructed responses are further used for modal identification to demonstrate the usability and accuracy of the reconstructed responses.

Chapter 6 proposes a segment-based generative adversarial network for dynamic response reconstruction under extreme loading conditions. The proposed network is a conditional generative adversarial network consisting of a bottlenecked generator and a segment-based discriminator. Numerical studies on a frame structure are conducted to evaluate the noise immunity. Another numerical study with a nonlinear building model is designed to validate the capability of the proposed approach on nonlinear response features extraction. Experimental study on a steel frame structure is conducted to investigate the effectiveness and robustness of the proposed method on real applications containing measurement noise and varying environmental and operational conditions.

1.4 References

Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M., & Inman, D. J. (2017). Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration, 388,* 154-170.

Bakir, P. G. (2011). Automation of the stabilization diagrams for subspace based system identification. *Expert Systems with Applications, 38*(12), 14390-14397.

Bao, Y., Li, H., Sun, X., Yu, Y., & Ou, J. (2013). Compressive sampling–based data loss recovery for wireless sensor networks used in civil structural health monitoring. *Structural Health Monitoring, 12*(1), 78-95.

Bao, Y., Shi, Z., Wang, X., & Li, H. (2018). Compressive sensing of wireless sensors based on group sparse optimization for structural health monitoring. *Structural Health Monitoring, 17*(4), 823-836.

Bao, Y., Tang, Z., & Li, H. (2020). Compressive-sensing data reconstruction for structural health monitoring: a machine-learning approach. *Structural Health Monitoring, 19*(1), 293-304.

Bi, F., Ma, T., & Wang, X. (2019). Development of a novel knock characteristic detection method for gasoline engines based on wavelet-denoising and EMD decomposition. *Mechanical Systems and Signal Processing, 117,* 517-536.

Boroschek, R. L., & Bilbao, J. A. (2019). Interpretation of stabilization diagrams using density-based clustering algorithm. *Engineering Structures, 178,* 245-257.

Braun, S. (2011). The synchronous (time domain) average revisited. *Mechanical Systems and Signal Processing, 25*(4), 1087-1102.

Cardoso, R., Cury, A., & Barbosa, F. (2017). A robust methodology for modal parameters estimation applied to SHM. *Mechanical Systems and Signal Processing, 95,* 24-41.

Cha, Y. J., Choi, W., Suh, G., Mahmoudkhani, S., & Büyüköztürk, O. (2018). Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. *Computer-Aided Civil and Infrastructure Engineering, 33*(9), 731-747.

Chegini, S. N., Bagheri, A., & Najafi, F. (2019). Application of a new EWT-based denoising technique in bearing fault diagnosis. *Measurement, 144*, 275-297.

Chen, W., Lu, Z., Lin, W., Chen, S., Ni, Y., Xia, Y., & Liao, W. (2011). Theoretical and experimental modal analysis of the Guangzhou New TV Tower. *Engineering Structures, 33*(12), 3628-3646.

Chen, Z., Bao, Y., Li, H., & Spencer Jr, B. F. (2018). A novel distribution regression approach for data loss compensation in structural health monitoring. *Structural Health Monitoring, 17*(6), 1473-1490.

Cross, E., Koo, K., Brownjohn, J., & Worden, K. (2013). Long-term monitoring and data analysis of the Tamar Bridge. *Mechanical Systems and Signal Processing, 35*(1-2), 16-34.

de Almeida Cardoso, R., Cury, A., & Barbosa, F. (2018). A clustering-based strategy for automated structural modal identification. *Structural Health Monitoring, 17*(2), 201-217.

Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence, 38*(2), 295-307.

Farrar, C. R., & Worden, K. (2012). *Structural health monitoring: a machine learning perspective*: John Wiley & Sons.

Gao, Y., & Mosalam, K. M. (2018). Deep transfer learning for image-based structural damage recognition. *Computer-Aided Civil and Infrastructure Engineering, 33*(9), 748-768.

He, J., Guan, X., & Liu, Y. (2012). Structural response reconstruction based on empirical mode decomposition in time domain. *Mechanical Systems and Signal Processing, 28*, 348-366.

He, Q., Wang, X., & Zhou, Q. (2014). Vibration sensor data denoising using a time-frequency manifold for machinery fault diagnosis. *Sensors, 14*(1), 382-402.

Hou, S., Liang, M., & Li, Y. (2011). An optimal global projection denoising algorithm and its application to shaft orbit purification. *Structural Health Monitoring, 10*(6), 603-616.

Huang, Y.-W., Wu, D.-G., & Li, J. (2011). Structural healthy monitoring data recovery based on extreme learning machine. *Computer Engineering, 16*, 241-243.

Jain, V., & Seung, S. (2009). *Natural image denoising with convolutional networks.* Paper presented at the Advances in Neural Information Processing Systems, Vancouver, Canada.

Kammer, D. C. (1997). Estimation of structural response using remote sensor locations. *Journal of Guidance, Control, and Dynamics, 20*(3), 501-508.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks.* Paper presented at the Advances in neural information processing systems, Lake Tahoe, USA.

Kuleshov, V., Enam, S. Z., & Ermon, S. (2017). Audio super resolution using neural networks. *https://arxiv.org/abs/1708.00853*.

Law, S., Li, J., & Ding, Y. (2011). Structural response reconstruction with transmissibility concept in frequency domain. *Mechanical Systems and Signal Processing, 25*(3), 952-968.

Lei, Y., Lin, J., He, Z., & Zuo, M. J. (2013). A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing, 35*(1-2), 108-126.

Li, J., & Law, S. (2011). Substructural response reconstruction in wavelet domain. *Journal of Applied Mechanics, 78*(4), 041010.

Li, J., & Law, S. (2012). Damage identification of a target substructure with moving load excitation. *Mechanical Systems and Signal Processing, 30*, 78-90.

Li, J., Law, S., & Ding, Y. (2012). Substructure damage identification based on response reconstruction in frequency domain and model updating. *Engineering Structures, 41*, 270-284.

Lin, J., & Qu, L. (2000). Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis. *Journal of Sound and Vibration, 234*(1), 135-148.

Lin, Y. Z., Nie, Z. H., & Ma, H. W. (2017). Structural damage detection with automatic feature-extraction through deep learning. *Computer-Aided Civil and Infrastructure Engineering, 32*(12), 1025-1046.

Ma, S., Li, J., Hao, H., & Jiang, S. (2018). Structural response recovery based on improved multi-scale principal component analysis considering sensor performance degradation. *Advances in Structural Engineering, 21*(2), 241-255.

Magalhaes, F., Cunha, A., & Caetano, E. (2009). Online automatic identification of the modal parameters of a long span arch bridge. *Mechanical Systems and Signal Processing, 23*(2), 316-329.

Mao, X., Shen, C., & Yang, Y. B. (2016). *Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections.* Paper presented at the Advances in neural information processing systems, Barcelona, Spain.

Nagayama, T., Sim, S.-H., Miyamori, Y., & Spencer Jr, B. (2007). Issues in structural health monitoring employing smart sensors. *Smart Structures and Systems, 3*(3), 299-320.

Neu, E., Janser, F., Khatibi, A. A., & Orifici, A. C. (2017). Fully Automated Operational Modal Analysis using multi-stage clustering. *Mechanical Systems and Signal Processing, 84, Part A*, 308-323.

Nguyen, T., Chan, T. H. T., Thambiratnam, D. P., & King, L. (2015). Development of a cost-effective and flexible vibration DAQ system for long-term continuous structural health monitoring. *Mechanical Systems and Signal Processing, 64-65*, 313-324.

Park, S. R., & Lee, J. (2016). A fully convolutional neural network for speech enhancement. *https://arxiv.org/abs/1609.07132*.

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). *Context encoders: feature learning by inpainting.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, USA.

Pathirage, C. S. N., Li, J., Li, L., Hao, H., Liu, W., & Ni, P. (2018). Structural damage identification based on autoencoder neural networks and deep learning. *Engineering Structures, 172*, 13-28.

Pathirage, C. S. N., Li, J., Li, L., Hao, H., Liu, W., & Wang, R. (2019). Development and application of a deep learning–based sparse autoencoder framework for structural damage identification. *Structural Health Monitoring, 18*(1), 103-122.

Peeters, B., & De Roeck, G. (2001). One-year monitoring of the Z24-Bridge: environmental effects versus damage events. *Earthquake engineering & structural dynamics, 30*(2), 149-171.

Poungponsri, S., & Yu, X.-H. (2013). An adaptive filtering approach for electrocardiogram (ECG) signal noise reduction using neural networks. *Neurocomputing, 117*, 206-213.

Rainieri, C., & Fabbrocino, G. (2010). Automated output-only dynamic identification of civil engineering structures. *Mechanical Systems and Signal Processing, 24*(3), 678-695.

Reynders, E., Houbrechts, J., & De Roeck, G. (2012). Fully automated (operational) modal analysis. *Mechanical Systems and Signal Processing, 29*, 228-250.

Ribeiro, A., Silva, J., & Maia, N. (2000). On the generalisation of the transmissibility concept. *Mechanical Systems and Signal Processing, 14*(1), 29-35.

Sun, M., Makki Alamdari, M., & Kalhori, H. (2017). Automated operational modal analysis of a cable-stayed bridge. *Journal of Bridge Engineering, 22*(12), 05017012.

Teng, J., Tang, D.-H., Zhang, X., Hu, W.-H., Said, S., & Rohrmann, R. (2019). Automated modal analysis for tracking structural change during construction and operation phases. *Sensors, 19*(4), 927.

Ubertini, F., Gentile, C., & Materazzi, A. L. (2013). Automated modal identification in operational conditions and its application to bridges. *Engineering Structures, 46*, 264-278.

Van Overschee, P., & De Moor, B. (2012). *Subspace identification for linear systems: Theory—Implementation—Applications*: Springer Science & Business Media.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.-A. (2010). Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research, 11*(Dec), 3371-3408.

Wan, H.-P., & Ni, Y.-Q. (2019). Bayesian multi-task learning methodology for reconstruction of structural health monitoring data. *Structural Health Monitoring, 18*(4), 1282-1309.

Wan, Z., Li, S., Huang, Q., & Wang, T. (2014). Structural response reconstruction based on the modal superposition method in the presence of closely spaced modes. *Mechanical Systems and Signal Processing, 42*(1-2), 14-30.

Xia, B., & Bao, C. (2013). *Speech enhancement with weighted denoising auto-encoder.* Paper presented at the Interspeech, Lyon, France.

Xia, Y., Chen, B., Zhou, X., & Xu, Y. (2013). Field monitoring and numerical analysis of Tsing Ma Suspension Bridge temperature behavior. *Structural Control and Health Monitoring, 20*(4), 560-575.

Xie, J., Xu, L., & Chen, E. (2012). *Image denoising and inpainting with deep neural networks.* Paper presented at the Advances in Neural Information Processing Systems, Lake Tahoe, USA.

Xu, Y., Wei, S., Bao, Y., & Li, H. (2019). Automatic seismic damage identification of reinforced concrete columns from images by a region-based deep convolutional neural network. *Structural Control and Health Monitoring, 26*(3), e2313.

Xu, Y., Zhang, X., Zhu, S., & Zhan, S. (2016). Multi-type sensor placement and response reconstruction for structural health monitoring of long-span suspension bridges. *Science bulletin, 61*(4), 313-329.

Yang, X.-M., Yi, T.-H., Qu, C.-X., Li, H.-N., & Liu, H. (2018). Automated eigensystem realization algorithm for operational modal identification of bridge structures. *Journal of Aerospace Engineering, 32*(2), 04018148.

Yang, Y., & Nagarajaiah, S. (2016). Harnessing data structure for recovery of randomly missing structural vibration responses time history: Sparse representation versus low-rank structure. *Mechanical Systems and Signal Processing, 74*, 165-182.

Ye, X., Jin, T., & Yun, C. (2019). A review on deep learning-based structural health monitoring of civil infrastructures. *Smart Structures and Systems, 24*(5), 567-585.

Yeh, R. A., Chen, C., Lim, T.-Y., Schwing, A. G., Hasegawa-Johnson, M., & Do, M. N. (2017). *Semantic image inpainting with deep generative models.* Paper presented at the The IEEE conference on computer vision and pattern recognition, Honolulu, USA.

Zhang, C. D., & Xu, Y. (2016). Optimal multi-type sensor placement for response and excitation reconstruction. *Journal of Sound and Vibration, 360*, 112-128.

Zhang, D., Bao, Y., Li, H., & Ou, J. (2012). Investigation of temperature effects on modal parameters of the China National Aquatics Center. *Advances in Structural Engineering, 15*(7), 1139-1153.

Zhang, G., Ma, J., Chen, Z., & Wang, R. (2014). Automated eigensystem realisation algorithm for operational modal analysis. *Journal of Sound and Vibration, 333*(15), 3550-3563.

Zhang, R., Chen, Z., Chen, S., Zheng, J., Büyüköztürk, O., & Sun, H. (2019). Deep long short-term memory networks for nonlinear structural seismic response prediction. *Computers & Structures, 220*, 55-68.

Zhang, Y., Miyamori, Y., Mikami, S., & Saito, T. (2019). Vibration-based structural state identification by a 1-dimensional convolutional neural network. *Computer-Aided Civil and Infrastructure Engineering, 34*(9), 822-839.

Zhang, Z., & Luo, Y. (2017). Restoring method for missing data of spatial structural stress monitoring based on correlation. *Mechanical Systems and Signal Processing, 91*, 266-277.

Zhao, M., & Jia, X. (2017). 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*, 129-147.

CHAPTER 2 IMPROVED AUTOMATED OPERATIONAL MODAL IDENTIFICATION OF STRUCTURES BASED ON CLUSTERING

Abstract: A robust automated method for operational modal analysis (OMA) is in a great demand for processing a large amount of structural health monitoring data from engineering structures. This chapter proposes an improved automated OMA approach based on data-driven Stochastic Subspace Identification (SSI) and clustering techniques with novel criteria. The framework of the proposed approach includes two main components, namely "Modal identification by SSI" and "Automated interpretation of SSI output". Three procedures including hard validation criteria removal, an improved statistics-based clustering procedure and a developed cluster merging procedure are combined in the second component for automatically interpreting the stabilization diagram from the SSI output, without a priori knowledge on the modal parameters and no manual tuning during interpreting. Numerical validation results on a frame structure model demonstrate that the proposed approach is capable of identifying the vibration modes accurately, under a significant noise effect. No spurious modes are observed and the physical modes can be accurately identified. Experimental studies on a steel frame structure in the laboratory and a real footbridge are conducted to demonstrate the robustness and applicability of using the proposed approach for automated OMA and modal tracking. Identification results are compared with baselines and those from an existing reference method to demonstrate the improvement and contribution made in the proposed approach on the automated OMA.

This chapter was published in *Structural Control and Health Monitoring* with the full bibliographic citation as follows: Fan G, Li J, Hao H. (2019). Improved automated operational modal identification of structures based on clustering. *Structural Control and Health Monitoring*, *26*(12): e2450. https://doi.org/10.1002/stc.2450.

2.1 Introduction

Long-term structural health monitoring (SHM) systems have been installed on a significant number of civil engineering structures, such as bridges (Peeters & De Roeck, 2001; Xia et al., 2013; Zhang et al., 2017) and buildings (Nayeri et al., 2008; Yuen & Kuok, 2010). Automated modal analysis method is of great interest for SHM applications to real engineering structures.

Identification of modal parameters by using the structural vibration responses under operational and environmental excitations is known as Operational Modal Analysis (OMA). With a minimum influence on the structural operation, OMA techniques have been developed to extract the modal parameters from continuous vibration monitoring data of operating structures. In the last decades, several automated procedures have been developed based on OMA techniques including the time domain parametric identification methods such as Stochastic Subspace Identification (SSI) (Van Overschee & De Moor, 2012) and Eigensystem Realization Algorithm (ERA) (Yang et al., 2018; Zhang et al., 2014), and the frequency domain non-parametric method, i.e. Frequency Domain Decomposition (FDD) (Rainieri & Fabbrocino, 2010). Among these methods, SSI has attracted a significant amount of attention owning to its clear mathematical background and steady performance with a number of successful applications to real civil engineering structures. Traditionally, modal analysis through SSI commonly involves two steps, namely, 1) Identification of mode candidates in a range of system orders; and 2) Interpretation of the SSI output via the empirical observation from the stabilization diagram and expert engineering knowledge to distinguish the physical modes and spurious modes. In terms of the automated modal identification procedures based on SSI, the automation processes mainly contribute to reducing the user intervention in modal analysis, especially in the interpretation of the SSI output to select the physical modes from the stabilization diagram (Bakir, 2011; Boroschek & Bilbao, 2019; Cardoso et al., 2017; Reynders et al., 2012; Teng et al., 2019; Ubertini et al., 2013).

The major challenge for the automated OMA procedures is to distinguish the physical modes and spurious modes from a number of observed modes in the stabilization diagram. To

address this challenge, clustering techniques including hierarchical clustering and *K*-means clustering, which can be used to group the physical modes across various system model orders into the same clusters, are several of the most popular methods to automatically interpret the SSI output and select the physical modes. Hierarchical clustering lists all the mode candidates from the parametric identification as single clusters, and keeps merging two closest clusters among all the clusters as one until the minimum distance is larger than a predefined cutoff threshold (de Almeida Cardoso et al., 2017; Magalhaes et al., 2009). The physical modes are selected from the final remaining clusters that contain more elements. *K*-means clustering collects modes with similar properties into a user-defined number of groups. Usually, *K* is set as two, which means that the mode candidates are grouped into two clusters with one for physical modes and another one for spurious modes (Neu et al., 2017; Sun et al., 2017). Those clustering methods are incorporated into the automated OMA procedures to distinguish the physical modes.

This chapter proposes an improved automated structural OMA approach based on clustering techniques. The framework of the proposed approach includes two main components, namely "Modal identification by SSI" and "Automated interpretation of SSI output". The data-driven SSI is used in the first component. Three procedures including hard validation criteria removal, an improved statistics-based clustering procedure and a developed cluster merging procedure are contained in the second component. This study contributes to defining and updating the clustering criteria based on statistical properties of grouped modal parameters. These self-adapting criteria automatically group the physical modes into the large cluster and diffuse the spurious modes as small clusters. The cluster containing elements less than a static threshold will be defined as a spurious-mode cluster, and therefore be eliminated from the SSI output. A cluster merging procedure is also developed and arranged after the statistics-based clustering to enhance the robustness of the proposed approach, which merges the physical-mode clusters as larger clusters and further remove spurious-mode clusters by a more critical threshold. Numerical and experimental studies are conducted to demonstrate the accuracy and performance of using the proposed approach for automated OMA and modal

tracking. The results will be compared with those reported in the literature and obtained from a reference method (Cabboi et al., 2017).

The rest of this chapter is organized as follows: firstly, the proposed automated OMA approach will be presented and the background of the reference method will be briefly mentioned in Section 2.2. Numerical studies are conducted in Section 2.3 to validate the accuracy and robustness of the proposed approach, particularly under the significant noise effect. The performance of using the proposed approach to remove the false modes in the identification results is investigated. Section 2.4 presents the experimental studies on a laboratory frame structure and a real footbridge to validate the applicability and performance of the proposed approach. Modal identification results from the proposed approach and a reference method are compared to demonstrate the improvement of the proposed approach. Finally, the conclusions and discussions on further study topics are provided in Section 2.5.

2.2 Methodology

2.2.1 Background of data-driven SSI

The data-driven SSI technique has been developed for modal identification of structures for more than two decades, therefore it is considered as a relatively mature OMA method with applications to various civil engineering structures. The theoretical derivation of data-driven SSI has been described in the literature (Van Overschee & De Moor, 2012) in details. For completeness of this chapter, it will be briefly reviewed here. SSI is developed based on the discrete time state-space form of motion of a linear time invariant system under ambient white noise excitations. The stochastic state-space model is expressed as

$$\begin{cases} x_{k+1} = Ax_k + w_k \\ y_k = Cx_k + v_k \end{cases} \tag{2.1}$$

where k is the sampling instant; $A \in \Re^{n\times n}$ is the discrete state matrix with n equal to two times of the system degrees of freedom (DOFs); $C \in \Re^{l\times n}$ is the output matrix with l defined as the number of measured output responses. $x_k \in \Re^{n\times 1}$ and $y_k \in \Re^{l\times 1}$ denote the discrete state vector and the output measurements at the k-th sampling instant, respectively; $w_k \in$

$\Re^{n\times 1}$ and $v_k \in \Re^{n\times 1}$ are two white noise vectors, representing the process noise due to disturbances and modelling inaccuracies, and the measurement noise respectively.

Modal identification by using data-driven SSI includes the following steps: 1) Form the block Hankel matrix H using the output measurements; 2) Calculate the orthogonal projection of the 'past' output measurements matrix Y_p to the 'future' output measurements matrix Y_f to identify the projection matrix $P \in \Re^{l\cdot i\times j}$; 3) Compute the observability matrix Γ using the projection matrix *P* to determine the system matrices *A* and *C*; 4) Identify the modal parameters from the extracted system matrices *A* and *C* by performing the eigenvalue analysis. Since the selection of the number of the half block rows of the block Hankel matrix *i* will be discussed in the next sections, the block Hankel matrix is given here

$$\mathrm{H}(i) = \begin{pmatrix} y(0) & y(1) & \dots & y(j-1) \\ y(1) & y(2) & \dots & y(j) \\ \vdots & \vdots & \ddots & \vdots \\ y(i-1) & y(i) & \dots & y(i+j-2) \\ \hline y(i) & y(i+1) & \dots & y(i+j-1) \\ y(i+1) & y(i+2) & \dots & y(i+j) \\ \vdots & \vdots & \ddots & \vdots \\ y(2i-1) & y(2i) & \dots & y(2i+j-2) \end{pmatrix} = \begin{bmatrix} Y_p \\ \hline Y_f \end{bmatrix} \tag{2.2}$$

2.2.2 Overview of the proposed automated OMA approach

The proposed improved automated OMA approach consists of the complete process from raw ambient data pre-processing to automated modal parameters identification, as shown in Figure 2-1. It consists of two main components, namely Component 1: Modal identification by SSI and Component 2: Automated interpretation of SSI output. The first component 'Modal identification by SSI' requires the user intervention and manual tuning for the pre-processing of vibration response data and the selection of the model parameters in SSI, including half the number of block rows of the Hankel matrix *i*, and the range of a model order range [n_{min}, n_{max}] to estimate the state space system matrix. It should be highlighted that the pre-processing of raw data and the definition of model parameters are important for the modal parameters identification by using data-driven SSI. Refined data and properly selected parameters lead to

a qualified SSI output, i.e. the genuine physical modes are properly identified and less spurious modes are observed, which is the basis of effective parametric identification.

In most of the previous studies, the interpretation of the stabilization diagram from the SSI output is performed manually based on the expert experience. The second component "Automated interpretation of the SSI output" of the proposed approach is the main contribution of this study. As shown in Figure 2-1, the second component in this proposed approach includes three procedures, that is, hard validation criteria removal, statistics-based clustering and cluster merging, which will be elaborated in Section 2.2.4.1, 2.2.4.2 and 2.2.4.3, respectively.

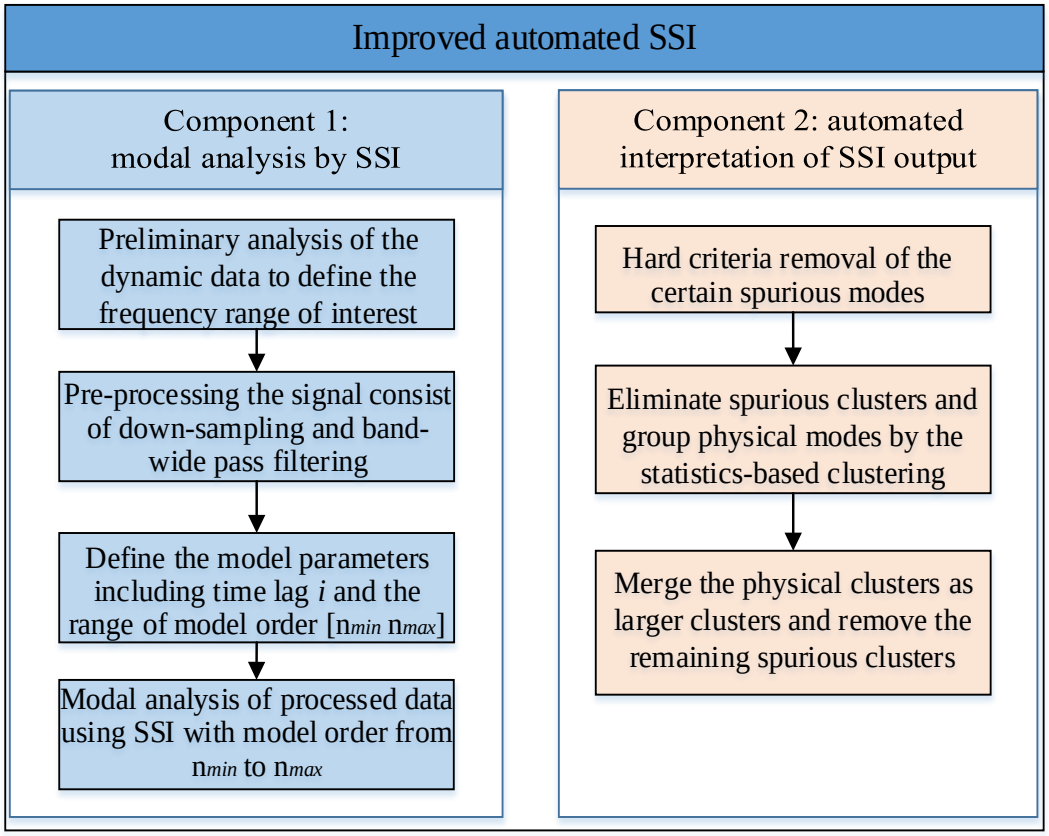


Figure 2-1. Flowchart of the proposed approach.

2.2.3 Model parameters selection in Component 1

The selection of the parameter *i* affects the accuracy of modal parameters identification results significantly. It may even fail if the duration of the time history signal ($i \times$ sampling

interval) is less than a period of the first structural mode. On the other hand, longer signals provide more information about the structural vibration, but without an upper limit, a large value of i enlarges the size of Hankel matrix and increase the computational demand significantly. Besides, it is also worth noting that with the inevitably measurement noise effect, using a larger value of i may introduce more spurious modes in the SSI output. Therefore, the parameter i should be selected carefully considering the period of the first mode of interest to control the number of spurious modes in the results (Priori et al., 2018; Reynders & De Roeck, 2008). In this study, the parameter i is chosen as

$$i \geq T_1/\Delta t \tag{2.3}$$

where T_1 is the period of the first structural mode and Δt is the sampling interval.

Modal information is identified at a large number of model orders observed from the SSI output, i.e. stabilization diagram which shows all the identified system poles against a pre-defined range of model orders. The physical modes and spurious modes are then distinguished based on the fact that physical modes are consistent in almost all the model orders, and spurious modes are scattered and observed only in some model orders. It is difficult to predict and anticipate which model order will give the best modal identification results from SSI. The lower bound of the system model order range n_{min} should be usually chosen twice as large as the number of the expected modes to be identified. On the other hand, the upper bound of the system model order n_{max} is usually high to ensure that the weakly excited modes can be successfully identified. However, n_{max} should be carefully determined because of the increasing computation cost and spurious modes with n_{max}. A selection of n_{max} between 120 and 160 in this study have no influence on the accuracy but the efficient. Based on the above selected model parameters, data-driven SSI can be conducted to produce the formed system modes matrix.

2.2.4 Procedures in Component 2

2.2.4.1 Removal of spurious modes by hard validation criteria

Despite the appropriate parameters being chosen following the specified rules as mentioned above, the system modes matrix may still exhibit considerable spurious modes, which increases the computational burden and affect the efficiency of the automated interpretation of SSI output. Certain spurious modes can be effectively removed from the identification results by using several hard validation criteria related to the damping ratios and mode shapes, which can be used to improve the identification efficiency and accuracy remarkably. For real civil structures, damping ratios ξ should be large than 0 and usually less than 10%. Identified modes with damping ratios out of this range can be deemed as spurious modes and removed from the system modes matrix. Further clearance is realized by evaluating the complexity of mode shapes measured with two well-known modal validation criteria, namely, Modal Phase Collinearity (MPC) and Mean Phase Deviation (MPD). MPC is an index to measure the spatial consistency of the identified mode that represents the mono-phase behaviour through computing the linear relation between the real part $\phi^R$and imaginary part ϕ^I of the identified mode shape vector (Pappa et al., 1993). MPC of the *j*th mode shape can be calculated as

$$\alpha = \frac{\phi_j^{I^T}\phi_j^I - \phi_j^{R^T}\phi_j^R}{2\phi_j^{R^T}\phi_j^I}, \beta = \alpha + \text{sign}\left(\phi_j^{R^T}\phi_j^I\right)\sqrt{\alpha^2+1} \tag{2.4}$$

$$MPC_j = \frac{\phi_j^{R^T}\phi_j^R - \phi_j^{I^T}\phi_j^I + \frac{1}{\beta}(\phi_j^{R^T}\phi_j^I)(2(\alpha^2+1)\sin^2\beta - 1)}{\phi_j^{R^T}\phi_j^R + \phi_j^{I^T}\phi_j^I} \tag{2.5}$$

MPC value ranges from 0 indicating a mode with the totally uncorrelated phase angle to 1 for a mono-phase behaviour, which is usually related to real normal modes.

MPD indicates the phase scatter for each identified mode shape vector. Similar to the principle of standard deviation, a lower value of MPD means that the phase of each element in the mode shape vector is close to the mean phase (MP). This may correspond to a normal mode. Instead of computing MP as the average of each phase angle, using the total least

squares method to find the angle of the linear regression model of the mode shape in the complex plane may provide more robust results (Reynders et al., 2012). MP of the *j*th mode shape is expressed as

$$MP_j = \arg_\theta \min \frac{\left\| \phi_j^I - \tan(\theta)\phi_j^R \right\|_2^2}{1 + \tan(\theta)} \tag{2.6}$$

where θ is the phase angle. It can be computed as

$$USV^T = \left[\phi_j^R \ \phi_j^I\right], MP_j = \arctan\left(\frac{-V_{12}}{V_{22}}\right) \tag{2.7}$$

where USV^T denotes the singular value decomposition. V_{12} and V_{22} are the elements of the matrix V at the position (1, 2) and (2, 2), respectively. MPD is then calculated as

$$MPD_j = \frac{\sum_{i=1}^{N_\phi} w_i \arccos\left|\frac{\phi_{ji}^R V_{22} - \phi_{ji}^I V_{12}}{\sqrt{V_{12}^2 + V_{22}^2}|\phi_{ji}|}\right|}{\sum_{i=1}^{N_\phi} w_i} \tag{2.8}$$

where N_ϕ is the number of elements in the *j*th mode shape vector; w_i is the weighting factor that is chosen as $|\phi_{ji}|$ in this study. A threshold value λ should be selected carefully based on the complexity of the structural vibration behaviour. The modes with MPD values large than λ and MPC values less than 1- λ will be considered as false modes and eliminated.

There are no standard threshold values for those criteria under ambient vibration testing. Setting a threshold value may depend on a number of factors, such as measurement noise, weak excitations and weakly excited modes and nonlinear behaviour of the structure. The selection is based on the experimental testing conditions and the expected characteristics of structures under investigation (Cabboi et al., 2017). Generally, for structures exhibiting clear linear behaviours with less weakly excited modes and measurement errors (signal synchronization problem), λ =0.3 is sufficient to ensure that the genuine physical modes will not be removed. On the other hand, setting the threshold λ as 0.7 for structures with complex behaviours or under complex experimental condition is conservative.

2.2.4.2 Statistics-based clustering

Even though the above hard validation criteria could remove certain spurious modes, it is still challenging to handle the scenarios for ambient structural vibration testing with a small signal-to-noise ratio and weakly excited modes. To further eliminate the spurious modes from the output, a novel automatic statistics-based clustering is proposed. The principle of this clustering strategy is to group the modes at different model orders with similar modal parameters including frequencies, damping ratios and mode shapes, to one cluster. Instead of setting constant clustering validation criteria based on engineers' experience for all tests, the proposed approach by the statistics-based clustering can automatically define the validation criteria to determine whether a new detected mode should be collected to a cluster without any user's manual intervention. These validation criteria are being updated in accordance with the statistical characteristics of all the previously collected modes and new detected modes to improve the quality of those criteria that represent more integrated information of the cluster. As a result, mode candidates that represent one structural mode with consistent modal parameters across most of the model orders are grouped to one or two large clusters. On the contrary, the discrete spurious modes are diffused to several clusters with very limited sizes.

With the data-driven SSI and filtering the certain spurious modes using the hard criteria, the system modes matrix containing the remaining mode candidates becomes the input for the clustering. The specific structure of the matrix is that modes of the system at the same model order are collected in one row, and the number of rows is equal to the total number of model order. One feature of the SSI output needs to be mentioned is that even the modal parameters of physical modes are consistent at most of the model orders (Neu et al., 2017), the identification of one mode in adjacent model orders tends to have particular similar modal parameters. This feature is widely used for validating the stability of modes and establish the 'stabilization diagram', which shows the stable modes at different model orders. However, the high similarity leads to an issue, that is, the first several modes collected in one cluster may demonstrate 'local' statistical characteristics of the cluster. The criteria defined in terms of 'local' statistical characteristics may lose the robustness, and are converged to very tiny values

such as 10^{-5} or even 10^{-6} because of the similarities in the selected modes. Consequently, the strict validation criteria may reject modes in successive model orders representing the same mode due to the slightly large discrepancy in one or more modal parameters. Therefore, a novel procedure for dealing with this limit is to randomize the model order in the system modes matrix before clustering so that the first several collected modes in a cluster represent more global statistical characteristics.

Before one cluster is grouped with a sufficient number of modes to calculate its statistic properties, the initial clustering validation criteria of frequency, damping ratio and MAC are defined referring to the stability limits of the conventional stabilization diagram in SSI based techniques. The threshold values of T_f and T_{MAC} are set as 1% and 2%, respectively. Due to the large scattering damping ratios, the threshold of the criterion on the damping ratio T_d is defined from 10% to 20% depending on the signal-to-noise ratio. MPC is found to be consistent at different model orders for the physical modes. Therefore, a criterion of MPC is added as a supplementary criterion to further filter out the spurious modes that may influence the statistical characteristics of the cluster. The validation criteria will be updated when the number of elements in this cluster is larger than 10% of the total model order used n_{total} ($n_{total} = (n_{max} - n_{min} + 1)$). Using 5% to 10% as the initial MPC criterion value T_{MPC} will have a limited influence on the final results. The number of modes as the sample of a cluster is chosen to be 10% of the total model order to stably represent the initial statistical characteristics of the cluster. Including the SSI parameters and hard criteria (necessary for all the OMA), all the parameters are predefined regarding the properties of the objective structure and test conditions. The interpretation of the SSI stabilization diagram is automated and no user intervention is involved during the interpretation. Figure 2-2 shows the flowchart of the proposed method with the detailed procedure of clustering and cluster merging. The pseudocode of the statistics-based clustering procedure is also attached in Appendix I to describe the procedure in detail. The steps in this clustering strategy are further explained as follows:

Step 1. Randomize the cleaned system modes matrix as *S*.

Step 2. The clustering process will group all the modes in the cleaned and random ordered system modes matrix to clusters. All the identified modes will be scanned with a sequence from the first to the last mode at all the model orders. Once an ungrouped mode is found, a new cluster C_k is created and this ungrouped mode is recorded as the first element of C_k. It is noted that k starts from 1 and ends up with the total number of clusters.

Step 3. Modes at the successive model order i are scanned, and the ungrouped mode m_{min} with the minimum distance to cluster C_k is selected. The distance between the mode and the cluster C_k is defined based on the variation between the mean frequency $\overline{f_k}$ and mode shape $\overline{\phi_k}$ of the cluster C_k and the mode frequency $f_{i,j}$ and mode shape $\phi_{i,j}$. The correlation in the mode shapes is evaluated by using Modal Assurance Criterion (MAC). The distance between the mode at the position $S[i, j]$ (the ith model order and the jth mode) to C_k is computed as

$$D_{i,j} = D_{f\,i,j} + D_{\phi\,i,j} = \frac{\left|\overline{f_k} - f_{i,j}\right|}{\overline{f_k}} + 1 - MAC(\overline{\phi_k}, \phi_{i,j}) \tag{2.9}$$

Step 4. Modal parameters of the mode $m_{\min}$ are extracted, and the variations of frequency, damping ratio, mode shape and MPC between the mode $m_{\min}$ and cluster C_k are computed. The most meaningful frequency, mode shape and MPC values for cluster C_k are chosen as the average of currently collected elements. The median of damping ratio is chosen because the identified damping ratios are usually relatively discrete within a cluster, the use of median can minimize the influence of outliers. Since the statistical characteristics of the cluster C_k are unknown in prior, the initial validation criteria for collected modes are defined as 1%, 20%, 2% and 10% for frequency, damping ratio, mode shape and MPC, respectively. Mode $m_{\min}$ will be collected to cluster C_k when all the following criteria are satisfied

$$\begin{aligned} &\frac{\left|\overline{f_k} - f_m\right|}{\overline{f_k}} \le \mathrm{T}_f; && \frac{\left|\mathrm{median}(\xi_k) - \xi_m\right|}{\xi_i} \le \mathrm{T}_d; \\ &1 - MAC(\overline{\phi_k}, \phi_m) \le \mathrm{T}_{MAC}; && \frac{\left|\overline{MPC_k} - MPC_m\right|}{\overline{MPC_k}} \le \mathrm{T}_{MPC}; \end{aligned} \tag{2.10}$$

Step 5. If mode m is collected to cluster C_k and the number of collected modes in cluster C_k is larger than 10% of n_{total} threshold, update the clustering criteria in Equation (2.10) every

time the new mode being collected considering the statistic properties of this cluster. The validation criteria of the cluster C_k are updated by computing the difference of frequency, damping ratio, mode shape and MPC between each element and the cluster C_k to form four 'difference vectors'. i.e. the difference vector for the natural frequency of modes in C_k is computed as:

$$d_f = \frac{|f_k - \overline{f_k}|}{f_k} \tag{2.11}$$

The mean value μ and standard deviation σ of each difference vector are obtained to update the threshold values of validation criteria T_f, T_d, T_{MAC}, and T_{MPC} respectively based on the 99.7% rule of their corresponding difference vectors. This corresponds to a physical explanation that 99.7% of the values lie within three standard deviations of the mean value, which can be expressed as

$$T = \mu + 3\sigma \tag{2.12}$$

Step 6. Continue grouping modes for cluster C_k until the last row of S by repeating Steps 3 and 4 and update the thresholds by Step 5 if necessary. If i reaches the last row of S, the clustering for C_k is terminated. Update k as $k=k+1$ and go back to Step 2 to start clustering for the next cluster.

Step 7. Once all the modes are collected, the clusters containing less than 10% of the total model order are eliminated.

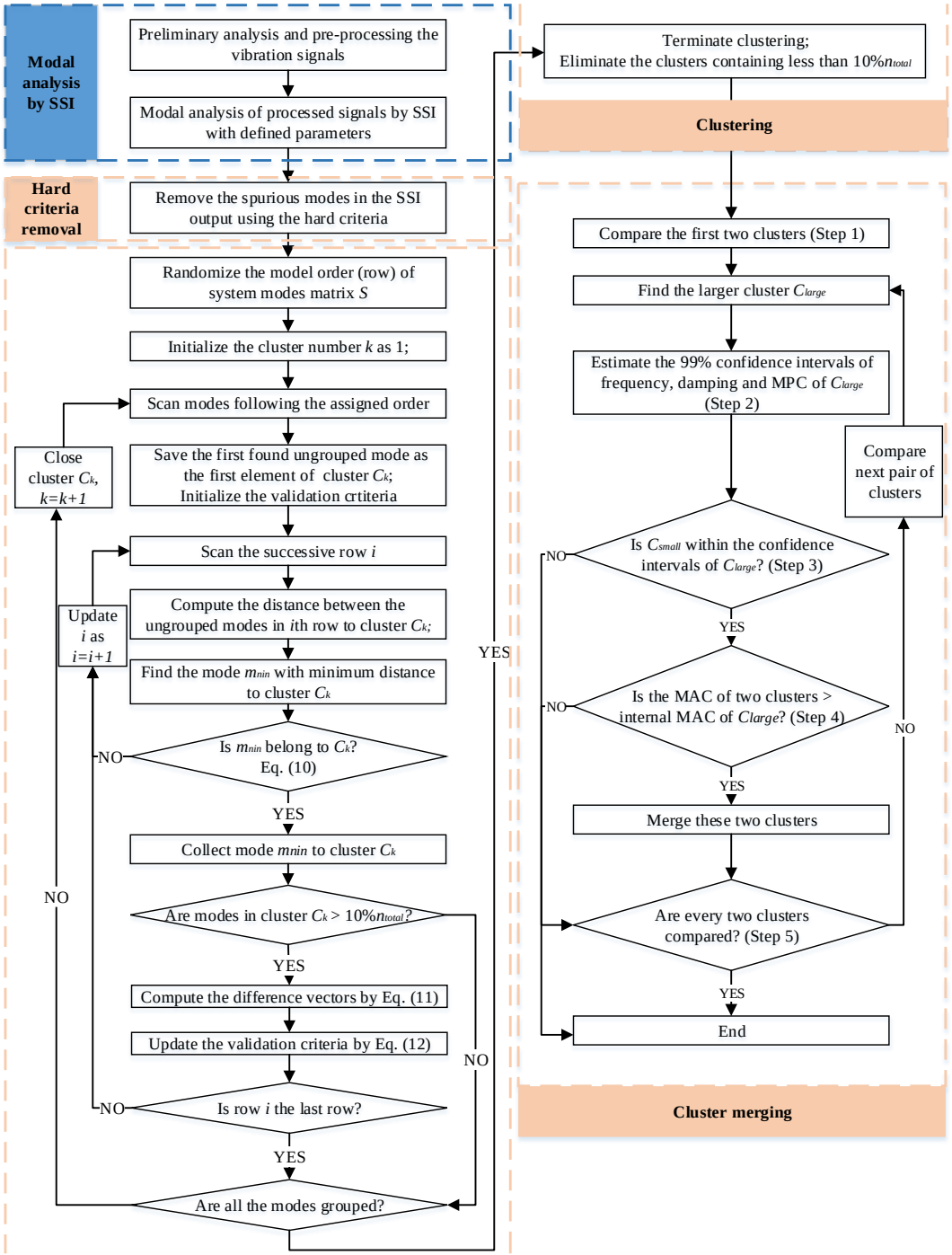


Figure 2-2. Flowchart of the proposed method with details of the clustering and cluster merging.

After performing the statistics-based clustering, clusters containing elements less than 10% of the total model order used n_{total} will be considered as spurious modes that rarely exist in the system modes matrix and will be removed. Setting 10% as the threshold is also small enough to conservatively keep all the clusters with genuine physical modes. Furthermore, 10% is

chosen as the starting point for updating the validation criteria, which implies that such clusters are assembled without prior knowledge on statistical characteristics of previous grouped modes.

2.2.4.3 Cluster merging

After the above statistics-based clustering procedure, all the modes are grouped into clusters based on their similar patterns. The number of spurious modes is significantly reduced after the elimination of clusters with less than 10% of n_{total}. Nevertheless, there are still two problems remaining among the current clusters. On the one hand, setting a small elimination threshold, i.e. 10% described in Section 2.2.4.2, results in that a few clusters consisting of spurious modes may possibly pass the threshold. These relative stable spurious modes are identified because of the existing periodic electric spikes in signals and some weakly excited local structure modes. On the other hand, some physical modes are not consistent in the damping or MPC values due to the influence of significant noise effect, which may be grouped as two or three relatively small separate clusters (but each of them is still large than the eliminating criterion). A possible way to avoid the separation of physical-mode related clusters is enlarging or cancelling the threshold of damping ratio or MPC when distinguishing the modes. However, it is found that the damping ratios and MPC for spurious-mode related clusters are more scattered than that of physical-mode related clusters, which will be shown in the subsequent data analysis results. In this case, setting the threshold values for damping ratio and MPC is essential for distinguishing the relatively stable spurious-mode clusters that are consistent in frequency and mode shape. Instead of merging the physical modes while tolerating more spurious-mode clusters by using compromised thresholds, a better solution for overcoming these two defects is to introduce a cluster merging step following the statistics-based clustering as described in Section 2.2.4.2. The cluster merging procedure will further merge the physical-mode clusters as lager clusters and eliminate the spurious-mode clusters.

The cluster merging procedure is developed based on the 99% confidence interval of frequency, damping ratio and MPC of each cluster. Consider that the modes in a large cluster are samples of a specific structural mode identified at the model orders n_{min} to n_{max}, the 99%

confidence interval provides an estimated range of values that the population of statistical parameters of a specific mode identified at the model order from n_{min} to n_{max} is most likely included. The cluster merging procedure is realized by the following steps

Step 1. Compare every two clusters and find out the larger cluster C_{large} with more elements.

Step 2. Estimate the 99% confidence interval for each criterion including frequency, damping and MPC of the larger cluster C_{large}. The confidence interval of each criterion involves a mean interval $[\mu_{lower}, \mu_{upper}]$.

Step 3. Compute the mean value of modal parameters and MPC for the small cluster C_{small}, if the mean frequency of the smaller cluster C_{small} and one of the mean damping ratio or MPC are located in the 99% confidence interval of the larger cluster C_{large}, these two clusters are highly possible presenting the same mode.

Step 4. Evaluate the MAC between the mean mode shapes of two clusters. If the MAC between two clusters is large than the mean of the internal MAC of C_{large}, where the internal MAC is the MAC between each mode shape and the mean mode shape of the cluster, merge these two clusters into a single cluster.

Step 5. Continue the merging process until every pair of clusters is compared and processed.

The cluster merging procedure groups the physical-mode clusters into larger clusters and remains the spurious-mode clusters in small sizes. A cutoff threshold at 50% of the total number of modes is used to distinguish the physical-mode clusters with the spurious-mode clusters. Cluster satisfying the threshold means that this mode is identified at more than half of system model orders with consistent modal parameters. This judgment is similar to the selection of physical modes by observing the stabilization diagram where modes with most consistent modal parameters are deemed as genuine physical modes. The most significant contribution of the proposed approach is an automated OMA approach which improves the traditional data-driven SSI based on clustering techniques, and does not require user intervention.

2.2.5 Reference method for comparison

Cabboi et al. (2017) presented an automated procedure for OMA by means of a self-adaptable clustering process, which is taken as a reference method in this study. The proposed approach adopts a similar idea but develops a different clustering strategy including statistics-based clustering and cluster merging procedures in the framework. Totally different criteria are applied for clustering in this study, which are defined carefully. The reference method updated the threshold values used for clustering by considering the sum of median distances of natural frequency and mode shape. Using the sum of median distances may have a limited physical meaning. The proposed approach groups the modes and updates the clustering criteria based on the statistical characteristics of modal parameters. The threshold value of each criterion is considered separately to ensure that it has a clearer physical meaning. In terms of the reference method, the distances between the current cluster and modes at the successive model order are computed utilizing Equation (2.9). A pre-defined distance threshold d_{max} is selected as recommended in (Cabboi et al., 2017) and used for selecting modes when the current cluster contains less than three elements. Among the modes that are satisfied by the threshold d_{max}, the one with the minimum distance to the cluster will be clustered. It is noted that the initial distance threshold d_{max} is defined as 0.02 in this study. Once the cluster collects more than two elements, the threshold will be updated as

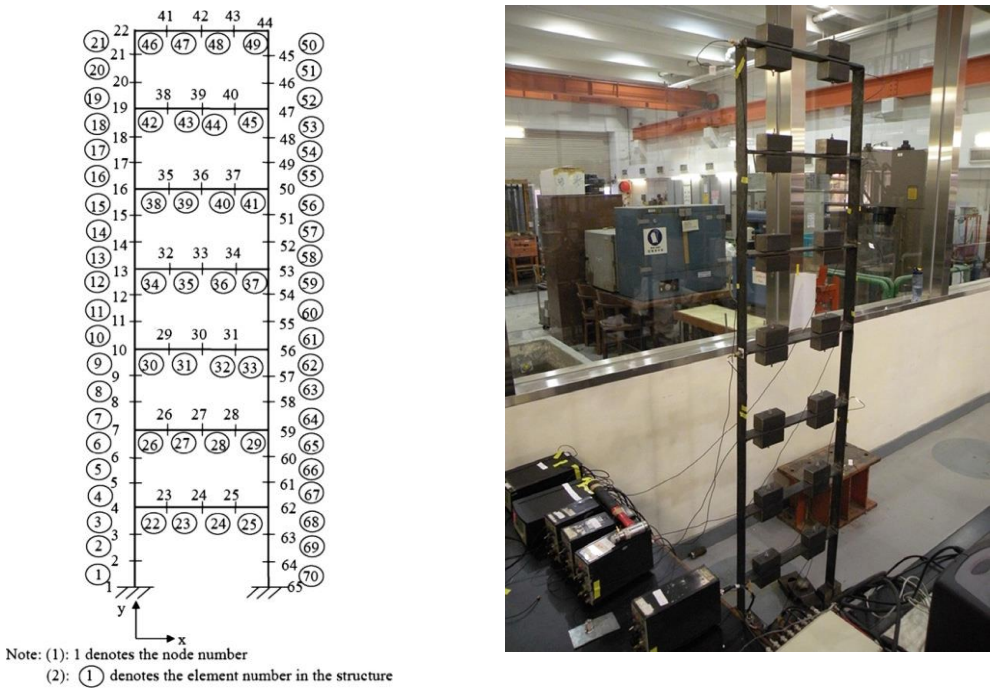
$$d_{max} = \sqrt{median(\left|d_{n,i} - median(d_{n,i})\right|)} \tag{2.13}$$

where $d_{n,i}$ is a distance vector containing the distance of each collected mode with the mean frequency and mode shape in the current cluster. Finally, the clusters containing less than one third of the total model order will be deleted, and the remaining clusters are considered as physical-mode clusters that represent the genuine structural vibration modes.

2.3 Numerical validations

Numerical studies on a seven-storey steel frame structure model are conducted to validate the accuracy and robustness of the proposed approach to identify the vibration characteristics of structures under ambient excitations. The finite element (FE) model of the frame structure

is shown in Figure 2-3(a), which is modelled based on an experimental frame as shown in Figure 2-3(b) by using two-node planar beam elements. Each node in the FE model consists of three DOFs, which are the translational displacements in the *x* and *y* directions and the rotational displacement in the *x-y* plane. The height and width of the frame structure model are 2.1m and 0.5m respectively, with 0.3m height in each floor. The cross sections of the column and beam elements are defined as 50mm × 5mm and 50mm × 9 mm, respectively. The mass density and elastic modulus of material are 7850kg/m^3 and 210GPa, respectively. A 4 kg lumped mass is added on the quarter and three-quarter locations of each floor to simulate the mass in the building. The FE model of the frame structure includes 70 planar elements and 65 nodes. With 3 DOFs at each node, the model has 195 DOFs in total. Two fixed supports at node 1 and 65 are modelled by restraining the corresponding DOFs.



(a) (b)

Figure 2-3. (a) The finite element model of the frame structure in the numerical study. (b) The reference laboratory frame.

Ambient excitation is simulated by applying a white noise with zero mean and 0.005m/s^2 standard deviation at both supports simultaneously in the *x* direction. Vibration responses of the numerical model are computed using the Newmark-beta method. The sampling frequency

is set as 500Hz, and 100 seconds vibration data are obtained. Acceleration responses at nodes 4, 7, 10, 13, 16, 19 and 22 in the *x*-direction are 'measured' for the modal analysis because they are the most representative nodes for the mode shape identification. The noise is added into the measured response $'\text{signal}_m'$ as follows

$$\text{signal}_{noise} = \text{signal}_m + N_p * Noise \tag{2.14}$$

where $'\text{signal}_{noise}'$ is the noisy response, N_p is the noise level of the measured signal, and $Noise$ is a random vector simulating the measurement noise effect, which can be assumed as a normally distributed random vector with zero mean and unit standard deviation of the vibration signal (Li & Law, 2009). In this study, 20% noise is added into the analytical acceleration responses to simulate the noisy responses with a low signal-to-noise ratio. These noisy responses will be used for modal identification and investigating the performance of the proposed approach against the existing method. Figure 2-4 shows the simulated ground excitation in the *x*-direction and acceleration response of node 22 with the noise smeared into the calculate response history.

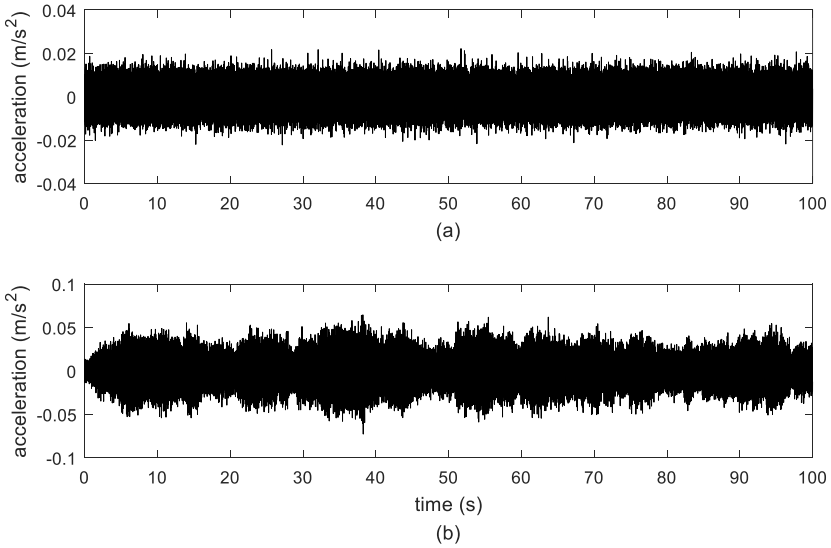


Figure 2-4. Applied ambient excitation and noisy response: (a) Simulated ambient excitation; (b) 'Measured' acceleration response at Node 22.

The vibration characteristics of the frame structure are calculated by performing the eigenvalue analysis based on the FE model, and confirmed with experimental data from

hammer impact vibration tests (Li et al., 2012). The first seven natural frequencies are obtained as 2.54Hz, 7.66Hz, 12.86Hz, 18.03Hz, 22.96Hz, 26.99Hz and 29.91Hz, respectively. It should be highlighted that the identified modal parameters from the previous study are not served as the available information for the automated OMA in this study but only as a baseline for the validation and comparison purposes. Fast Fourier Transform (FFT) is adopted to roughly understand the frequency range of excited modes in the measured responses. As shown in Figure 2-5, the fundamental frequency is around 2.54Hz and most excited modes are within 30Hz. In the Fourier spectrum, the amplitude of the peak at around 29.9Hz is small and very close to the amplitude of the background noise. Therefore, this mode is considered as a weakly excited mode. The period of the first mode is calculated as 1/2.54=0.39s. As mentioned previously in Equation (2.1), the choice of the minimum value of *i* is significantly influenced by the sampling frequency. Since the interested frequency range is below 30Hz in this study, the vibration response is firstly filtered by a 50Hz low-pass filter and then down sampled to 100Hz to better represent all the frequencies of interest and remove the high frequency noises. Consequently, the parameter i is chosen slightly larger than the criterion as 50, where the criterion is calculated as $i \geq T_1/\Delta t = 0.39/0.01 = 39$. The system model order range is selected from 30 to 120 to prevent the identification of the weakly excited modes from being disturbed and meanwhile without introducing too many spurious modes. This frame model is not a complicated system so that less weakly excited modes could be expected, and its dynamic behaviour is linear. For numerical simulation or laboratory tests with minor synchronization issue, using λ =0.3 as the threshold for MPD and 0.7 (1- λ) for MPC is more than sufficient.

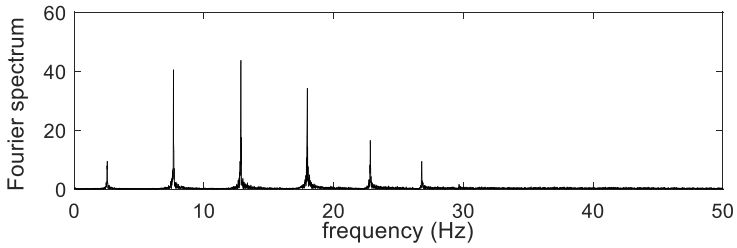


Figure 2-5. Fourier spectrum of a simulated vibration response at node 4 in the x-direction.

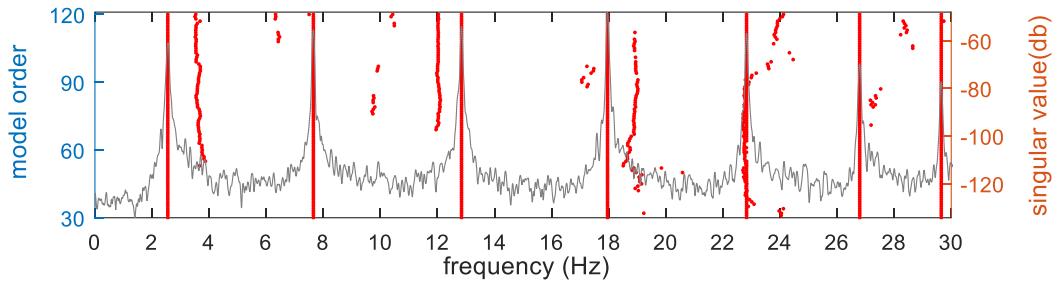
Figure 2-6(a) shows the remaining modes after removing certain spurious modes by using the hard validation criteria as shown in Equation (2.10). Through observing the stabilization diagram, several physical modes can be clearly identified because they are consistent in all the model orders. While some modes are unclear which are possibly physical or spurious modes. The results in the stabilization diagram are compared with another widely used OMA technique, namely FDD (Brincker et al., 2001), with the first singular values of spectral matrix drawn behind the stabilization diagram. In the results from FDD, the peaks observed in the spectrum most likely represent the nature frequencies of the structure. Corresponding to the SSI output, the first singular values consist of high peaks at the frequencies of the certain physical modes, and also several weak peaks which may be the weakly excited modes or due to the introduced significant noise that confuses the identification. Analysing the system modes matrix by using the statistics-based clustering procedure, Figure 2-6(b) shows the grouped clusters containing elements more than 10% of the total model order. It should be noted that modes belonging to different clusters are illustrated in different colours. Most of the spurious modes are eliminated and the physical modes representing the same mode are separated as two or three clusters due to the severe criteria and noise. Compared to the results as shown in Figure 2-6(a), it should be highlighted that the physical modes in Figure 2-6(b) at different model orders are not as consecutive as those in Figure 2-6(a). This means that the outliers of the physical modes are rejected by the physical-mode clusters, which result in a more accurate identification of modal parameters. Figure 2-6(c) shows the final remaining clusters by implementing the cluster merging procedure based on the results from the statistics-based clustering. Totally seven distinct clusters represented by different colours correspond to the first seven physical modes of the frame model, which are denoted as P1 to P7 and demonstrated in the final modal identification results of using the proposed approach, as shown in Figure 2-6(c). The mean frequencies of each cluster are listed in Table 2-1, and compared with the results obtained from the reference method (Cabboi et al., 2017) as shown in Figure 2-6(d). The reference method produces the same identification results of the first seven modes compared with those from the proposed approach, except a spurious mode at 12.03Hz denoted as S1, which is not effectively removed. This spurious mode has a relative consistent frequency

across a number of model orders, as shown in Figure 2-6(d). By using the reference method, this may be taken as a large cluster and wrongly identified as a mode. As demonstrated in Table 2-1, the proposed approach accurately identifies the first seven modes of the frame model without any spurious modes. The maximum difference of the frequency is observed at the 7th mode within a 1% relative error. Considering the significant noise effect at 20%, the results demonstrate that the proposed approach has a robust and accurate performance in the modal identification of structures under ambient excitations.

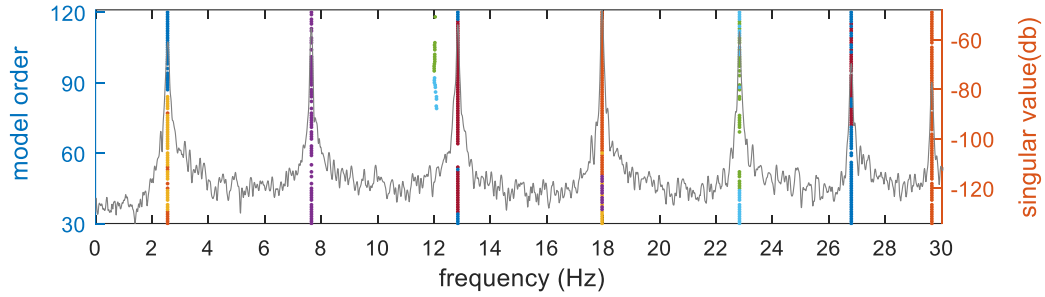
To further demonstrate how the proposed approach improves the results based on the clustering techniques, Figure 2-7 shows the eight grouped clusters regarding their number of modes and corresponding mean frequencies when implementing the cluster merging procedure. A red dash line lies on the 50% of the total number of orders represents the threshold for spurious-mode clusters removal. It can be seen that all the physical-mode clusters representing the same mode are merged as a larger cluster, meanwhile, two spurious-mode clusters at 12.03Hz are also merged as a single cluster. The reason for the merging of this two clusters can be concluded as the discrete modes in varying model orders representing this mode are rejected, the collected modes in these two clusters are relatively stable. Even though, this merged cluster is still deleted since it contains a small number of stable modes which is far from the defined 50% threshold.

In summary, the main differences in modal identification results between the proposed approach and the reference methods can be concluded as

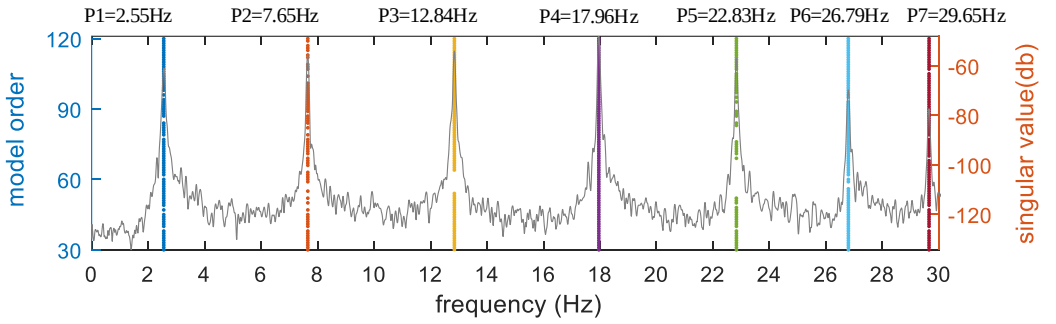
1) The proposed approach considers more physically reasonable criteria to remove the spurious modes in a comprehensive manner; and
2) In Equation (2.13), the square root of the median difference value may increase the threshold to a considerable value when the collected elements are discrete. For this numerical study, the threshold of the spurious-mode related cluster (S1) has been enlarged to a maximum value 8.1%, where for most of the physical-mode clusters the threshold is only below 1%. In contrast, the proposed approach using Equation (2.12) can provide more reasonable threshold values based on physical meanings.



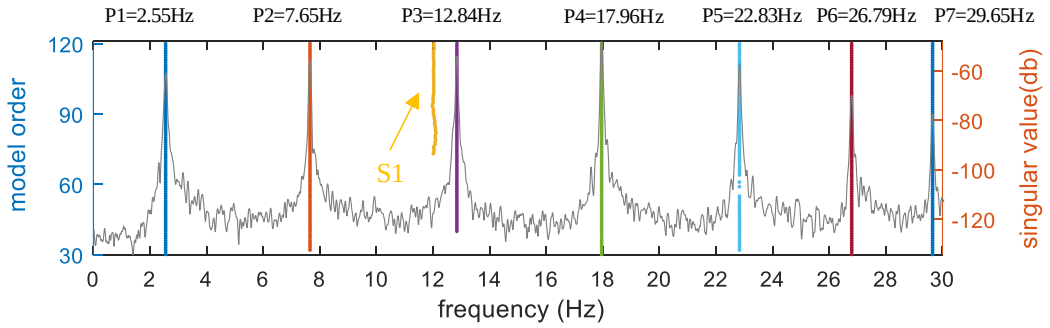
(a)



(b)



(c)



(d)

Figure 2-6. Comparison of modal identification results: (a) Hard validation criteria; (b) After statistics-based clustering; (c) After cluster merging; (d) Using Reference method.

Table 2-1. Comparison of the identified nature frequencies of the numerical model from the proposed approach and reference method.

Mode	Baseline (Li et al., 2012)	The proposed approach	Relative error (%)	Reference method (Cabboi et al., 2017)	Relative error (%)
1st (P1)	2.54	2.55	0.39	2.55	0.39
2nd (P2)	7.66	7.65	0.13	7.65	0.13
(S1)	-	-	-	12.03	-
3rd (P3)	12.86	12.84	0.16	12.84	0.16
4th (P4)	18.03	17.96	0.39	17.96	0.39
5th (P5)	22.96	22.83	0.57	22.83	0.57
6th (P6)	26.99	26.79	0.74	26.79	0.74
7th (P7)	29.91	29.65	0.87	29.65	0.87

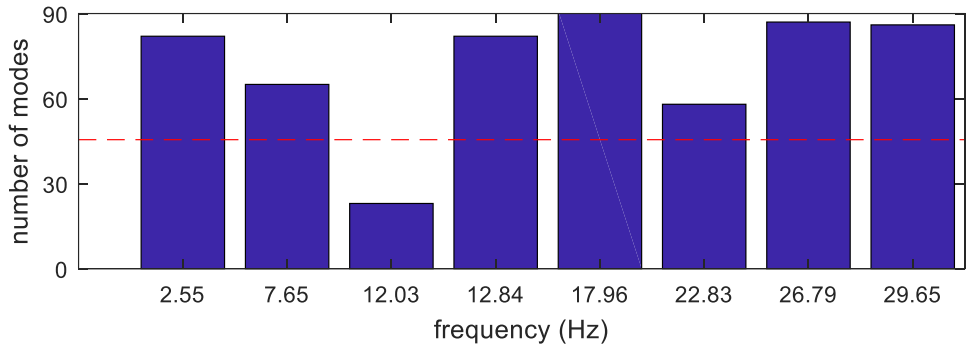


Figure 2-7. The number of modes for each cluster in the numerical study.

2.4 Experimental validations

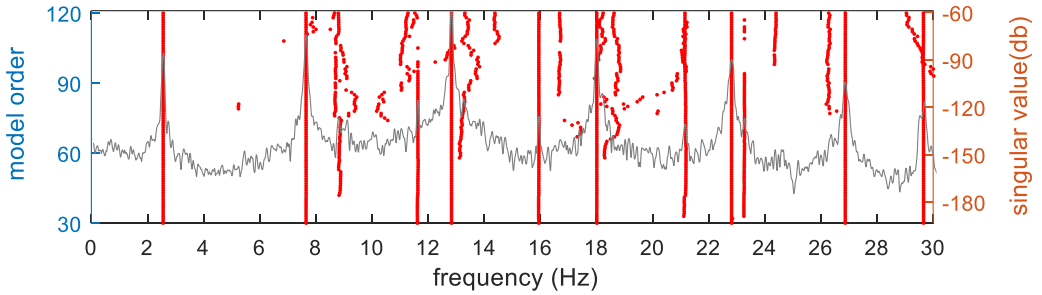
2.4.1 Validations on a laboratory frame structure

Experimental validations are conducted on a laboratory steel frame structure under ambient excitations as shown in Figure 2-3(b) to verify the performance of using the proposed approach for operational modal identification. The frame structure has the same dimensions, material properties and boundary conditions as presented in the numerical study in Section 2.3. Seven accelerometers are installed on the left side of beam-column joints of each floor to measure the lateral vibration responses of the frame structure under ambient excitations. National Instruments data acquisition system is used to record the acceleration responses with a sampling frequency of 2000Hz for a period of 10 minutes.

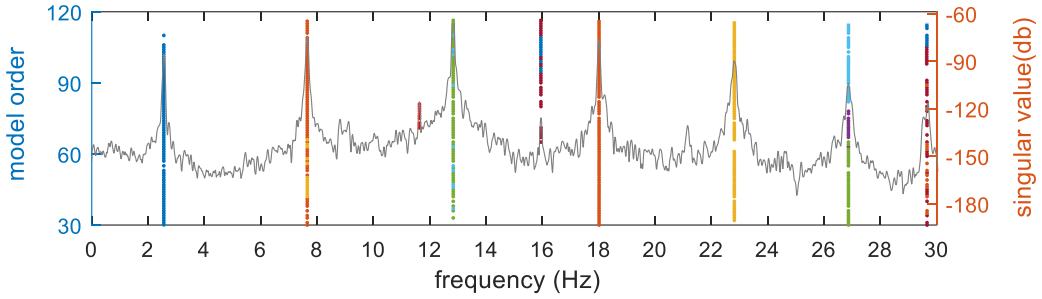
The same model parameters associated with SSI that selected in the numerical validations are used for analysing the experimental data measured from this laboratory frame structure. The procedures as described in Sections 2.2.3 and 2.2.4 are followed to perform the clustering and identify the vibration characteristics. Figure 2-8(a) shows the stabilization diagram from SSI after the hard validation criteria removal. It can be seen that a massive number of spurious modes are still mixed in the system modes matrix. However, compared to the original system mode matrix, this step decreases in fact more than half of the total number of modes, which enhanced the computational efficiency of the proposed approach more than 4 (2^2) times. By conducting the statistics-based clustering process, the obvious spurious modes are grouped in small clusters and effectively eliminated by using the 10% threshold. It should be noted that the relatively stable spurious modes are either removed with the 10% threshold such as modes at frequencies 8.81Hz, 21.18Hz and 23.27Hz, or separately collected into bigger clusters, such as modes at frequencies 11.64Hz and 15.95Hz, as shown in Figure 2-8(b). On the other hand, some weakly excited modes or noise disturbed physical modes are also collected into two or three clusters such as modes at 7.65Hz and 29.65Hz. Figure 2-8(c) shows the final results after the cluster merging procedure. All the first seven true physical-mode related clusters are merged together, while the spurious-mode clusters remain their original sizes in this step. This means that the spurious-mode clusters representing the same mode have larger discrepancies in one or more modal parameters than physical-mode clusters. Comparing Figures 2-8(a) and (c), even some relatively stable spurious modes such as modes at frequency 8.81Hz, 11.64Hz, 15.95Hz and 21.18Hz are effectively removed in the final identification results. Such modes may be stable in one or two modal parameters. When all the four validation criteria are used for clustering, they are separated into several small clusters and removed by the proposed clustering techniques. Figure 2-9 shows the corresponding number of identified mode in each cluster. The defined threshold can be used to successfully remove the spurious modes after the clustering merging.

The results obtained from the reference method are shown in Figure 2-8(d). Comparing the results from the proposed approach and the reference methods as shown in Figures 2-8(c) and (d) respectively, both methods produce accurate identification results on the first seven

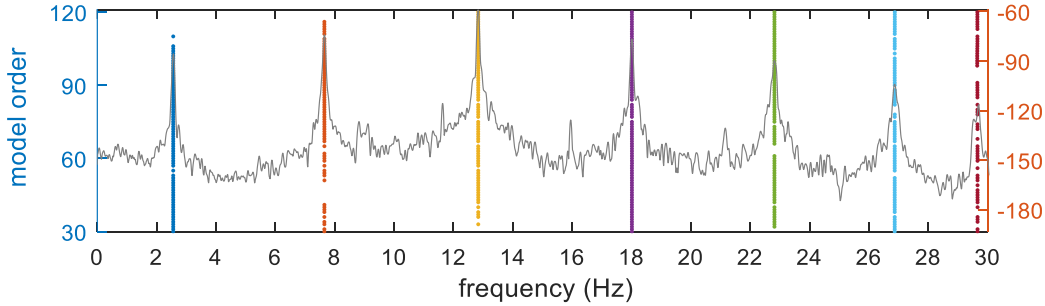
modes. The identified frequencies are almost the same except the 6th mode with a slightly better result from the proposed approach. However, when examining all the identification modes from the reference method, there are four spurious modes (S1 to S4) with comparatively consistent frequencies and mode shapes, which are not eliminated. This maybe because of the incomprehensive consideration of setting thresholds for validation criteria. The proposed approach can improve the performance by using the statistics-based clustering with reasonable criteria in grouping and merging the clusters based on physical meaning and statistics from the identified vibration properties. This will support to generate better identification results without spurious modes in this study as shown in Figure 2-8(c). The accuracy of these identification results can be further demonstrated in Figures 2-10(a) and (b), where the relationships between the damping ratios and MPCs with respect to frequencies of all the remaining modes after the statistics-based clustering procedure are shown respectively. Spurious-mode clusters representing the same mode have consistent frequency but discrete damping ratios and MPC values. For example, the damping ratios of two spurious-mode clusters at frequency 15.95Hz are separated with a clear gap. The physical-mode clusters are consistent with all the modal parameters even under noise effect. Seven large clusters are finally reserved in the system modes matrix after the two clustering procedures. The reserved seven clusters contain modes corresponding to the first seven modes of the frame structure respectively. The most representative natural frequency of each cluster is compared with that from a previous study (Li et al., 2012) and the reference method (Cabboi et al., 2017), as listed in Table 2-2. It should be noted that the responses measured under ambient excitations usually have a relatively high noise effect, which could be more challenging for operational modal identification compared with that using forced vibration test data. It can be observed from Table 2-2 that natural frequencies are identified accurately. Minor differences are observed since 1) Different data analysis methodologies are used; and 2) The test conditions for recorded data used in the previous study (Li et al., 2012) and this study are different. The discrepancies in frequencies may be caused by the environmental effect such as temperature and humidity variations. The identified seven mode shapes are shown in Figure 2-11. These modes represent strong linear behaviours leading to very tiny MPC values, as shown in Figure 2-10(b).



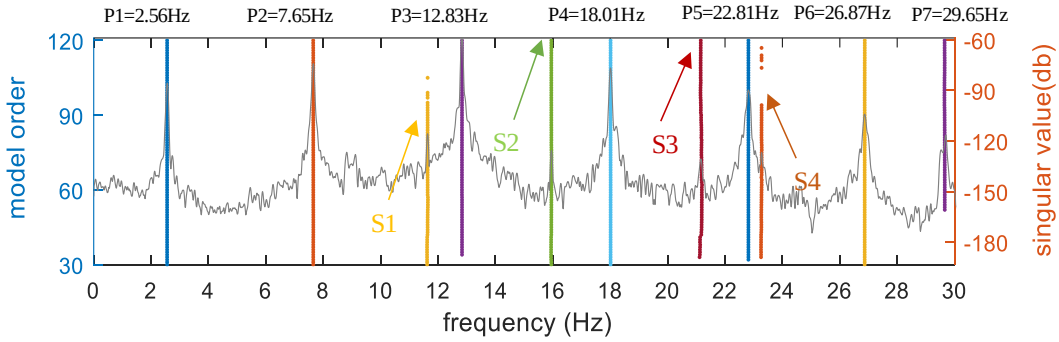
(a)



(b)



(c)



(d)

Figure 2-8. Comparison of modal identification results on the laboratory steel frame: (a) Hard validation criteria; (b) After statistics-based clustering; (c) After cluster merging; (d) Using Reference method.

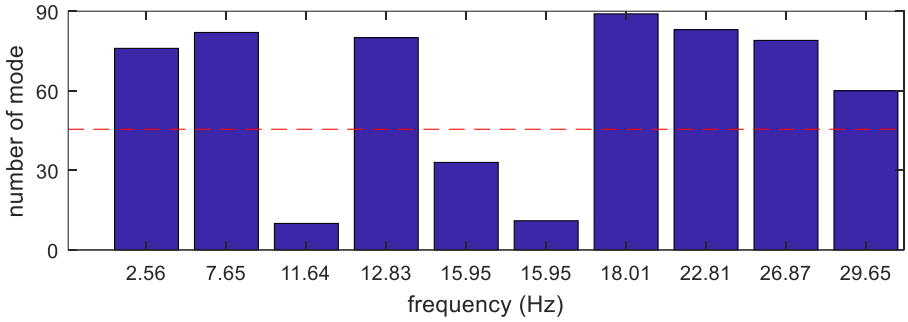
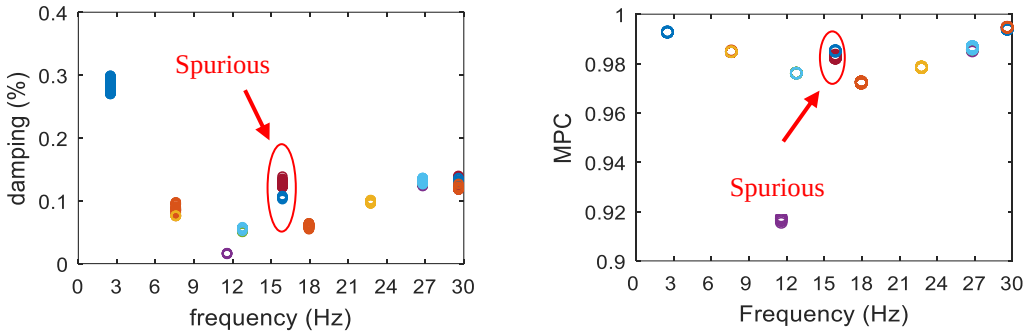


Figure 2-9. The number of modes for each cluster in the experimental study.



(a) (b)

Figure 2-10. Relationships among various variables: (a) Damping ratio vs frequency; (b) MPC vs frequency.

Table 2-2. Comparison of the identified nature frequencies of the laboratory frame structure from the proposed approach and the reference method.

Mode	Baseline (Li et al., 2012)	The proposed approach	Relative error (%)	The reference method (Cabboi et al., 2017)	Relative error (%)
1st (P1)	2.54	2.56	0.79%	2.56	0.79%
2nd (P2)	7.66	7.65	0.13%	7.65	0.13%
(S1)	-	-		11.64	
3rd (P3)	12.86	12.83	0.23%	12.83	0.23%
(S2)	-	-		15.95	
4th (P4)	18.03	18.01	0.11%	18.01	0.11%
(S3)	-	-		21.15	
5th (P5)	22.96	22.81	0.65%	22.81	0.65%
(S4)	-	-		23.27	
6th (P6)	26.99	26.87	0.39%	26.87	0.44%
7th (P7)	29.91	29.65	0.87%	29.65	0.87%

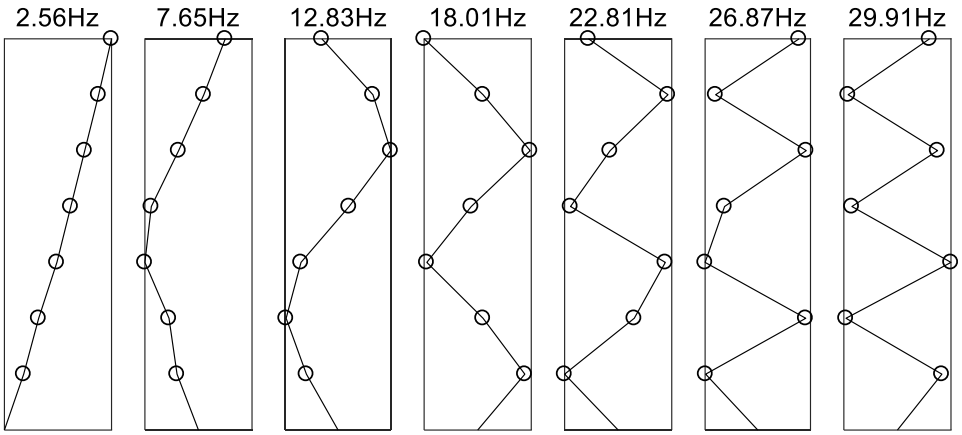


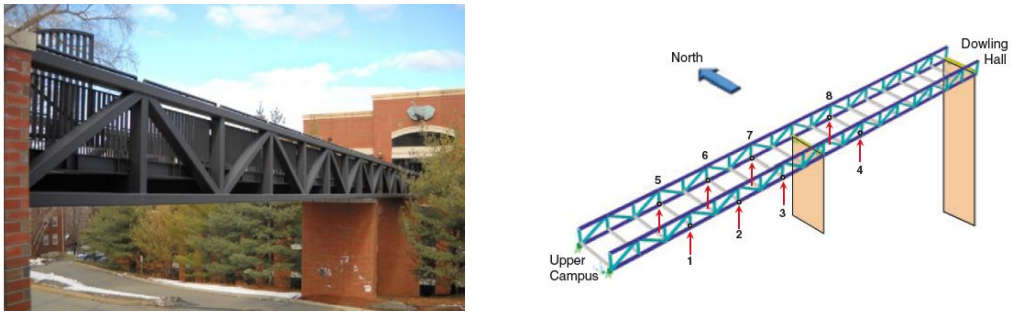
Figure 2-11. Identified mode shapes of the first seven modes.

2.4.2 Validations on Dowling Hall footbridge

Dowling Hall footbridge is located on the Medford, MA, campus of Tufts University. It serves as the pedestrian bridge connecting the main campus and the student services offices. As shown in Figure 2-12(a), this bridge is a two-span continuous steel frame bridge. Dowling Hall footbridge is 44m long and 3.7m wide. A long-term structural health monitoring system has been installed to perform the continuous vibration monitoring. The sensory system layout on the bridge is shown in Figure 2-12(b). Vibration responses from 8 accelerometers were measured over 17 weeks from January 2010 to May 2010. The detailed introduction of the bridge and the structural health monitoring system can be found in Ref. (Moser & Moaveni, 2013). The first six nature frequencies and mode shapes of the footbridge identified in the literature (Moaveni & Behmanesh, 2012; Moser & Moaveni, 2011) will be used as a reference for assessing the accuracy and performance of the proposed approach.

The acceleration data in the vertical direction from the Dowling Hall footbridge under the ambient excitations recorded on 4th January 2010 are used for the analysis. From the Fourier spectrum of the acceleration data as shown in Figure 2-13, the fundamental frequency is around 4.6Hz and the frequency range of interest is within 30Hz. The original sampling rate of the acceleration data is 128Hz, which is only slightly higher than the interested frequency range. Consequently, no down-sampling process is conducted in order to keep the data integrity. The period for the first mode is 1/4.6=0.22s so that the parameter i is chosen as 40 to satisfy Equation (2.3) where $i \geq T_1/\Delta t$ =0.22/0.008=28. The system model order range is

selected from 40 to 140 for the same purpose of remaining the weakly excited mode without compromising the number of spurious modes. Since the test is conducted on a real structure with long signal transmission cables, synchronization issue could be expected. Meanwhile, weakly excited modes may also be observed in this case because of the low level ambient excitations, the threshold values for MPD and MPC are chosen as 0.7 and 0.3 respectively to retain all the physical modes.



(a) (b)

Figure 2-12. (a) South view of Dowling Hall Footbridge; (b) Sensor layout.

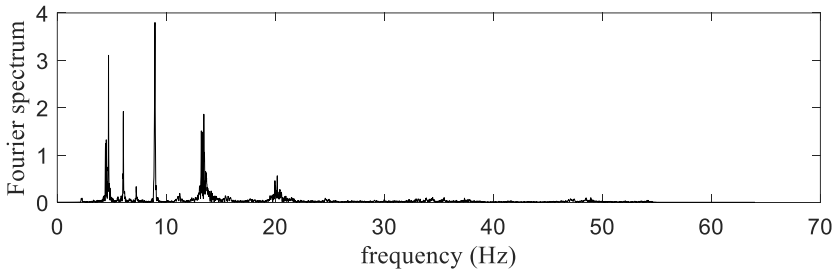


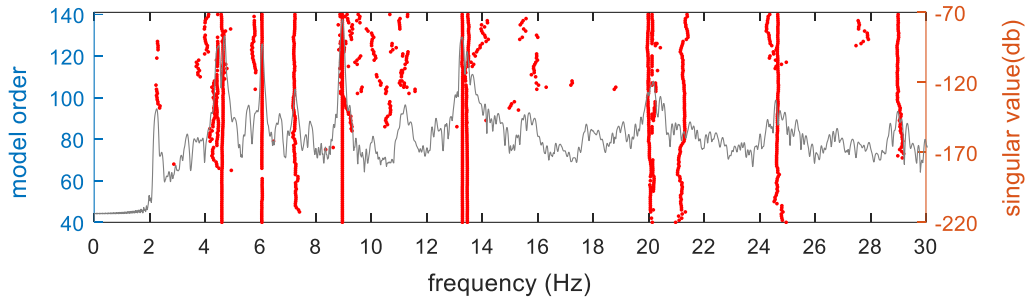
Figure 2-13. Fourier spectrum of a vibration signal from the footbridge.

The proposed appraoch described in Section 2.2.4 is used to analyse the testing data. Figure 2-14(a) shows all the remaining identified modes in the system modes matrix within 30Hz after the hard criteria removal. FDD is also applied to analyse the same testing data, and its spectrum is drawn behind the stabilization diagram. It is observed that FDD fails to identifiy the modal information effectively since the spectrum contains numerous peaks, which make identifying the the real modes diffiuclt. It can be observed from the stabilization diagram that 13.28Hz and 13.46Hz are two very closely spaced modes. For OMA of civil engineering structures, identifying the closely spaced modes and distinguishing them with spurious modes

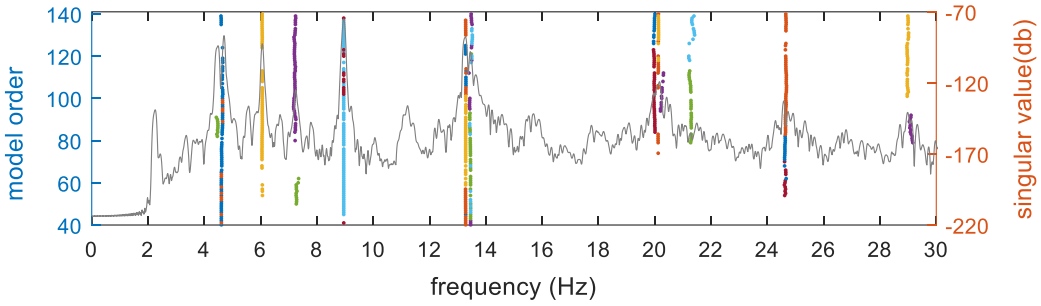
is always a serious challenge. Figures 2-14(b) and (c) show the modal identification results after conducting the statistics-based clustering and cluster merging procedures. It can be seen that the statistics-based clustering procedure greatly decreases the number of spurious modes, and the following cluster merging procedure groups the physical-mode clusters and cleans the stabilization diagram with only eight remaining stable modes. Only the first six modes within 15Hz are reported in a previous study (Moser & Moaveni, 2011) and there is no reference for the rest two modes around 20Hz and 24.5Hz. Therefore in this study, the identified natural frequencies of the first six modes, which are marked as P1 to P6 in Figure 2-14(c), are compared and listed in Table 2-3. The proposed apporach performs the modal parameters identification slightly better than the reference method for most of those six modes. It is worth mentioning that modes at 13.18Hz (P5) and 13.46 Hz (P6) are closely spaced modes. The reason for this two modes are well seperated is that except the frequency, other vibration characteristics including damping ratio, MPC and mode shape are dissimilar as shown in Figures 2-15 and 2-16. The proposed clustering and cluster merging procedure clearly distiguished these two modes based on the comprehensivie comsideration of the vibration characteristics.

Table 2-3. Comparison of the identified nature frequencies of the footbridge from the proposed approach and the reference method.

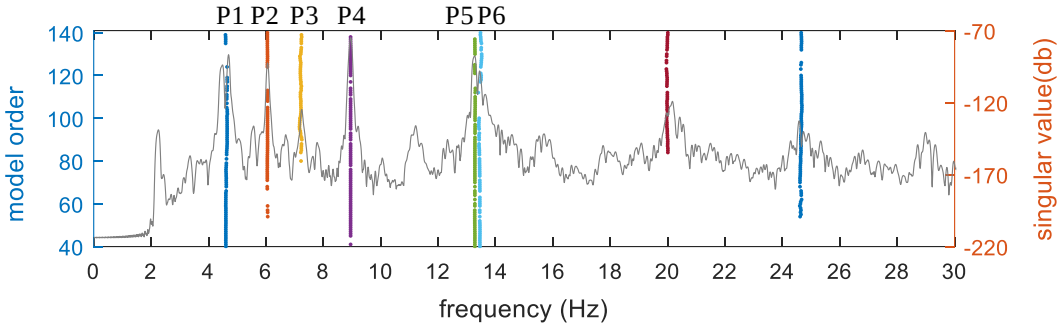
Mode	**Baseline (Moser & Moaveni, 2011)**	**The proposed approach**	**Relative error (%)**	**The reference method (Cabboi et al., 2017)**	**Relative error (%)**
(S1)	-	-		4.44	
1st (P1)	4.66	4.62	0.76%	4.63	0.69%
2nd (P2)	6.21	6.05	2.50%	6.04	2.72%
3rd (P3)	7.08	7.21	1.90%	7.23	2.08%
4th (P4)	8.88	8.95	0.81%	8.95	0.81%
5th (P5)	13.23	13.28	0.41%	13.28	0.41%
6th (P6)	13.62	13.46	1.17%	13.45	1.23%



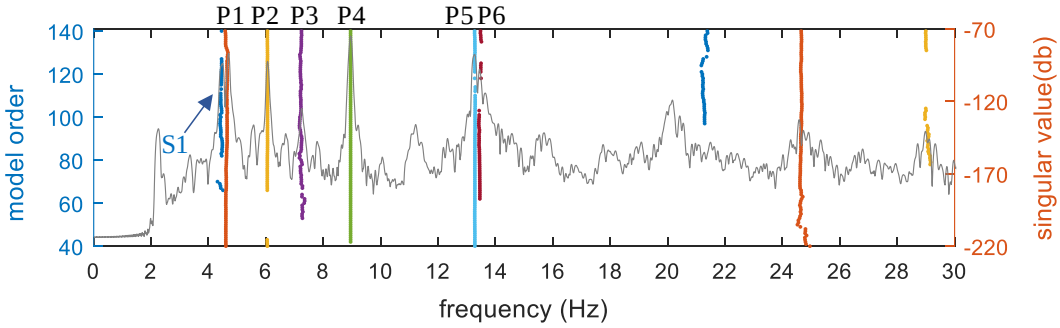
(a)



(b)

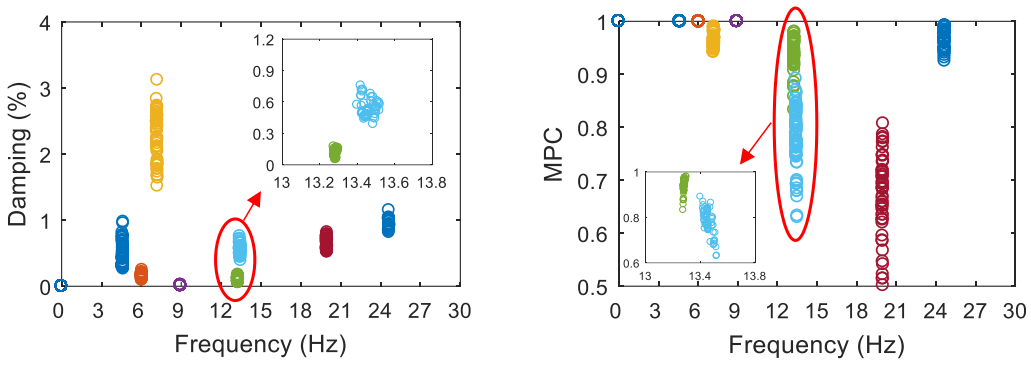


(c)



(d)

Figure 2-14. Modal identification results of Dowling Hall footbridge: (a) Hard validation criteria; (b) After statistics-based clustering; (c) After cluster merging; (d) Using Reference method.



(a) (b)

Figure 2-15: (a) Damping ratio vs frequency. (b) MPC vs frequency.

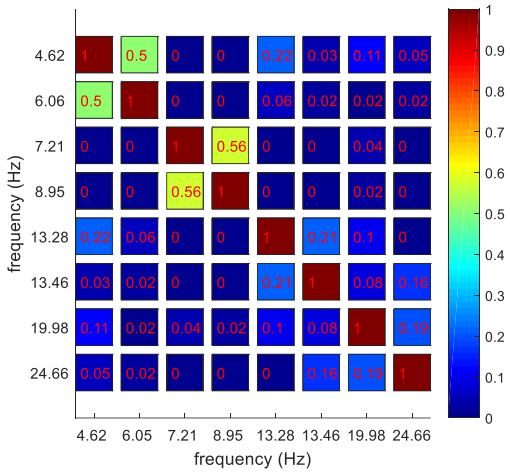


Figure 2-16. MAC value between each two of the final remained clusters.

The results from the reference method are shown in Figure 2-14(d). It should be noted that the reference method fails to eliminate a spurious mode at 4.44Hz, which is denoted as S1. This may be because it is very close to the first natural frequency with a relatively high peak in FDD spectrum and a realtively stable and consistent apperance across a number of model orders in the stabilization diagram. This is a significant issue in most OMA methods to distiguish the genuine physical modes and spurious modes, particualry when the modes are closely spaced. The maximum difference between the results in the literature and identified frequencies is observed in the second mode with a 2.5% relative error. This discrepancy could be explained that these two tests were conducted in different seasons. In Ref. (Moser & Moaveni, 2011), the measurement data were collected in April 2009. In this study, the data used for modal analysis is from January 2010. Accounting for the temperature and humidity variations, the discrepancies in the frequencies could be expected. Overall, the proposed

approach can successfully identify the first six vibration modes of this footbridge, which are shown in Figure 2-17. The mode shapes for the second span are different in the higher order modes because the current study only uses four available long-term monitoring sensors on each side of the footbridge. While in Ref. (Moser & Moaveni, 2011), the data are measured from a short period test with six sensors on each side.

To investigate whether the proposed approach can be further applied for analyzing long-term monitoring data, also known as modal tracking, a study with measured data sets collected at every 12:00 pm for 7 days is conducted. Same model parameters of the SSI as the above experimental validations are selected for the modal identification. The tracked natural frequencies under 15Hz are shown in Figure 2-18. It can be observed that all these six physiscal modes are successfully tracked and identified without any spurious modes. The results demonstrate that the proposed approach is capable of analysing different data sets measured from structures under varying testing conditions without any manual intervention and tunning. This indciates that when massive vibration monitoring data are collected, the proposed approach is robust and suitable for the automated OMA and modal tracking.

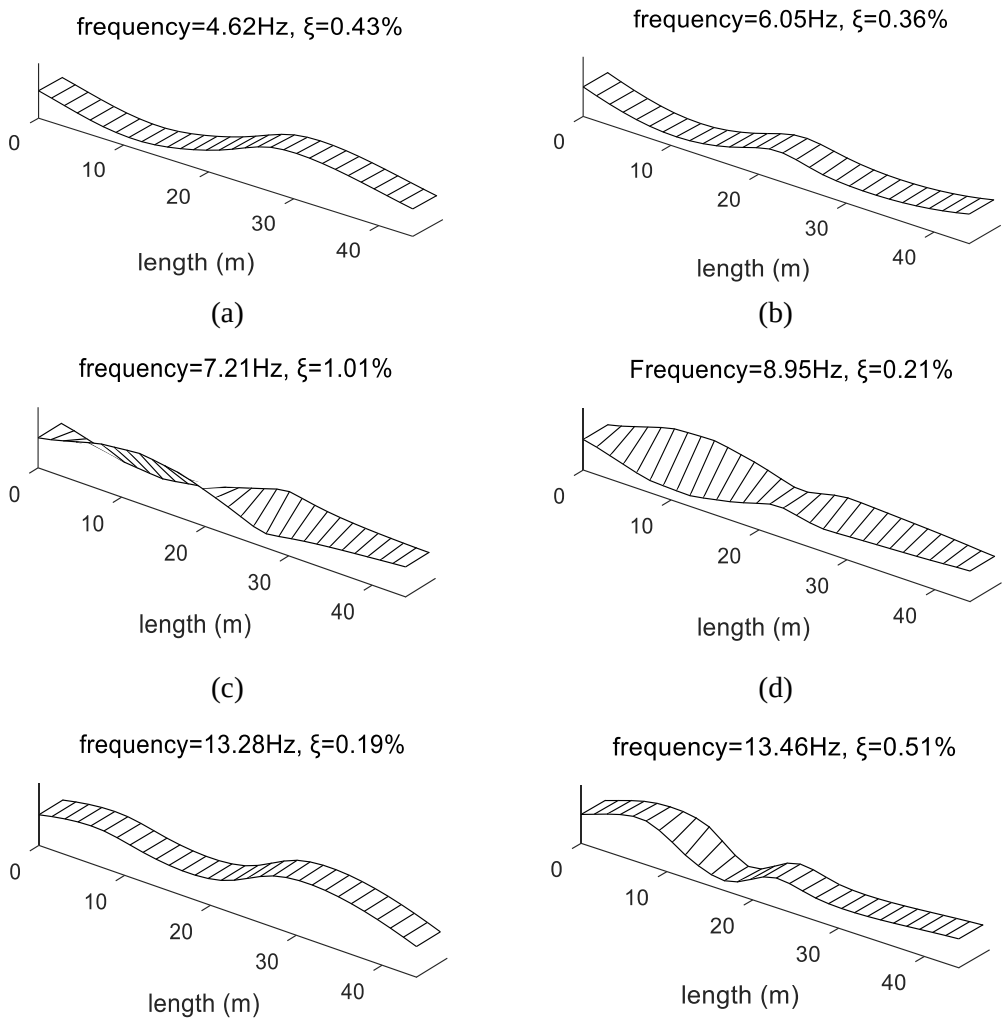


Figure 2-17. Identified vibration mode shapes of the Dowling Hall footbridge: (a) Mode 1; (b) Mode 2; (c) Mode 3; (d) Mode 4; (e) Mode 5; (f) Mode 6.

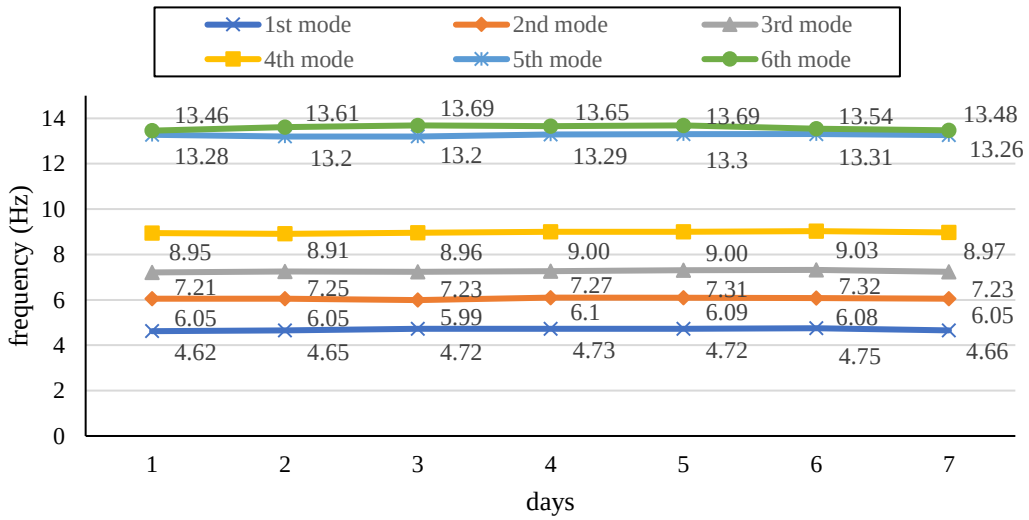


Figure 2-18. Identified natural frequencies of the first six modes using the first week's measured acceleration data.

2.5 Conclusions and discussions

This chapter proposes an improved automated OMA approach based on data-driven SSI and clustering techniques. The framework of the proposed approach consists of two main components, namely Component 1: Modal identification by SSI and Component 2: Automated interpretation of SSI output. Data-driven SSI is used in Component 1. Three procedures including hard validation criteria removal, a statistics-based clustering procedure and a cluster merging procedure are involved in Component 2 to automatically interpret the stabilization diagram from the SSI output, without any previous knowledge on the modal parameters and user intervention during process. Original contribution is made in the improvement of the statistic-based clustering and the development of cluster merging to identify the weakly excited modes and remove the false modes effectively. Numerical validations on a frame structure model are conducted, and the results demonstrate that the proposed approach is capable of identifying the vibration modes accurately and resisting a high noise disturbance in the measured signals effectively. Experimental studies on a seven-storey steel frame structure and a real footbridge are conducted to demonstrate the robustness and performance of using the proposed approach for automated OMA. In both the numerical and experimental verifications, the physical modes across varying model orders representing the same mode are collected to a single large cluster. The identified nature frequencies are compared with those in the literature and obtained from the reference method. The results show that the proposed approach can accurately identify the modal parameters of physical modes, and concurrently eliminate all the spurious modes especially the stubborn modes with relative stable modal parameters at some model orders. The weakly excited mode and two closely spaced modes are clearly distinguished in the analysis of the frame and the footbridge.

It should be noted that few parameters including the model orders, MPC/MPD thresholds and the initial criteria are manually tuned or selected based on the understanding of the structural vibration behaviour and the testing conditions. Properly selecting these parameters can improve the identification quality and the computational efficiency. In the proposed approach, the validation criteria threshold of each cluster is considered and updated

individually, which is more closely related to the characteristics of the corresponding cluster based on the grouped modal parameters. Compared with the reference method, the proposed approach provides a more comprehensive clustering strategy to select the physical modes more strictly.

Taking all the numerical and experimental validation results into consideration, the proposed approach exhibits a good robustness and performance for modal analysis of engineering structures under ambient excitations. In addition, the proposed approach is applicable to interpret results from any parametric time domain identification techniques but not limited to SSI method. Identification results from the vibration monitoring data over 7 days demonstrate that the proposed approach can be extended for the modal tracking purpose.

2.6 References

Bakir, P. G. (2011). Automation of the stabilization diagrams for subspace based system identification. *Expert Systems with Applications, 38*(12), 14390-14397.

Boroschek, R. L., & Bilbao, J. A. (2019). Interpretation of stabilization diagrams using density-based clustering algorithm. *Engineering Structures, 178*, 245-257.

Brincker, R., Zhang, L., & Andersen, P. (2001). Modal identification of output-only systems using frequency domain decomposition. *Smart Materials and Structures, 10*(3), 441.

Cabboi, A., Magalhães, F., Gentile, C., & Cunha, Á. (2017). Automated modal identification and tracking: Application to an iron arch bridge. *Structural Control and Health Monitoring, 24*(1), e1854.

Cardoso, R., Cury, A., & Barbosa, F. (2017). A robust methodology for modal parameters estimation applied to SHM. *Mechanical Systems and Signal Processing, 95*, 24-41.

de Almeida Cardoso, R., Cury, A., & Barbosa, F. (2017). A clustering-based strategy for automated structural modal identification. *Structural Health Monitoring: An International Journal*, 147592171668923.

Li, J., Law, S. S., & Ding, Y. (2012). Substructure damage identification based on response reconstruction in frequency domain and model updating. *Engineering Structures, 41*, 270-284.

Li, X. Y., & Law, S. S. (2009). Identification of structural damping in time domain. *Journal of Sound and Vibration, 328*(1-2), 71-84.

Magalhaes, F., Cunha, A., & Caetano, E. (2009). Online automatic identification of the modal parameters of a long span arch bridge. *Mechanical Systems and Signal Processing, 23*(2), 316-329.

Moaveni, B., & Behmanesh, I. (2012). Effects of changing ambient temperature on finite element model updating of the Dowling Hall Footbridge. *Engineering Structures, 43*, 58-68.

Moser, P., & Moaveni, B. (2011). Environmental effects on the identified natural frequencies of the Dowling Hall Footbridge. *Mechanical Systems and Signal Processing, 25*(7), 2336-2357.

Moser, P., & Moaveni, B. (2013). Design and Deployment of a Continuous Monitoring System for the Dowling Hall Footbridge. *Experimental Techniques, 37*(1), 15-26.

Nayeri, R. D., Masri, S. F., Ghanem, R. G., & Nigbor, R. L. (2008). A novel approach for the structural identification and monitoring of a full-scale 17-story building based on ambient vibration measurements. *Smart Materials and Structures, 17*(2), 025006.

Neu, E., Janser, F., Khatibi, A. A., & Orifici, A. C. (2017). Fully Automated Operational Modal Analysis using multi-stage clustering. *Mechanical Systems and Signal Processing, 84, Part A*, 308-323.

Pappa, R. S., Elliott, K. B., & Schenk, A. (1993). Consistent-mode indicator for the eigensystem realization algorithm. *Journal of Guidance, Control, and Dynamics, 16*(5), 852-858.

Peeters, B., & De Roeck, G. (2001). One-year monitoring of the Z24-Bridge: environmental effects versus damage events. *Earthquake engineering & structural dynamics, 30*(2), 149-171.

Priori, C., De Angelis, M., & Betti, R. (2018). On the selection of user-defined parameters in data-driven stochastic subspace identification. *Mechanical Systems and Signal Processing, 100*, 501-523.

Rainieri, C., & Fabbrocino, G. (2010). Automated output-only dynamic identification of civil engineering structures. *Mechanical Systems and Signal Processing, 24*(3), 678-695.

Reynders, E., & De Roeck, G. (2008). Reference-based combined deterministic–stochastic subspace identification for experimental and operational modal analysis. *Mechanical Systems and Signal Processing, 22*(3), 617-637.

Reynders, E., Houbrechts, J., & De Roeck, G. (2012). Fully automated (operational) modal analysis. *Mechanical Systems and Signal Processing, 29*, 228-250.

Sun, M., Makki Alamdari, M., & Kalhori, H. (2017). Automated operational modal analysis of a cable-stayed bridge. *Journal of Bridge Engineering, 22*(12), 05017012.

Teng, J., Tang, D.-H., Zhang, X., Hu, W.-H., Said, S., & Rohrmann, R. (2019). Automated modal analysis for tracking structural change during construction and operation phases. *Sensors, 19*(4), 927.

Ubertini, F., Gentile, C., & Materazzi, A. L. (2013). Automated modal identification in operational conditions and its application to bridges. *Engineering Structures, 46*, 264-278.

Van Overschee, P., & De Moor, B. (2012). *Subspace identification for linear systems: Theory—Implementation—Applications*: Springer Science & Business Media.

Xia, Y., Chen, B., Zhou, X., & Xu, Y. (2013). Field monitoring and numerical analysis of Tsing Ma Suspension Bridge temperature behavior. *Structural Control and Health Monitoring, 20*(4), 560-575.

Yang, X.-M., Yi, T.-H., Qu, C.-X., Li, H.-N., & Liu, H. (2018). Automated eigensystem realization algorithm for operational modal identification of bridge structures. *Journal of Aerospace Engineering, 32*(2), 04018148.

Yuen, K.-V., & Kuok, S.-C. (2010). Ambient interference in long-term monitoring of buildings. *Engineering Structures*, *32*(8), 2379-2386.

Zhang, G., Ma, J., Chen, Z., & Wang, R. (2014). Automated eigensystem realisation algorithm for operational modal analysis. *Journal of Sound and Vibration*, *333*(15), 3550-3563.

Zhang, Y., Kurata, M., & Lynch, J. P. (2017). Long-term modal analysis of wireless structural monitoring data from a suspension bridge under varying environmental and operational conditions: system design and automated modal analysis. *Journal of Engineering Mechanics*, *143*(4), 04016124.

CHAPTER 3 LOST DATA RECOVERY FOR STRUCTURAL HEALTH MONITORING BASED ON CONVOLUTIONAL NEURAL NETWORKS

Abstract: Signal transmission loss of using wireless sensors for structural health monitoring (SHM) is a usual case, which undermines the reliability of the sensors for monitoring the structural conditions. The measured vibration data with a high data loss ratio can hardly be used for the analysis, i.e. modal identification, as it will lead to significant errors in the results. This chapter proposes a novel approach based on convolutional neural networks (CNN) for recovering the lost vibration data for SHM. The used network is a fully feed-forward CNN with bottleneck architecture and skip connection, which constructs the nonlinear relationships between the incomplete signal with data loss measured from the sensors with the transmission loss and the complete true signal. The trained network extracts the robust higher representation features of the measured incomplete signals using the compression layers, and expands those features gradually throughout the reconstruction layers to recover and obtain the complete true signals. The long-term vibration data from Dowling Hall Footbridge are employed to validate the effectiveness and robustness of the proposed approach for the lost data recovery. Two case studies are conducted to validate the recovery accuracy for single-channel and multiple-channel cases, respectively. The effect of sampling rate on the recovery accuracy is also investigated. The proposed approach exhibits the outstanding capability of lost data recovery, even when the signals have severe data loss ratios up to 90%. To further demonstrate the reliability of the recovered signals for data analysis, modal identification results by using the recovered signals with different data loss ratios show a very good agreement with those obtained from the complete true data.

This chapter was published in *Structural Control and Health Monitoring* with the full bibliographic citation as follows: Fan G, Li J, Hao H. (2019). Lost data recovery for structural health monitoring based on convolutional neural networks. *Structural Control and Health Monitoring, 26*(10): e2433. https://doi.org/10.1002/stc.2433.

3.1 Introduction

Safe and reliable operations of civil infrastructure, such as buildings and bridges, are of significant importance to public. Long-term structural health monitoring (SHM) has been rapidly developed in recent years, owing to its capability of tracking the structural health conditions and reporting the abnormal status of structures at an earlier stage to prevent structural failures and calamity. For real engineering applications, a comprehensive long-term SHM system demands a large number and various types of sensors, data acquisition (DAQ) systems, transmission (long cables or wireless) and other equipment. The conventional SHM systems that use wired sensors to provide a stable signal measurement and transmission have been widely adopted in the last several decades. However, these SHM systems are expensive and the installation of such systems is labour intensive and time consuming for large-scale civil engineering structures, such as super-tall buildings and long-span bridges (Ni et al., 2011). On the other hand, the research on using wireless sensors and wireless transmission technology in SHM has attracted significant attention. The wireless sensor based SHM systems without the complicated cabling system can provide a more flexible selection of sensor locations and is more economical to install and maintain. Furthermore, wireless sensors can provide additional computation capability to process the local measurement on board (Lynch & Loh, 2006). Consequently, wireless sensing techniques have been intensively developed from different aspects, such as the sensor hardware, the network deployment, communication technique and data acquisition, to improve the performance and usability for civil engineering SHM (Chen & Casciati, 2014; Kane et al., 2014; Linderman et al., 2013; Lynch, 2005; Sun et al., 2016; Winter & Swartz, 2017).

A remarkable number of civil infrastructures have been monitored by using the wireless sensor based SHM systems (Kim et al., 2007; Ni et al., 2009; Zhang et al., 2017). However, data loss still occurs in wireless transmission and has seriously affected the development and application of wireless sensors for SHM. It has been reported that a 0.4% data loss is close to injecting a 5% noise on the power spectrum density (Nagayama et al., 2007). The measured vibration data with a high data loss ratio can hardly be used for the modal analysis as it may

lead to significant errors in the identification results, which consequently influences the accuracy of structural condition assessment. The potential reasons for the signal loss are complex including the radio interaction, ambient condition variation, sensor fault, and signal attenuation in the long-distance transmission. The influence of those factors can be mitigated but it is impossible to totally remove from the wireless transmission process at present.

Data recovery can be considered as a remedy when the data loss occurs. To address the lost data recovery issue in SHM, approaches based on the finite element model (FEM) have been proposed (He et al., 2012; Law et al., 2011). Other than those model-based methods which demand an accurate FEM, response recovery conducted directly from the measurements is more attractive and applicable for large-scale infrastructure with numerous uncertainties. Bao et al. (2013) proposed an innovative approach based on the compressive sampling technique for the signal loss recovery of wireless sensor networks installed on a bridge structure. The recovery accuracy is good but depends significantly on the selected domain that the time-domain signal is transferred to, and the threshold value that makes the output matrix sparse to overcome the ill-posed issue. Wan and Ni (2019) leveraged the Bayesian multi-task learning methodology to recover the blocked SHM missing data including temperature data and acceleration responses. It is reported that the choice of the covariance function has a great influence on the recovery performance.

Deep learning is one of the emerging techniques in computer science and has been applied for civil SHM in recent years. The development and applications of deep learning based algorithms such as Convolutional Neural Network (CNN) and Auto-Encoder for SHM mainly focus on damage detection using the crack images (Cha et al., 2017), vibration signals (Abdeljaber et al., 2017; Lin et al., 2017) and vibration characteristics (Pathirage et al., 2018; Pathirage et al., 2019). Nevertheless, deep learning has also been proved to be effective in addressing the signal recovery problems such as image super-resolution (Dong et al., 2015) and audio super-resolution (Kuleshov et al., 2017). The trained deep learning models predict the high-resolution components of images or audios through the known low-resolution ones. Another interesting area attracting extensive attention is the damaged photo recovery (Yeh et

al., 2017), which is also known as in-painting. It recovers the damaged part of the photo from the remaining part, which is a two-dimensional case of lost signal recovery.

In this study, the application of deep learning based techniques is extended to recover the lost time domain measurement data during the transmission in SHM systems. The proposed deep learning based algorithm to perform the lost SHM data recovery is based on a fully CNN, which has been reported to be efficient to implement the dense predictions for pixel to pixel tasks such as semantic segmentation (Long et al., 2015), image segmentation (Ronneberger et al., 2015) and image restoration (Mao et al., 2016). In the rest of this chapter, Section 3.2 is organized to present the detail configuration of the used CNN including the architecture of the network and the relevant parameters. The techniques including skip connection, dropout, the selected activation function and the sub-pixel shuffling operation used to improve the performance of the proposed network will also be clearly described. The preparation of the datasets for training the CNN will be introduced in Section 3.3. Dowling Hall Footbridge which has been installed with a long-term SHM system is selected as the test bed for validating this method. In Section 3.4, two networks are trained and tested with signals of 10% to 90% random loss ratios for evaluating and comparing the performance of the proposed approach for lost data recovery from one channel and from all eight channels of the whole SHM system. In Section 3.5, modal identification by using the original and recovered signals is conducted to investigate the usability and effectiveness of using the proposed approach to obtain the accurate and reliable recovered signals for condition assessment and damage detection.

3.2 Methodology

The proposed CNN is a fully convolutional feed-forward network consisting of a number of sequential convolutional layers to implement the gradually higher-level features extraction and signal reconstruction, by using the compression and reconstruction layers respectively. It has a bottleneck structure as originated from the Deep Auto-Encoders. The bottleneck structure is beneficial for automatically extracting the higher-level features gradually from the input data, meanwhile mitigating the noise effect and reducing the dimensionality of features before feeding to data recovery training. The proposed CNN keeps the advantage of CNN

while sharing the benefit of bottleneck architecture by replacing all the encoder and decoder layers with convolutional layers. It is noted that the convolutional layers consist of multiple convolutional kernels, which can extract more robust features than the encoder layer.

Another significant merit of the proposed approach is making use of residual blocks to overcome the vanishing gradients, improve the training efficiency of deeper networks and potentially tune the number of layers during training. The compression and reconstruction layers are symmetrically linked with skip connections, and therefore the network training converges much faster and obtains high quality optimization results. The skip connections allow the signals to be back-propagated to bottom layers directly and tackle the problem of gradient vanishing, which improves the training efficiency and accuracy of deep neural networks. In this study, for lost vibration data recovery, the input and output data are similar especially when only a small proportion of data are lost. In this situation, the layer in the symmetric position shares a large number of features, such as the amplitude of the sampling points and the waveform. The skip connections between each pair of layers effectively shuttle the features extracted in the shallow layers to deeper layers, which enriches the features in deeper layers and consequently improves the information flow and efficacy of the network.

3.2.1 The architecture of the used CNN

The used fully CNN consists of an input layer, several convolutional layers and one output layer. The architecture of the CNN is illustrated in Figure 3-1, and the detailed parameters of each layer are shown in Table 3-1. It should be mentioned that the length of feature maps as shown in Figure 3-1, and the input and output shapes of each layer listed in Table 3-1 are defined according to the training datasets, where each sample is equally segmented as 1024 points. In fact, the input data length of this fully CNN can be arbitrary. Consequently, in the testing stage, other network parameters excluding the input and output dimensions, are kept as constant. The input and output layers have the same number of neurons, which are equal to the length of the signal to be recovered. The input and output feature map shapes of each hidden layer vary with the input data size.

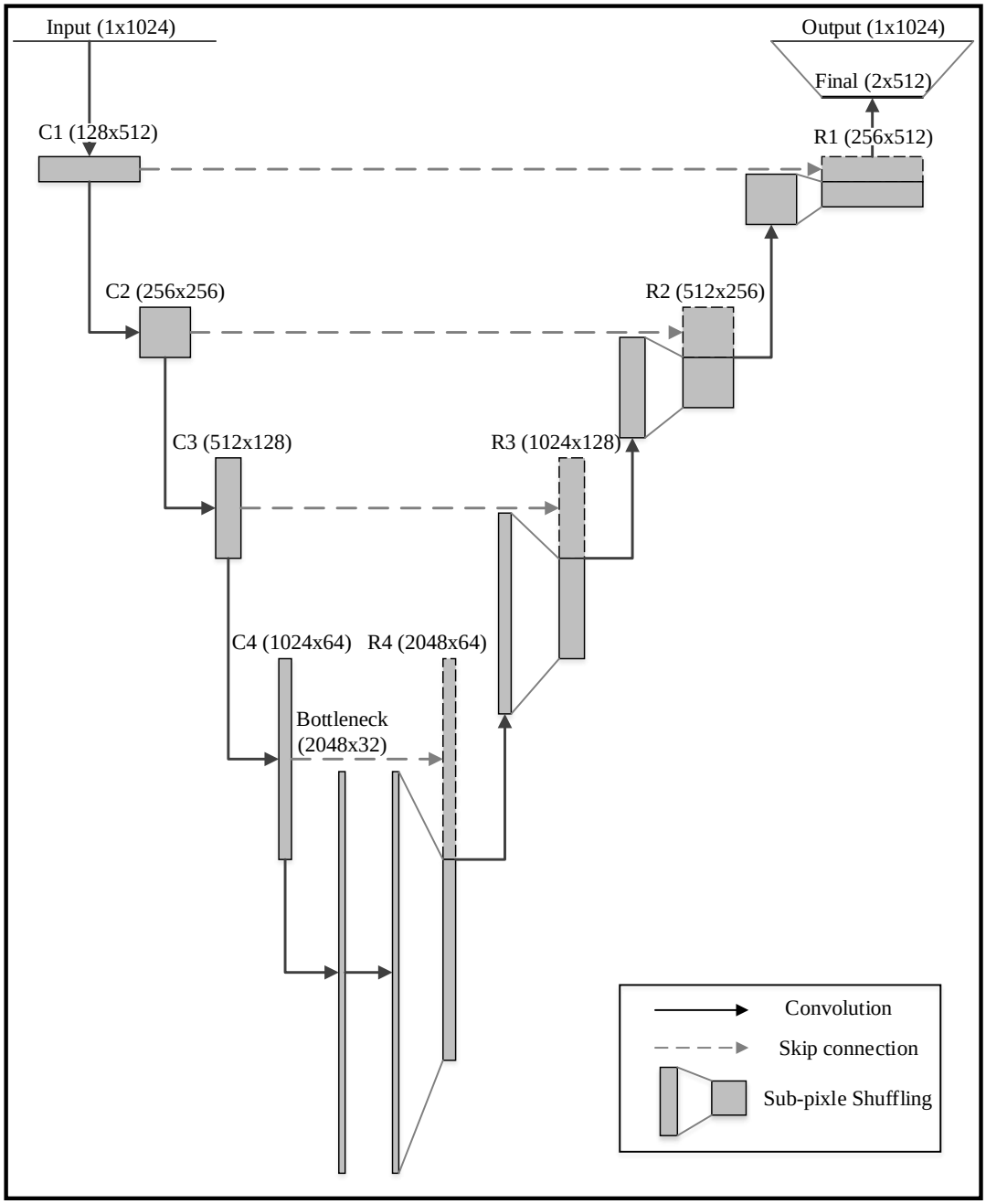


Figure 3-1. The architecture of the used fully convolutional neural network.

Table 3-1. The detailed parameters for each layer of the used CNN

Layer	Kernel number	Kernel size	Stride	Padding	Input shape	Output shape	sub-pixel shuffling
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
C4	1024	8	2	Same	(512,128)	(1024, 64)	N
Bottle neck	1024	8	1	Same	(1024,64)	(2048, 32)	N
R4	2048	8	1	Same	(2048,32)	(2048, 64)	Y
R3	1024	16	1	Same	(2048,64)	(1024, 128)	Y
R2	512	32	1	Same	(1024,128)	(512, 256)	Y
R1	256	64	1	Same	(512,256)	(256, 512)	Y
Final	2	8	1	Same	(256,512)	(1, 1024)	Y

The bottleneck architecture employed in this network consists of the compression layers, bottleneck layer and reconstruction layers. The first compression layer C1 extracts the low-level features from the input data and the following compression layers obtain half of the length of their input feature maps by using convolution kernels with a stride of 2 to extract the higher dimensional representation of the input. Meanwhile, it doubles the number of feature maps using two times the number of kernels to enrich the number of higher-level features. On the other hand, the reconstruction layers enlarge the feature maps gradually and reduce the number of feature maps simultaneously until reaching the same dimension as the first compression layer. Skip connection is implemented between the compression layers and reconstruction layers in the symmetric positions to build the residual blocks, promote the information flow and alleviate the gradient vanishing as shown in Figure 3-1. The output of the reconstruction layer is concatenated with the delivered output features of the reconstruction layer. A layer between the compression layers and the reconstruction layers is called 'bottleneck layer', which keeps the smallest dimensionality but represents the highest level of features of the input data. The bottleneck layer has the same operation as the compression layers. For the audio signal, the bottom layers (close to the input) learn the wavelet features of the input, and the top layer (the bottleneck layer) may learn the phonemes of the audio (Aytar et al., 2016). Transferring the theory to structural vibration lost data recovery, the bottom layers may also learn the wavelet features in the time domain, and the higher layers may extract

the vibration characteristics, i.e. natural frequencies of the structure (Lin et al., 2017). A final convolutional layer that concentrates and shuffles the features as the same dimension of the output layer, is proposed between the last reconstruction layer and the output layer. The dropout technique is implemented to prevent over-fitting which will be elaborated in section 3.2.2.

To sum up, the successive compression layers progressively extract the higher-level features of the incomplete data until the bottleneck layer. After the bottleneck layer, the process back-calculates where the output features at reconstruction layers are concatenated with the output features at compression layers and gradually reconstruct the signal to approach the original one. The compression layers and bottleneck layer involve two operations, that is, (1) The convolution of the input; and (2) Nonlinear activation of the output using a Leaky rectifier. The reconstruction layers involve four operation procedures, namely, (1) Convolution of the input features; (2) Nonlinear activation of the output using a Leaky rectifier; (3) Resize the activated output; and (4) Finally concatenate the activated output with the features from the compression layer in the mirrored position. In addition, the last layer contains, (1) The convolution; and (2) Resize operations. The details of the activation function, resizing the output features by a one-dimension subpixel shuffling operation are elaborated in the following sections.

Since the incomplete and original signals that share the same dimension, the objective function selected to optimize the CNN parameters is the mean of the L2-norm error of each sample in a mini batch as

$$Loss = \frac{1}{N}\sum_{k=1}^{N}\left\|y_k - f_\phi(x_k)\right\|_2^2 \tag{3.1}$$

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

3.2.2 Dropout

Dropout is a technique to address the overfitting issues in deep neural networks. Instead of dealing with the overfitting by training several neural networks to start from the different initial conditions and average the predictions from each neural network, dropout is a more efficient technique (Srivastava et al., 2014). Dropout means dropping out neurons and disconnects those neurons with the adjacent input and output layers. It drops out neurons randomly when training with a different batch of samples to break-up the co-adapted sets of neurons and makes those remained neurons more robust. Each neuron with an independent probability *p* to be dropped, where *p* is set as 0.5 (Srivastava et al., 2014). This is the optimal dropout probability for a variety of neural networks and tasks. It should be mentioned that when testing a neural network trained using dropout technique, the weight of each neuron should be scaled-down by multiplying *p* because all the neurons will participate in predicting the output in the testing process.

3.2.3 Leaky rectified linear unit (ReLU)

The activation function follows the convolution operations that transfer the mapping between the input to the output nonlinearly. It is essential for the recovery task since the relationship between the incomplete and original data cannot be simply presented by a linear function. Leaky ReLU (Maas et al., 2013) as an improved version of ReLU (Nair & Hinton, 2010) has attracted significant attention recently. Leaky ReLU has a nonlinear gradient over its whole domain as shown in Equation (3.2). Leaky ReLU prevents the gradient vanishing that potentially happens when using *'sigmoid'* as the activation function. Meanwhile, it allows a non-zero small gradient to overcome the defect of ReLU, that is, the neuron could potentially never be activated once a large gradient flows through it (Maas et al., 2013).

$$f(x) = \begin{cases} x, & if\ x > 0 \\ 0.01x, & \text{otherwise} \end{cases} \tag{3.2}$$

3.2.4 Sub-pixel shuffling operation

As shown in Figure 3-1, the original output feature maps from the reconstruction layers have half length of the feature maps at the corresponding compression layer. To keep those feature maps with a consistent size, the output feature maps from the reconstruction layer are resized by a sub-pixel shuffling operation, which doubles the length and half divide the number of feature maps. The shuffling operation is implemented by splitting the feature maps as two groups and combining two feature maps at the same position of each group as one by interpolating one to another as shown in Figure 3-2. It is a one-dimensional case of the sub-pixel convolution layer (Shi et al., 2016). This is an efficient sub-pixel shuffling operation that takes less computation load compared to the general deconvolution. The sub-pixel shuffling operation has also been attested to introduce less artefact in the output (Odena et al., 2016).

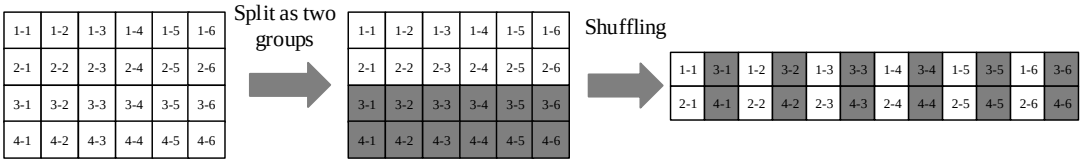


Figure 3-2. An example of one-dimensional sub-pixel shuffling with 4 feature maps and 6 elements in each map.

3.2.5 Metrics

Given the original signal as *y* and the recovered signal *y'*, the recovery error is evaluated by comparing the difference between two signals using L2-norm, and is calculated as

$$\xi = \frac{\|y - y'\|}{\|y\|} = \frac{\sum_1^n (y_i - y_i')^2}{\sum_1^n {y_i}^2} \tag{3.3}$$

When evaluating the error in the frequency spectrum, *y and y'* become the original and recovered signal spectra in the frequency domain.

3.2.6 Datasets

A robust neural network requires abundant datasets to train and fine-tune the network parameters. Structural vibration data especially the long-term monitoring vibration data fit

well with the requirement of ‘big data’. The original measured acceleration data are processed to serve as the input and output of the neural network. The objective of this study is to recover the original signal using the incomplete signal, hence the input of the neural network is the acceleration signal with transmission loss and the output is the original acceleration signal to be recovered. In practice, the recorded data in the receiver or base station are numbered sequentially (Bao et al., 2013). This means that the time point of the lost data is known. At the time point that the signal is lost, the record will be empty or set as zero. Consequently, the input incomplete signal is generated from the original signal, where the transmission data loss is simulated by setting a proportion of the sampling points as zero entries corresponding to the lost proportion (Bao et al., 2018). Figure 3-3 shows an example of the input signal with 50% data loss and the corresponding original signal measured by the SHM system of Dowling Hall Footbridge which will be detailed in Section 3.3.

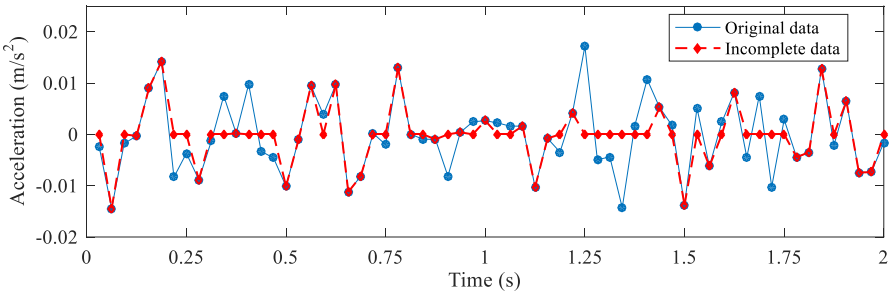


Figure 3-3. A comparison between the incomplete signal with 50% data loss and the corresponding original signal.

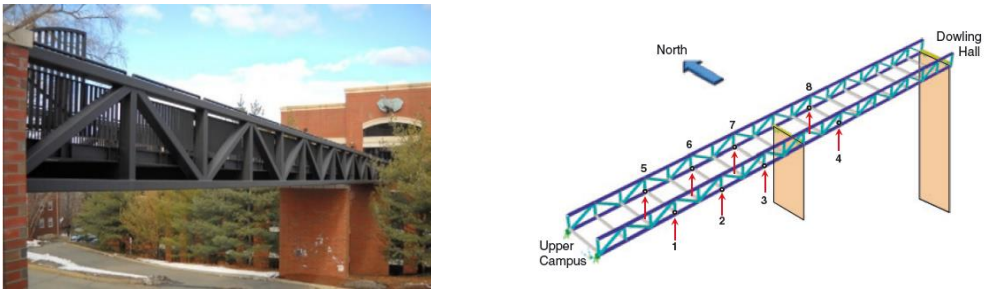
3.3. Experimental validations on lost data recovery

Experimental studies are conducted to validate the effectiveness and robustness of using the proposed approach for SHM lost data recovery. The experimental studies include three scenarios, namely, a single channel signal recovery case, a multiple channel signals recovery case and a comparison study for investigating the effect of the sampling rate. These studies share the same network architecture but trained by different datasets. In this work, the performances of lost data recovery by using a CNN trained based on data from a single channel and another network trained from multiple channels data are firstly investigated. Network for

a single channel can be trained faster, and it can be only used to recover the lost data of this channel. In contrast, the network training for multiple channels is relatively time-consuming but has more comprehensive and practical applications to recover the data of more channels. The recovery accuracy may be influenced by the significantly reduced sampling rate. Therefore, a comparison study to investigate the effect of sampling rate is also conducted using the same data with single channel case but different sampling rates.

3.3.1 Dowling Hall Footbridge and the installed SHM system

The data used for validating the proposed method are obtained from a long-term SHM system installed on the Dowling Hall Footbridge, which is located on the Medford, MA, campus of Tufts University. It is a pedestrian bridge connecting the main campus and the student services offices as shown in Figure 3-4(a). This bridge is a two-span continuous steel frame bridge with a reinforced concrete slab. The total length of Dowling Hall Footbridge is 44 m and the width is 3.7 m. A long-term SHM system has been installed to monitor the vibration responses and ambient conditions of the bridge continuously. 16 weeks vibration response from 8 accelerometers as shown in Figure 3-4(b) are collected from January 2010 to May 2010. Eight PCB 393B04 uniaxial accelerometers are installed on the bridge and connected to National Instruments cRIO-9074 DAQ system including integrated chassis and controller. This sensor model has a frequency range from 0.06-450Hz, with 3 μg Root Mean Square broadband resolution. The communications with the DAQ system is conducted through a wireless-G network at Tufts University. A wireless bridge with an external Hawking HAO14SDP antenna featuring a 14dBi gain factor and all-weather construction. More detailed descriptions of the footbridge and the SHM system can be found in the literature (Moser & Moaveni, 2013). System modal identification of this bridge was conducted in a previous study (Moser & Moaveni, 2011), and the identified first six nature frequencies, damping ratios and the corresponding mode shapes are used as reference baselines in this study for assessing the reliability and accuracy of modal identification results obtained from the recovered data.



(a) (b)

Figure 3-4. (a) South view of Dowling Hall Footbridge. (b) Accelerometer layout.

3.3.2 Data preprocessing

Preprocessing data to satisfy the purpose of the network is necessary to improve the training process. In this study, to optimize the performance of the used CNN, the measured accelerations are preprocessed based on signal filtering techniques and data normalization. As introduced in Ref. (Moser & Moaveni, 2013), the provided data have a sampling rate of 128 Hz and have already been filtered by a band-pass filter from 2 Hz to 55 Hz. However, the frequencies of interest are within 4 Hz to 14 Hz as reported in a previous study (Moser & Moaveni, 2011). To further minimize the redundant information contained in the acceleration data, the acceleration data for the single and multiple channel cases are processed by a low-pass filter with a cut-off frequency of 16 Hz. It is admitted that less information contained in the training dataset means fewer features need to be learned by the networks. With limited neurons in the neural networks, the filtering in fact assists in improving the machine learning process where the neurons can memorize more about the 'useful' features. After filtering, a downsampling step is implemented to the measured acceleration data to reduce the sampling rate to 32 Hz. This step shrinks the data size as one-quarter of the original data, which enhances the training efficiency of the used CNN significantly. Finally, the acceleration data were normalized to the range -1 to 1 by using the below equation

$$accleration_{normalized} = \frac{accleration}{max\ (|accleration|)} \tag{3.4}$$

3.3.3 Datasets generation

The measured acceleration data in the first week are chosen, in which 5 minutes data in each hour are utilized to generate the training and validation datasets for the lost data recovery. Four 5-minutes long segmental signals from four randomly selected hours in the first and second week are used to generate the testing data for investigating the performance of the trained CNN. The total length of data used for training the CNN is 14 hours. Compared to the large size of data measured from the long term SHM systems, training a robust CNN for recovering the lost data by using only 14 hours data is practical and effective.

The normalized acceleration data from the second channel of the SHM system, as shown in Figure 3-4(b), are selected to evaluate the performance of single channel data loss recovery and the comparison study. Since the monitoring systems are always working under uncertain conditions, the lost points and the lost proportion are also random, i.e., in different periods, the lost pattern and the lost proportion can vary greatly. To simulate the data loss pattern, 5 groups of data with different lost proportions using the normalized 14 hours acceleration data are generated. Each group of data contains the original signal and the incomplete signal with a certain lost proportion from 10% to 90% with a 20% increase. The lost signal points are randomly selected from the time sequence. The training and validation samples are small patches segmented from the full time series of the acceleration data. A window with a length of 1024 points is designed to produce the samples by scanning these 5 groups of data from the beginning to the end with a 50% overlap. The length of each sample is 1024 (32 seconds) and totally around 400,000 samples are produced and recorded in the datasets. The proportion of samples used for training and validation are 90% and 10% respectively.

In the data loss recovery of multiple channels signals, eight channels of data are fed to the CNN together. The way to generate the datasets is slightly different from the single channel case. For the sake of keeping the mode shapes information of the data, the eight channels signals are firstly conducted the ‘data loss’ process individually and then combined as an array following the order of channels. The positions of the lost data points are different for each channel which is more realistic in real applications. The combined time series are then

normalized using the maximum absolute value among this array. The same window and scanning strategy following the single channel case are used, where the number of segmented samples is around 3,200,000 in total. Four 5-minutes long normalized segmental signals are processed as four groups to serve as the testing data (Test data 1 to Test data 4) consisting of incomplete data with 10% to 90% data loss ratios. For the single channel case, the signal from the second channel of the Test data 1 to Test data 4 are used. On the other hand, the testing data are all the signals from eight channels in Test data 1 to Test data 4 for the multiple channel case. The multiple channel testing data also form as an array before using as the input to the networks, and then resizing to its original form after obtaining the output. This CNN accepts input with a random size, so the 5-minutes long testing data are directly inputted to the trained CNN to examine the accuracy of the trained networks.

3.4. Recovery performance evaluation

3.4.1 Results of recovering the single channel signal

The computer used for training this network is built with a GTX1060 Graphics card with a 6GB GPU, an i7-6700K CPU and 8GB memory. The training time for the single channel case is 2 hours and for the multiple channel case is around 16 hours. Since the architecture of the used CNN keeps the same for both studies, the training time is mainly dependent on the size of training data. After the training of the network for single channel signal recovery, four groups of testing data containing the second channel signal only are inputted to the network. Processing each group of testing data can be completed instantaneously, therefore a real-time signal recovery is realizable when the trained CNN is ready and could be embedded in the data processing algorithms. The single channel lost data recovery results of the Test data 1 in the time domain are shown in Figure 3-5. A 100 points segment is presented in this figure, which demonstrates the gradually increasing discrepancy between the original and recovered signals with the increase of the data loss ratio. The randomly remained signal points used to recover the original signal are also drawn in Figure 3-5.

Recovered signals by using incomplete signals with 10% to 80% loss ratios have exhibited a strong consistency with the original signals. The recovered and original signals are almost overlapped within these loss ratios range. For the recovered signal using only 10% remained signal points (90% loss), it shows a very similar waveform pattern but for some points, the signal amplitudes are varying around the original signal in a tolerable extent. Figure 3-6 shows the recovery errors quantified using Equation (3.3). The recovery error progressively increases with the increasing severity of data loss. In addition, the recovery performance of Test data 1 and 4 are better than that of Test data 2 and 3, especially when the loss ratios become higher. This phenomenon can be explained as that the Test data 1 and 4 are recorded in busy periods. A large number of pedestrians walking on the footbridge provide strong operational excitations to excite more vibration modes and induce many free-decay formed vibrations in the recorded responses. Those responses cover mainly the first several structural modes of the bridge with a relatively higher signal to noise ratio. Compared with the low level ambient excitations induced responses, the data loss recovery of the responses measured from well excited scenarios is more accurate. Figure 3-7 shows the signal recovered from incomplete data with a 90% loss in a free-decay form in Test data 1. The error of this signal segment is shown in Figure 3-8, where the recovery error is much smaller than that from the Test data 2 and 3 as shown in Figure 3-6. The errors in the data loss recovery are only 5% and 9% for the cases with 10% and 20% loss ratios, respectively.

The recovery accuracy of the proposed approach is also investigated in the frequency domain. The time domain signals are transferred into the frequency domain by using Fast Fourier Transform (FFT). Figure 3-9 shows the Fourier spectra of the recovered and original signals of the Test data 1 in a logarithmic form. The first four modes are well excited at this sensor location during the testing period. The trained CNN has demonstrated a strong ability to recover the lost signal with a 70% loss ratio. By observing the results of 80% data loss, the Fourier spectrum of the recovered signal shows a weak amplitude compared to the original signal in the 3rd mode. Meanwhile, when recovering the 90% lost incomplete signal, an attenuation of the signal strength at the second mode frequency is found. Even though, the first four structural natural frequencies can be clearly identified from the Fourier spectra of those

recovered signals. The natural frequencies identified from the recovered signals are almost the same as those obtained from the original signals. The observation is consistent with the finding from the time domain where the waveforms keep the same but have some differences in the amplitudes.

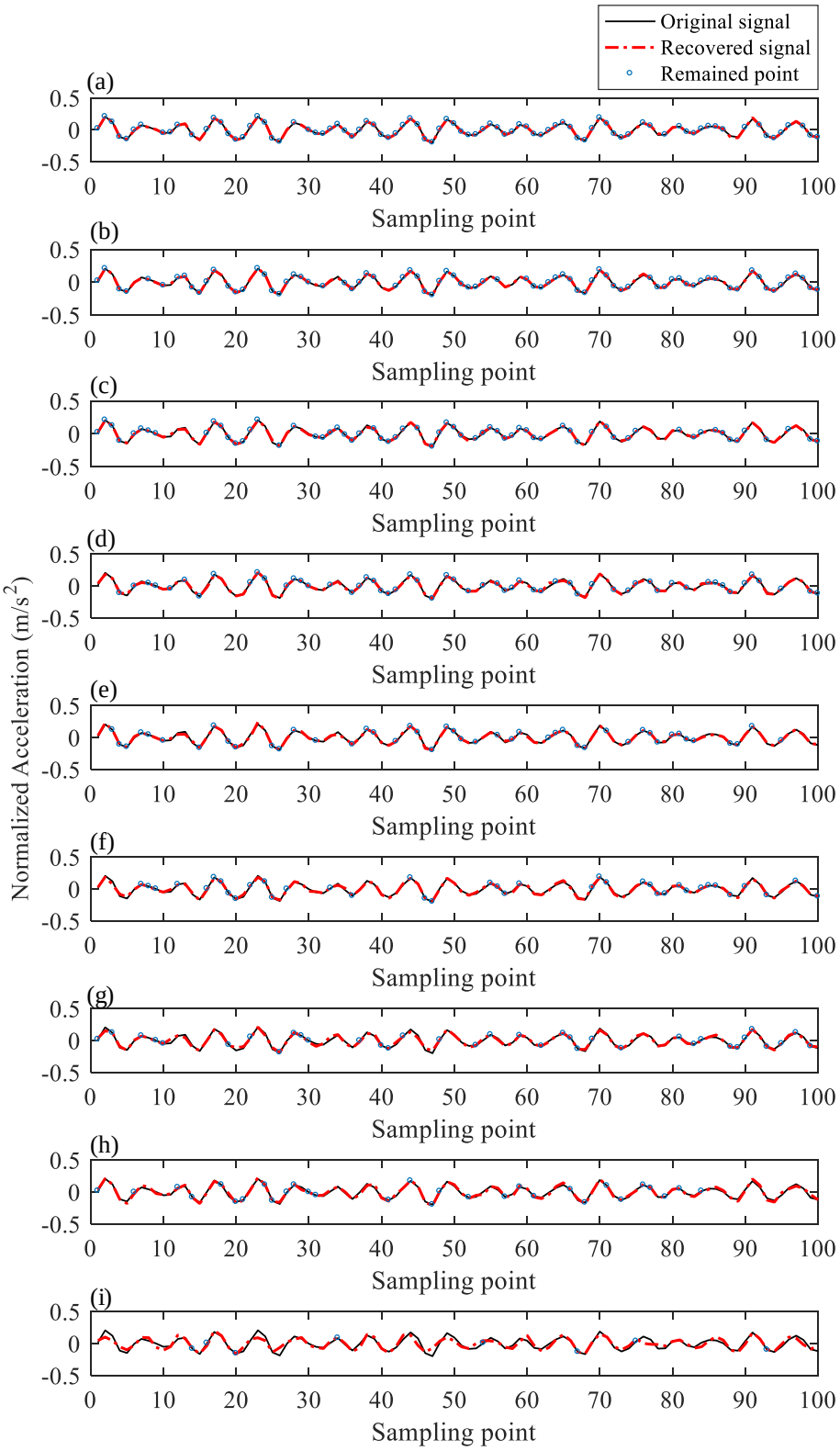


Figure 3-5. Comparisons between the original and recovered signals of 100 sampling points in Test data 1 with different data loss ratios: (a) 10%; (b) 20%; (c) 30%; (d) 40%; (e) 50%; (f) 60%; (g) 70%; (h) 80%; (i) 90%.

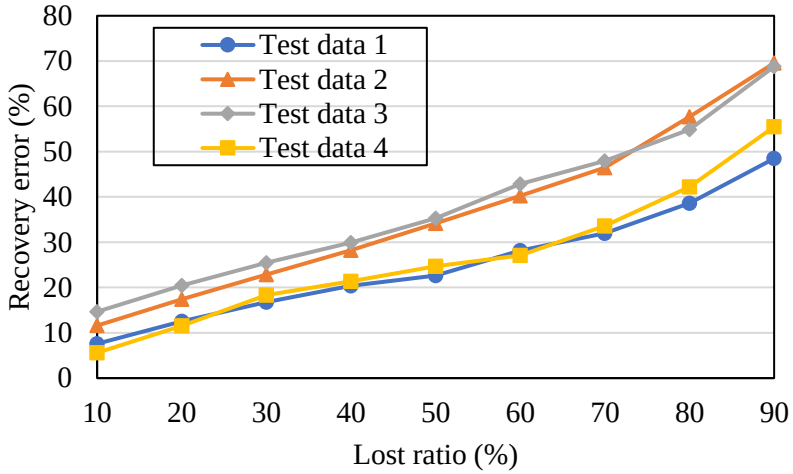


Figure 3-6. The recovery errors of the four groups of testing data.

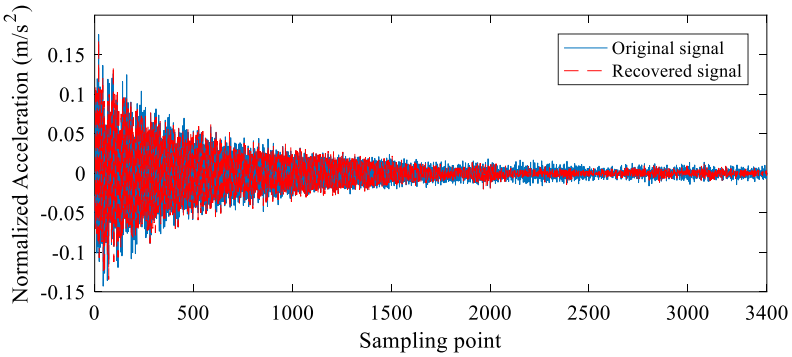


Figure 3-7. The comparison between the original and recovered signals of a free-decay form signal segment in Test data 1 originated from the incomplete data with 90% data loss.

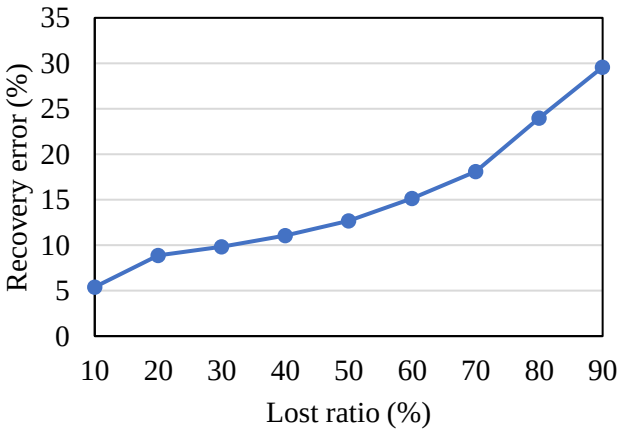


Figure 3-8. The recovery error of the free-decay formed signal segment in Test data 1.

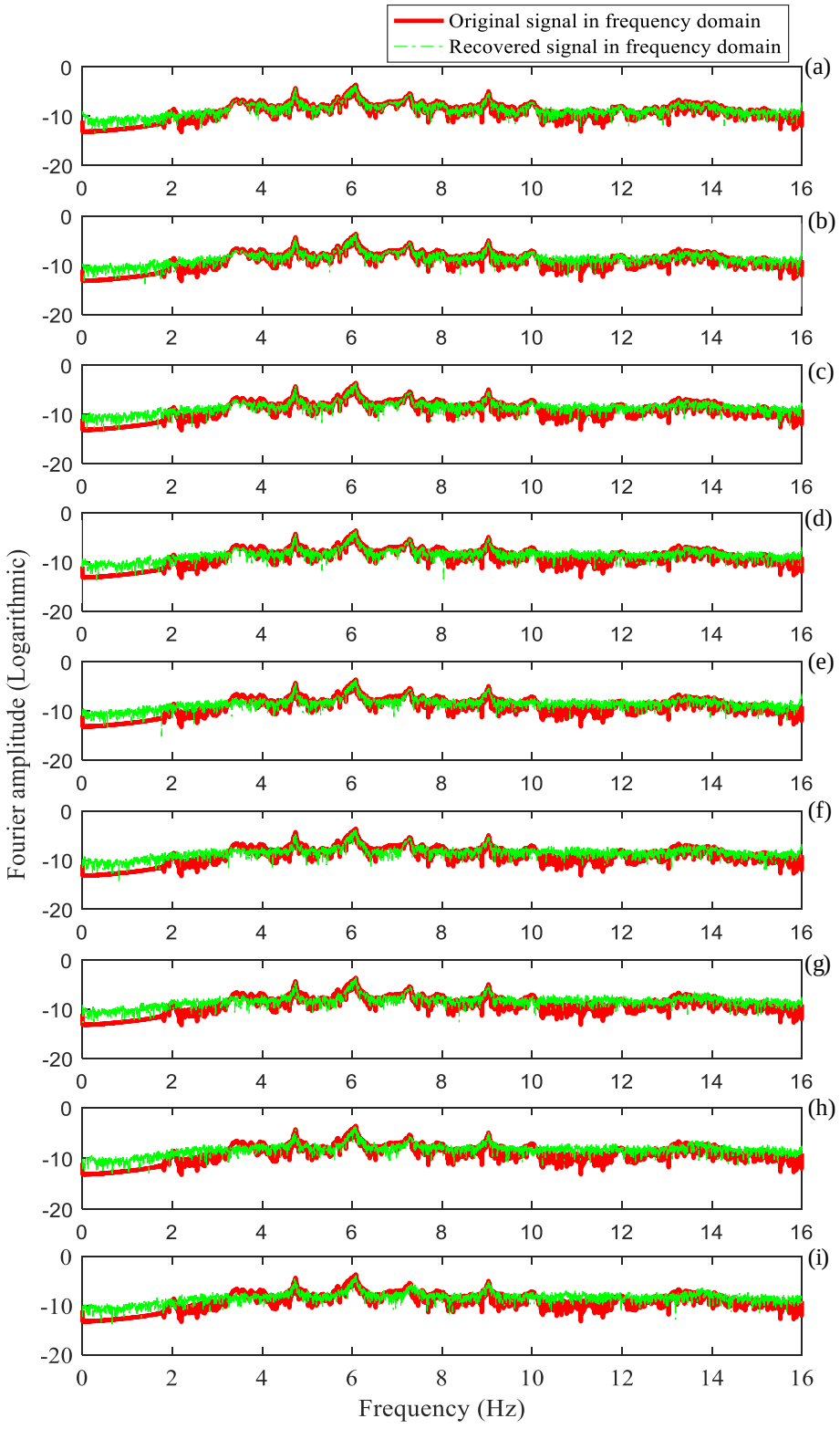


Figure 3-9. Comparisons between the original and recovered signals (Test data 1) in the frequency domain based on incomplete data with different data loss ratios: (a) 10%; (b) 20%; (c) 30%; (d) 40%; (e) 50%; (f) 60%; (g) 70%; (h) 80%; and (i) 90%.

Figure 3-10 shows the recovery errors of the four groups of testing data in the frequency domain with different data loss ratios. The effectiveness and robustness of the proposed approach are validated in these experimental studies, where the structural vibration data have severely lost ratios but can be successfully recovered with small errors and accurate frequency identification results.

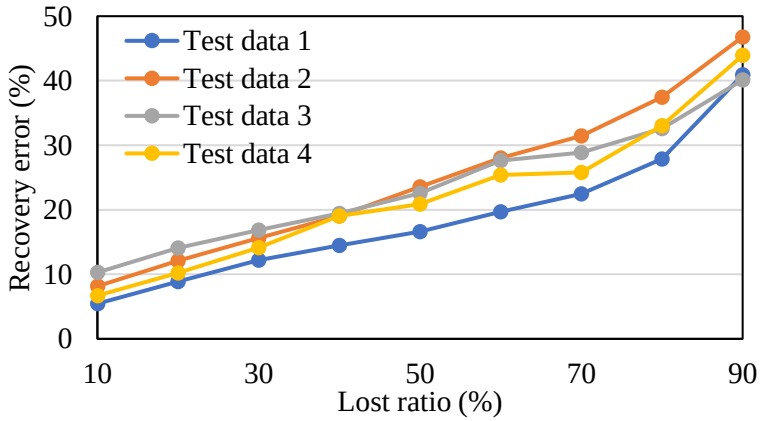


Figure 3-10. The recovery errors of the four groups of testing data in the frequency domain.

3.4.2 Results of recovering the multiple channel signals

When the network for recovering multiple channel signals is successfully trained, four groups of incomplete signals with data loss ratios from 10% to 90% are used as the input to the network with its original length to examine the performance of lost data recovery. An example of the recovery results from the Test data 1 is shown in Figure 3-11. The recovery accuracy of this network is slightly worse than that for the single channel case presented in Section 3.4.1. The recovered signals almost overlap with the original signals with data loss ratios from 10% to 60%. Small errors can be observed between some of the sampling points when the loss ratio increases to 70% and 80%. The recovered signal obtained from the incomplete data with 90% data loss shows relatively larger differences, as shown in Figure 3-11(i). However, the recovered signal still follows the same waveform as the original one with similar amplitudes and frequencies. The quantified recovery errors corresponding to different loss ratios of four groups of testing data are shown in Figure 3-12. For all the four groups of testing data, the recovery errors gradually increase with the loss ratio. The trends of the errors are similar to those between the recovered and original signals in time domain.

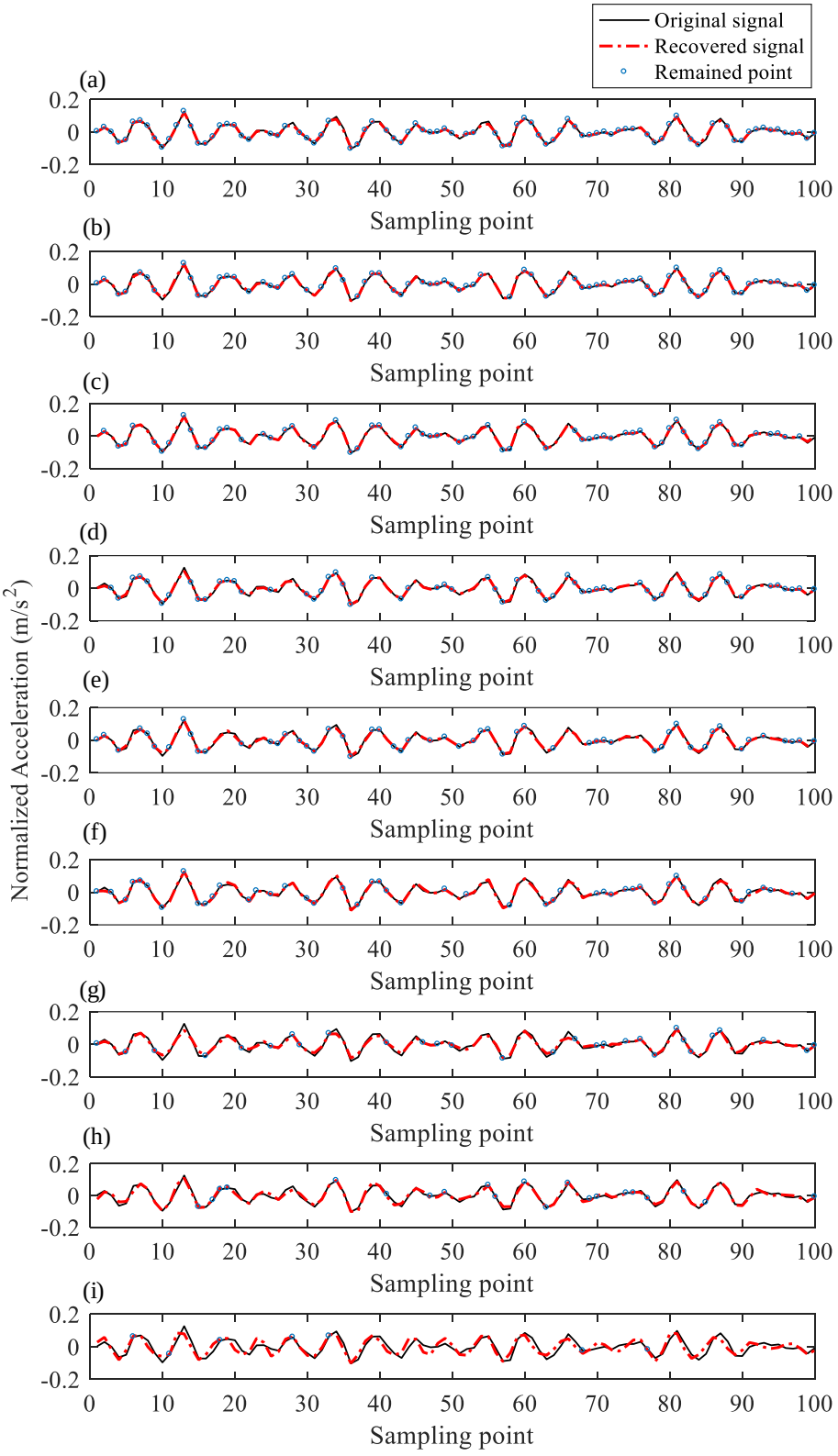


Figure 3-11. The comparison between the original and the recovered signals of 100 sampling points in Test data 1 originated from incomplete data with: (a) 10% data loss, (b) 20% data loss, (c) 30% data loss, (d) 40% data loss, (e) 50% data loss, (f) 60% data loss, (g) 70% data loss, (h) 80% data loss, and (i) 90% data loss.

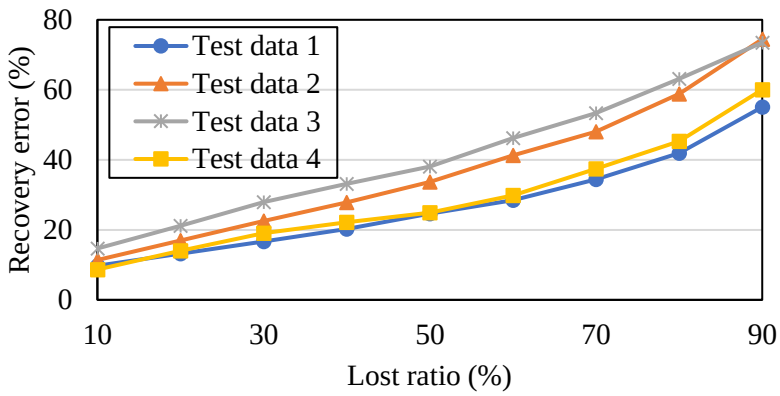


Figure 3-12. The recovery errors of the four groups of testing data.

Figure 3-13 shows the original and recovered signals in the frequency domain originated from incomplete data with the increasing loss ratios from 10% to 90%. The Fourier spectra of the recovered signals are very similar as those of the original signals, when the data loss ratio is between 10% and 30%. The discrepancies in the spectra at the non-structural vibration frequencies points are progressively increasing for the cases from 40% to 70% data loss, but the spectra at structural frequencies are still highly consistent with those of the original signals. The recovery accuracy at the structural frequencies is influenced obviously when the loss ratio raises to 80% and 90%. With such severe data loss levels, recovering the complete information of the original signal is very tough. As a result, the amplitude of the spectrum peaks is underestimated but the frequencies are properly reserved, which means that the natural frequencies can be reliably identified but the mode shapes may be affected in modal identification. The errors between the original and recovered signals in the frequency domain are shown in Figure 3-14. The frequency domain results provide a clearer explanation of the recovery error, where in general the presented CNN learns the structural frequencies features from the training samples accurately and only recovers the critical information of the vibration signals, i.e. structural natural frequencies and vibration modes. At the same time, the redundant information such as noise and disturbance in the vibration signal during the transmission is not recovered. From the neural network aspect, the signals with the information of structural frequencies exist in most of the data samples while the noises that are randomly injected in the response have no clear and consistent pattern to be learned as a significant feature in the data. One of the most significant merits of the proposed approach based on CNN is that it can

automatically learn the best features from the vibration data. Considering the natural frequencies and vibration modes are the most important fundamental vibration characteristics, in this study, they are well recognized and learned by the proposed network and used to capture the patterns in the vibration data. This also means that features representing the signals with structural frequencies are allocated with large weights while the weight of other neurons associated with noises is minimized as it is rarely activated.

Compared with the single channel recovery results, the recovery accuracy of multiple channel signals is slightly worse. It is worth noting that the same network architecture is used for both cases, but the complexity of the multiple channels case is obviously higher which means more features need to be extracted and more nonlinear relationships need to be constructed between the input and output. As a result, the constant number of neurons in the CNN may not be able to record all the features, but higher weights are assigned to the neurons representing the structural vibration characteristics.

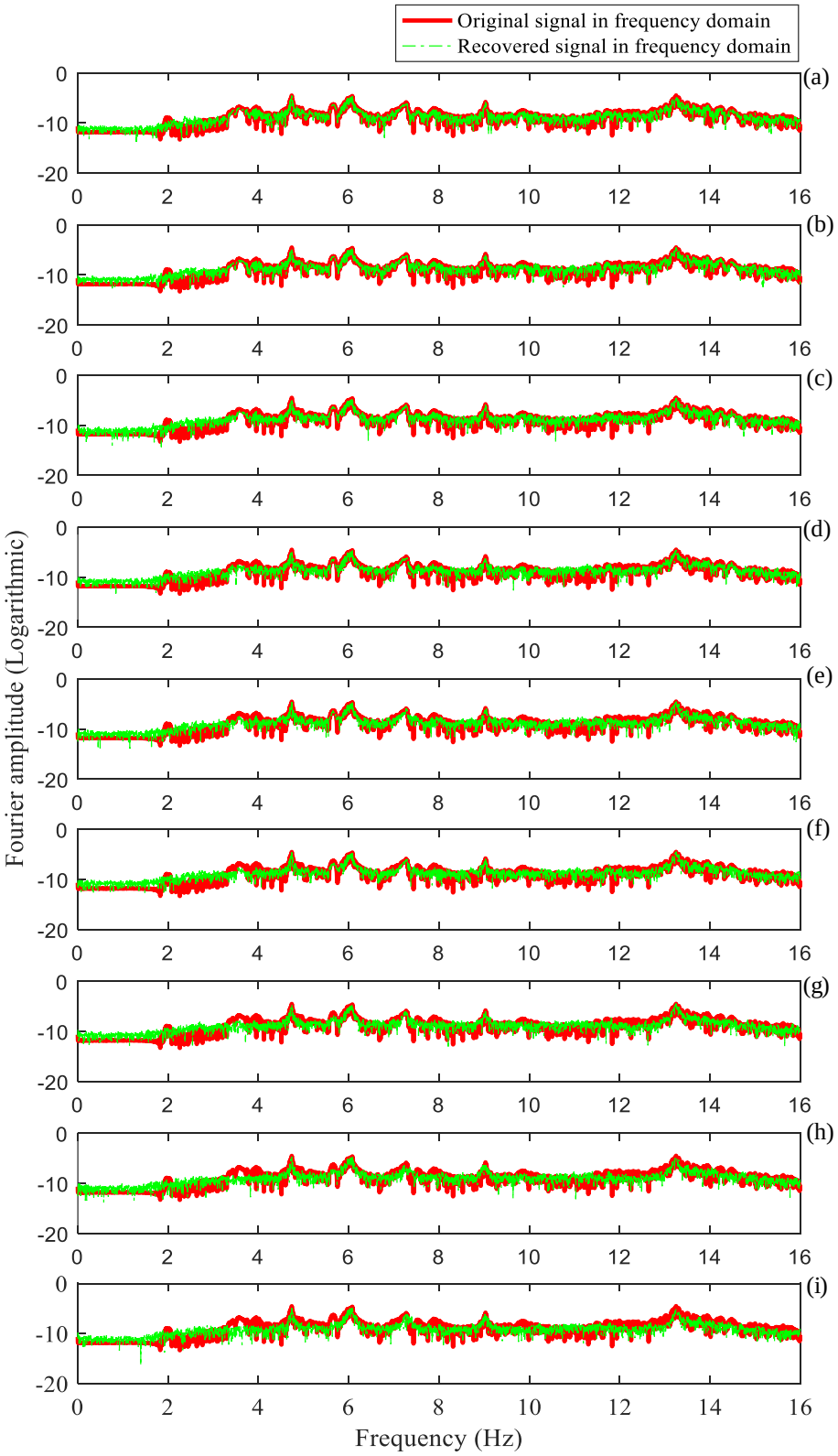


Figure 3-13. Comparisons between the original and recovered signals (Test data 1) in the frequency domain based on incomplete data with different data loss ratios: (a) 10%, (b) 20%, (c) 30%, (d) 40%, (e) 50%, (f) 60%, (g) 70%, (h) 80%, (i) 90%

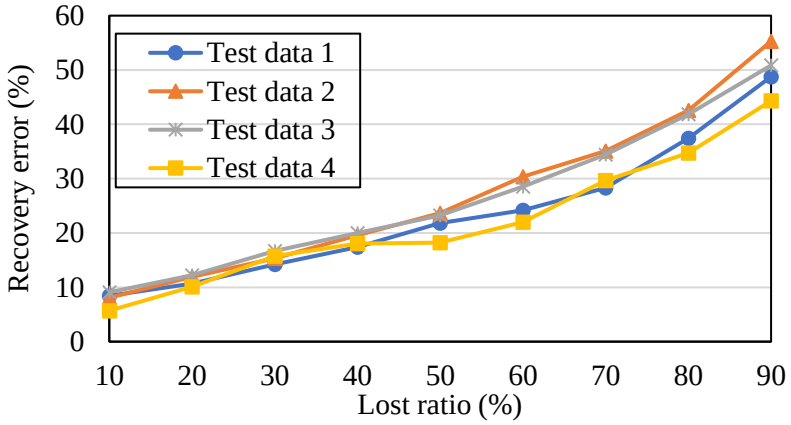
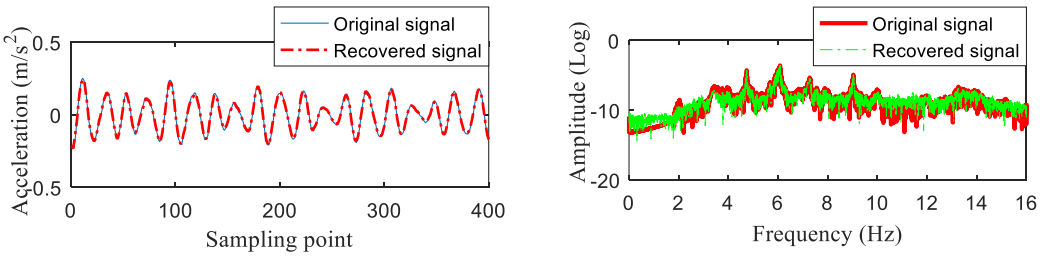


Figure 3-14. The recovery errors of the four groups of testing data in the frequency domain.

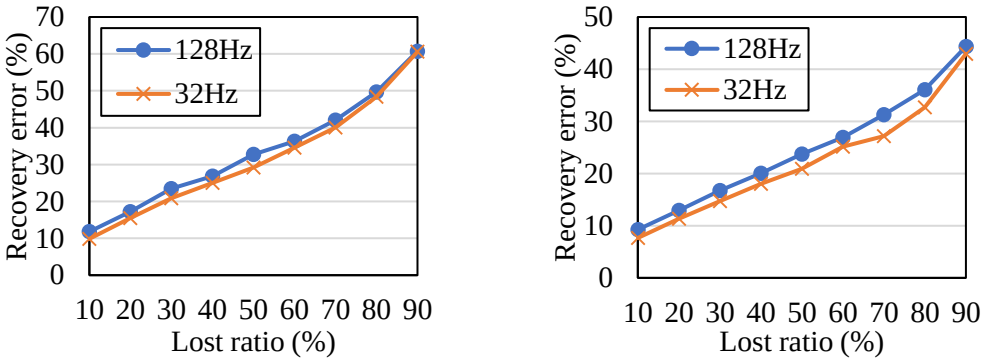
3.4.3 Effect of different sampling rates

The effect of sampling rate is investigated by training two network models with the same configuration while using the same training and testing datasets with different sample rates. Studies by using the same data but with sampling rates of 32 Hz and 128 Hz respectively are conducted. The original and recovered signals in time and frequency domains for the case with 50% loss ratio are shown in Figures 3-15(a) and (b). The recovery errors in both time and frequency domains are shown in Fig. 16. The comparison results show that the recovery performance is slightly improved by reducing the sampling rate. As mentioned that the excited vibration modes of the bridge are within 14 Hz. For signals with a higher sampling rate, redundant information such as noise and local high frequency structural modes may introduce additional errors in the data recovery. The proposed method shows a good performance when using the data sampled at only 32 Hz in this study.



(a) (b)

Figure 3-15. The comparison between the original and recovered signals from incomplete data with 50% loss ratio in: (a) Time domain; (b) Frequency domain



(a) (b)

Figure 3-16. The comparison of average recovery errors in: (a) Time domain; (b) Frequency domain.

3.5 Modal identification using original and recovered signals

Modal identification of this footbridge by using the original and recovered signals is further conducted to demonstrate the feasibility and accuracy of using the recovered signals for modal identification. Stochastic Subspace Identification (SSI) (Van Overschee & De Moor, 2012) is employed to identify the natural frequencies, damping ratios and mode shapes of this footbridge. SSI is a time domain parametric identification method that uses the output-only data for the modal identification. Another operational modal analysis method named Frequency Domain Decomposition (FDD) (Brincker et al., 2001) is used to produce the first singular values of the response spectrum matrix, where the peaks of the singular values plot are most likely presenting the natural frequencies. The identified modes by using SSI method with both the original and recovered signals reconstructed from the incomplete data with 90% data loss are shown in the stabilization diagrams presented in Figures 3-17(a) and (b), respectively. The first six modes can be identified in both stabilization diagrams, which means that the first six modes are excited and the proposed approach effectively recovers the signals at those structural frequencies, although an extremely severe data loss ratio of 90% is considered. Comparing the results in Figures 3-17(a) and (b), it can be observed that the stabilization diagram of the original signals contains more spurious modes than that of the recovered signal. The spurious modes especially those at frequencies around 3.8 Hz and 13.5 Hz that are relatively stable along the system model orders, are potentially reserved in the

identification results. This may lead to false identification results. Besides, the physical mode at the frequency around 13.8 Hz represents the sixth structural mode, and is inconsistent along the system model orders, particularly when comparing the damping ratios. Meanwhile, many spurious modes are also generated around this physical mode which may result in the false identification of this mode if no prior knowledge is provided. The singular values obtained from the FDD analysis of the original and recovered signals as shown in Figures 3-18(a) and (b) also confirm the results of SSI. A massive number of relatively weak peaks representing the spurious modes can be observed in Figure 3-18(a), and several peaks are closely spaced around the sixth mode. On the other side, the singular values plot of the recovered signal is much cleaner, i.e., fewer spurious mode peaks and very distinguished physical mode peaks are observed in Figure 3-18(b).

The errors in the identified modal parameters from the original signals and recovered signals, which are obtained from the incomplete data with 30%, 60% and 90% lost ratios, are summarized in Table 3-2, Table 3-3 and Table 3-4, respectively. It can be observed that the identified natural frequencies using the recovered signals are very accurate for the first six modes. The maximum discrepancy is about 0.6% when the loss ratio reaches 90%, which is acceptable for further analysis such as anomaly detection and damage detection. It should be highlighted that the testing signals are measured under different ambient conditions compared with training datasets, which means that the network has learned robust features of the vibration signals. It can be used for recovering the new measured incomplete data, which are not included in the training datasets. The identified damping ratios from the recovered signals show relatively large errors compared with the ones from the original signals. However, when analysing the original signals, the identified damping ratios at different system orders in SSI analysis output exhibit a large discrepancy, especially at the sixth mode. Such a significant difference may also be caused by the identification methodology error.

Table 3-2. Comparison between the modal identification results from original and recovered signals originated from the incomplete signal with 30% data loss.

Mode	Frequency (original)	Frequency (recovered)	Error	Damping ratio (original)	Damping ratio (recovered)	Error	MAC
1st	4.74Hz	4.74Hz	0.0%	0.0016	0.0019	18.75%	1.00
2nd	6.03Hz	6.03Hz	0.0%	0.0034	0.0044	29.41%	1.00
3rd	7.24Hz	7.24Hz	0.0%	0.0023	0.0017	26.09%	1.00
4th	9.04Hz	9.04Hz	0.0%	0.0018	0.0021	16.67%	1.00
5th	13.24Hz	13.24Hz	0.0%	0.0019	0.0017	10.53%	1.00
6th	13.87Hz	13.87Hz	0.0%	0.0014	0.0019	35.71%	0.99

Table 3-3. Comparison between the modal identification results from original and recovered signals originated from the incomplete signal with 60% data loss.

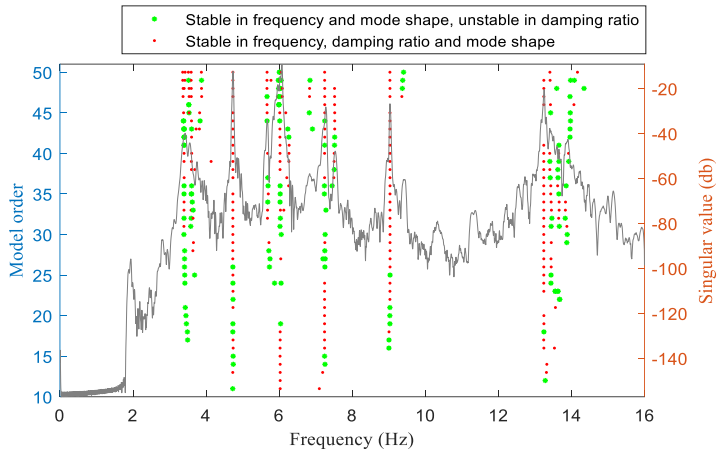
Mode	Frequency (original)	Frequency (recovered)	Error	Damping ratio (original)	Damping ratio (recovered)	Error	MAC
1st	4.74Hz	4.74Hz	0.0%	0.0016	0.0021	31.25%	1.00
2nd	6.03Hz	6.04Hz	0.3%	0.0034	0.0038	11.76%	1.00
3rd	7.24Hz	7.25Hz	0.1%	0.0023	0.0028	21.74%	1.00
4th	9.04Hz	9.03Hz	0.1%	0.0018	0.0024	33.33%	0.99
5th	13.24Hz	13.25Hz	0.1%	0.0019	0.0018	5.26%	0.99
6th	13.87Hz	13.86Hz	0.1%	0.0014	0.0024	71.43%	0.98

Table 3-4. Comparison between the modal identification results from original and recovered signal originated from the incomplete signal with 90% data loss.

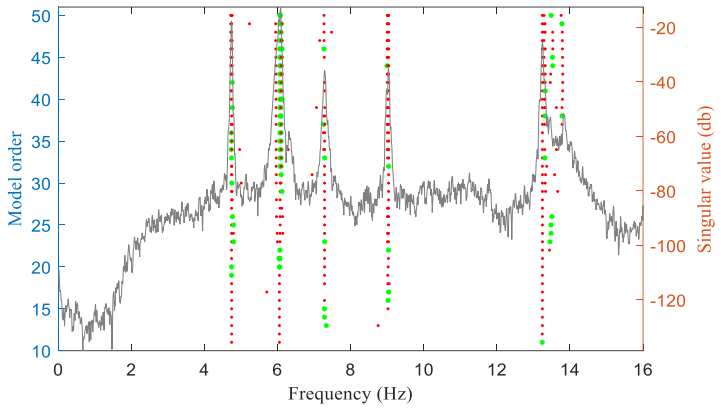
Mode	Frequency (original)	Frequency (recovered)	Error	Damping ratio (original)	Damping ratio (recovered)	Error	MAC
1st	4.74Hz	4.74Hz	0.0%	0.0016	0.0009	43.75%	0.99
2nd	6.03Hz	6.04Hz	0.2%	0.0034	0.0022	35.29%	0.99
3rd	7.24Hz	7.28Hz	0.6%	0.0023	0.0017	26.09%	0.98
4th	9.04Hz	9.04Hz	0.0%	0.0018	0.0014	22.22%	0.97
5th	13.24Hz	13.25Hz	0.1%	0.0019	0.0017	10.53%	0.98
6th	13.87Hz	13.83Hz	0.3%	0.0014	0.0025	78.57%	0.96

The original and recovered first six mode shapes of Dowling Hall Footbridge identified by using SSI method are shown in Figure 3-19. In this case, the recovered signals are reconstructed from the incomplete signals with 90% data loss. The mode shapes in Figure 3-19 match well with all the true modes. Only a minor discrepancy between the mode shapes

from original and recovered signals can be observed. The accuracies of the recovered mode shapes are further evaluated by using the Modal Assurance Criterion (MAC) and summarized in Tables 3-2 to 3-4. The worst identification is the sixth mode in Table 3-4, where the MAC value is around 96%. By observing the singular values plot of the original signals, it can be found that the sixth mode is weakly excited in this measurement period. It is also confirmed by the fact that the stable poles of the sixth mode begin to appear at the higher system orders in the stabilization diagram. It is worth noting that a MAC value of 0.96 for mode shapes of such a weakly excited mode, which are obtained from the recovered signals reconstructed based on only 10% information of the original signal, is very good. To investigate the recovery effectiveness of each channel, the accuracy of the recovered modal shape for each measurement channel is examined by using the Co-ordinate Modal Assurance Criterion (COMAC). The COMAC results for all the measurement channels under different levels of data loss are shown in Figure 3-20. Very good agreement between the original and recovered mode shapes is observed, as evidenced by the obtained COMAC values. The lowest COMAC values for the cases with 30%, 60% and 90% data loss levels are 99.6%, 98.6% and 95.5%, respectively. As consistent with MAC values, COMAC values using the recovered signals with 30% data loss are highest for all the channels among these three scenarios. Recovered mode shapes with 60% data loss are better than those with 90% data loss. In general, the proposed approach effectively recovers the lost signals for all the channels, which can be used for obtaining accurate modal identification results.

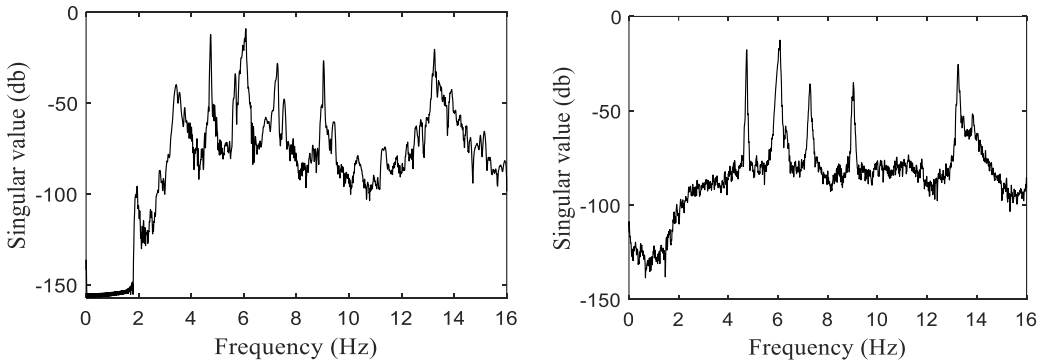


(a)



(b)

Figure 3-17. SSI results: (a) The stabilization diagram obtained from the original signals in Test data 1; (b) The stabilization diagram from the recovered signals in Test data 1 (90% data loss).



(a) (b)

Figure 3-18. FDD results: (a) The singular values plot of the original signals in Test data 1; (b) The singular values plot of the recovered signals in Test data 1 (90% data loss case).

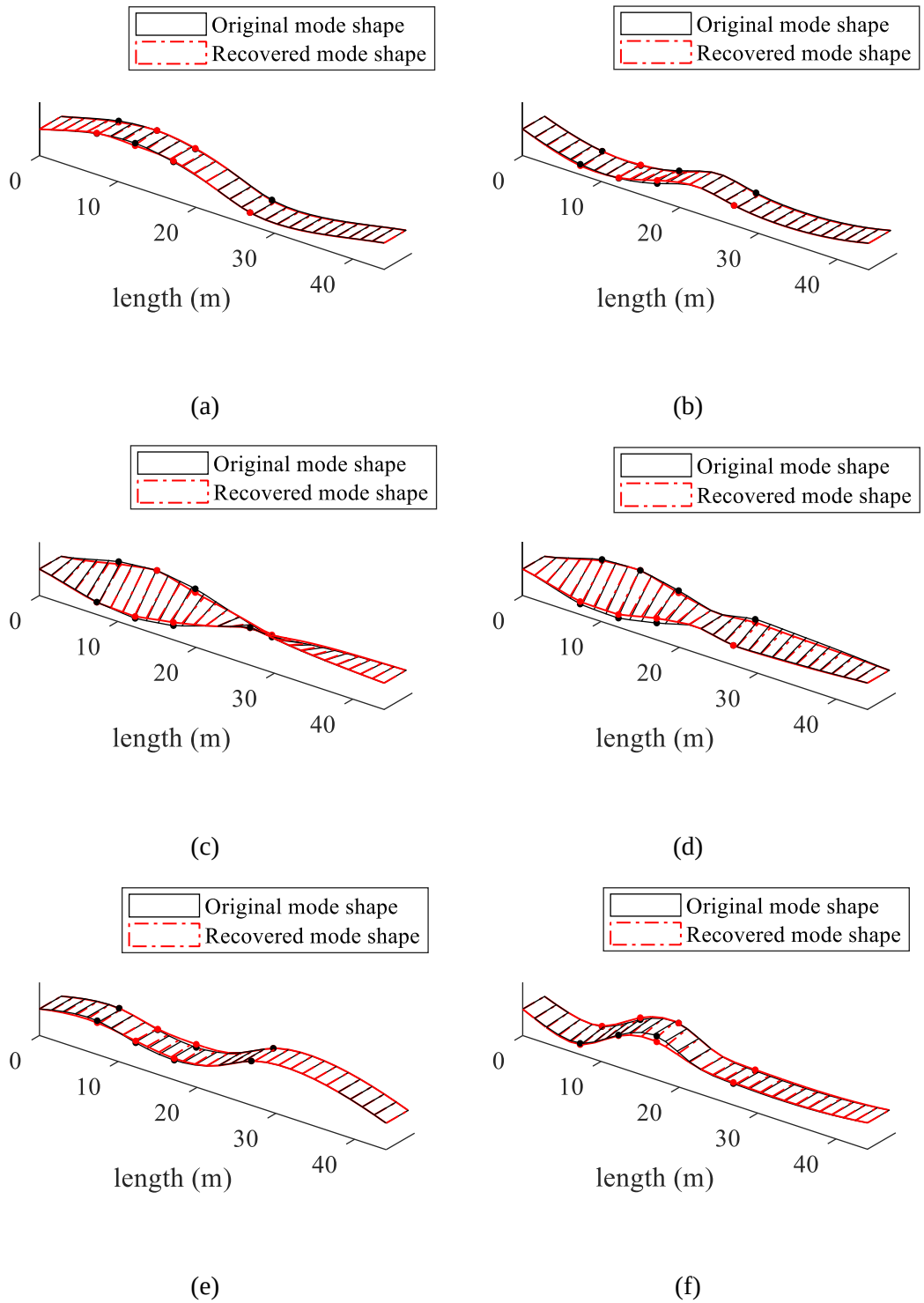


Figure 3-19. The comparison between the mode shapes extracted from the original and recovered signals originated from the incomplete signal with 90% data loss. (a) 1st mode. (b) 2nd mode. (c) 3rd mode. (d) 4th mode. (e) 5th mode. (f) 6th mode.

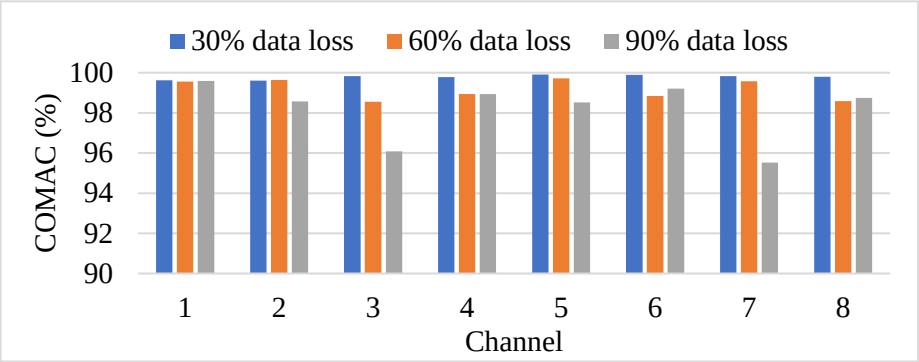


Figure 3-20. COMAC values for each channel under varying data loss ratio.

3.6 Conclusions

This chapter proposes an innovative approach based on special designed CNN for recovering the lost SHM data during the transmission process. The architecture and detailed parameters of the proposed CNN, as well as the entire procedure to implement the lost data recovery are elaborated. The accuracy and robustness of the proposed method are investigated by two case studies, where one experimental study based on data from a single measurement channel and another one based on all eight channels of a SHM system installed on the Dowling Hall Footbridge. The testing results demonstrate that the recovery error of the time domain signal increases with the data loss ratio. However, the increment of error can be mainly contributed to the growing difference of the noise components between the original and recovered signals. A comparison study is conducted with different sampling rates. The results demonstrate that the proposed approach can accurately recover signals with a low sampling rate, as long as the excited vibration modes are covered in the target frequency range of interest as half of the sampling rate. The data lost recovery results and modal identification accuracy using the recovered signals demonstrate that the proposed CNN has a decent performance to effectively recover the lost data in SHM, even with 90% data loss ratio. In the testing to validate the performance of the proposed approach, the trained CNNs recover the 5-minutes long testing data instantly. However, it should be noted that there is no specific length limit for the signal served as the input. The testing data are measured under different ambient conditions compared with the training data, but the recovered signals still present a high degree

of agreement with the true complete ones. The proposed approach can be used for recovering the new measured incomplete data, which are not included in the training datasets.

3.7 References

Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M., & Inman, D. J. (2017). Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration, 388*, 154-170.

Aytar, Y., Vondrick, C., & Torralba, A. (2016). *Soundnet: Learning sound representations from unlabeled video*. Paper presented at the Advances in Neural Information Processing Systems, Barcelona, Spain.

Bao, Y., Li, H., Sun, X., Yu, Y., & Ou, J. (2013). Compressive sampling–based data loss recovery for wireless sensor networks used in civil structural health monitoring. *Structural Health Monitoring, 12*(1), 78-95.

Bao, Y., Shi, Z., Wang, X., & Li, H. (2018). Compressive sensing of wireless sensors based on group sparse optimization for structural health monitoring. *Structural Health Monitoring, 17*(4), 823-836.

Brincker, R., Zhang, L., & Andersen, P. (2001). Modal identification of output-only systems using frequency domain decomposition. *Smart Materials and Structures, 10*(3), 441.

Cha, Y. J., Choi, W., & Büyüköztürk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering, 32*(5), 361-378.

Chen, Z., & Casciati, F. (2014). A low-noise, real-time, wireless data acquisition system for structural monitoring applications. *Structural Control and Health Monitoring, 21*(7), 1118-1136.

Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence, 38*(2), 295-307.

He, J., Guan, X., & Liu, Y. (2012). Structural response reconstruction based on empirical mode decomposition in time domain. *Mechanical Systems and Signal Processing, 28*, 348-366.

Kane, M., Zhu, D., Hirose, M., Dong, X., Winter, B., Häckell, M., Lynch, J. P., Wang, Y., & Swartz, A. (2014). *Development of an extensible dual-core wireless sensing node for cyber-physical systems.* Paper presented at the Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2014.

Kim, S., Pakzad, S., Culler, D., Demmel, J., Fenves, G., Glaser, S., & Turon, M. (2007). *Health monitoring of civil infrastructures using wireless sensor networks.* Paper presented at the Proceedings of the 6th international conference on Information processing in sensor networks.

Kuleshov, V., Enam, S. Z., & Ermon, S. (2017). Audio super resolution using neural networks. *https://arxiv.org/abs/1708.00853*.

Law, S., Li, J., & Ding, Y. (2011). Structural response reconstruction with transmissibility concept in frequency domain. *Mechanical Systems and Signal Processing, 25*(3), 952-968.

Lin, Y. Z., Nie, Z. H., & Ma, H. W. (2017). Structural damage detection with automatic feature-extraction through deep learning. *Computer-Aided Civil and Infrastructure Engineering, 32*(12), 1025-1046.

Linderman, L. E., Mechitov, K. A., & Spencer Jr, B. F. (2013). TinyOS-based real-time wireless data acquisition framework for structural health monitoring and control. *Structural Control and Health Monitoring, 20*(6), 1007-1020.

Long, J., Shelhamer, E., & Darrell, T. (2015). *Fully convolutional networks for semantic segmentation.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.

Lynch, J. P. (2005). Design of a wireless active sensing unit for localized structural health monitoring. *Structural Control and Health Monitoring, 12*(3-4), 405-423.

Lynch, J. P., & Loh, K. J. (2006). A summary review of wireless sensors and sensor networks for structural health monitoring. *Shock and Vibration Digest, 38*(2), 91-130.

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). *Rectifier nonlinearities improve neural network acoustic models.* Paper presented at the International conference on machine learning, Atlanta, USA.

Mao, X., Shen, C., & Yang, Y. B. (2016). *Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections.* Paper presented at the Advances in neural information processing systems, Barcelona, Spain.

Moser, P., & Moaveni, B. (2011). Environmental effects on the identified natural frequencies of the Dowling Hall Footbridge. *Mechanical Systems and Signal Processing, 25*(7), 2336-2357.

Moser, P., & Moaveni, B. (2013). Design and deployment of a continuous monitoring system for the dowling hall footbridges. *Experimental Techniques, 37*(1), 15-26.

Nagayama, T., Sim, S.-H., Miyamori, Y., & Spencer Jr, B. (2007). Issues in structural health monitoring employing smart sensors. *Smart Structures and Systems, 3*(3), 299-320.

Nair, V., & Hinton, G. E. (2010). *Rectified linear units improve restricted boltzmann machines.* Paper presented at the Proceedings of the 27th international conference on machine learning, Haifa, Israel.

Ni, Y. Q., Li, B., Lam, K. H., Zhu, D. P., Wang, Y., Lynch, J. P., & Law, K. H. (2011). In-construction vibration monitoring of a super-tall structure using a long-range wireless sensing system. *Smart Structures and Systems, 7*(2), 83-102.

Ni, Y. Q., Xia, Y., Liao, W. Y., & Ko, J. M. (2009). Technology innovation in developing the structural health monitoring system for Guangzhou New TV Tower. *Structural Control and Health Monitoring, 16*(1), 73-98.

Odena, A., Dumoulin, V., & Olah, C. (2016). Deconvolution and checkerboard artifacts. *Distill, 1*(10), e3.

Pathirage, C. S. N., Li, J., Li, L., Hao, H., Liu, W., & Ni, P. (2018). Structural damage identification based on autoencoder neural networks and deep learning. *Engineering Structures, 172*, 13-28.

Pathirage, C. S. N., Li, J., Li, L., Hao, H., Liu, W., & Wang, R. (2019). Development and application of a deep learning–based sparse autoencoder framework for structural damage identification. *Structural Health Monitoring, 18*(1), 103-122.

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-net: convolutional networks for biomedical image segmentation.* Paper presented at the International conference on medical image computing and computer-assisted intervention, Munich, Germany.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., & Wang, Z. (2016). *Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, USA.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*(1), 1929-1958.

Sun, Z., Li, B., Dyke, S. J., Lu, C., & Linderman, L. (2016). Benchmark problem in active structural control with wireless sensor network. *Structural Control and Health Monitoring, 23*(1), 20-34.

Van Overschee, P., & De Moor, B. (2012). *Subspace identification for linear systems: Theory—Implementation—Applications*: Springer Science & Business Media.

Wan, H.-P., & Ni, Y.-Q. (2019). Bayesian multi-task learning methodology for reconstruction of structural health monitoring data. *Structural Health Monitoring, 18*(4), 1282-1309.

Winter, B. D., & Swartz, R. A. (2017). Wireless structural control using multi-step TDMA communication patterning bandwidth allocation. *Structural Control and Health Monitoring, 24*(12), e2025.

Yeh, R. A., Chen, C., Lim, T.-Y., Schwing, A. G., Hasegawa-Johnson, M., & Do, M. N. (2017). *Semantic image inpainting with deep generative models.* Paper presented at the The IEEE conference on computer vision and pattern recognition, Honolulu, USA.

Zhang, Y., Kurata, M., & Lynch, J. P. (2017). Long-term modal analysis of wireless structural monitoring data from a suspension bridge under varying environmental and operational conditions: system design and automated modal analysis. *Journal of Engineering Mechanics, 143*(4), 04016124.

CHAPTER 4 VIBRATION SIGNAL DENOISING FOR STRUCTURAL HEALTH MONITORING BY RESIDUAL CONVOLUTIONAL NEURAL NETWORKS

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 chapter 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.

This chapter was published in *Measurement* with the full bibliographic citation as follows: Fan G, Li J, Hao H. (2020). Vibration signal denoising for structural health monitoring by residual convolutional neural networks. *Measurement*, 107651. https://doi.org/10.1016/j.measurement.2020.107651.

4.1 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 (Cross et al., 2013; Xia et al., 2013) and high rise buildings (Nguyen et al., 2015; Ni et al., 2009). 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 (Li & Law, 2012; Li et al., 2012) 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 (He et al., 2014). Conventional time domain averaging methods are more suitable for periodic signals (Braun, 2011). 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 (Bi et al., 2019; Lin & Qu, 2000). 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 (Chegini et al., 2019; He et al., 2014). Other methods based on singular value decomposition (Zhao & Jia, 2017), empirical mode decomposition (Lei et al., 2013) and global projection (Hou et al., 2011) 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) (Vincent et al., 2010) and Convolutional Neural Networks (CNN) (Krizhevsky et al., 2012) have been extended to denoise image (Jain & Seung, 2009; Xie et al., 2012), audio (Park &

Lee, 2016; Xia & Bao, 2013) and medical domain signals (Poungponsri & Yu, 2013) 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 (Zhao et al., 2019) such as damage and crack detection using images (Cha et al., 2017; Jang et al., 2019; Xu et al., 2019), vibration signals (Abdeljaber et al., 2017; Guo et al., 2016; Jing et al., 2017; Lin et al., 2017; San Martin et al., 2018; Zhang et al., 2019) and vibrational characteristics (Pathirage et al., 2018; Pathirage et al., 2019), abnormal measurement diagnosis (Bao et al., 2019) and vibration lost data recovery (Fan et al., 2019). 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 chapter 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 chapter, 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 4.2. 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 4.3. To further verify the robustness of the proposed method, Section 4.4 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 4.5.

4.2 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 (Kuleshov et al., 2017) is a fully convolutional feedforward network. The developed ResNet has a bottleneck structure which originates from Deep Auto-Encoder (Hinton & Salakhutdinov, 2006) and is named as Convolutional Encoder-Decoder Network (Badrinarayanan et al., 2017). 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 chapter 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.

4.2.1 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 4-1. 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 4-1. 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 (Isola et al., 2017). 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 (Mao et al., 2016). 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 4-1. The kernel number and kernel size are initialized similarly as Ref. (Kuleshov et al., 2017) 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 4-1 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 4.2.4). 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 4-1 and the input and output shapes of each layer are listed in Table 4-1, 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 4-1. 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)	-

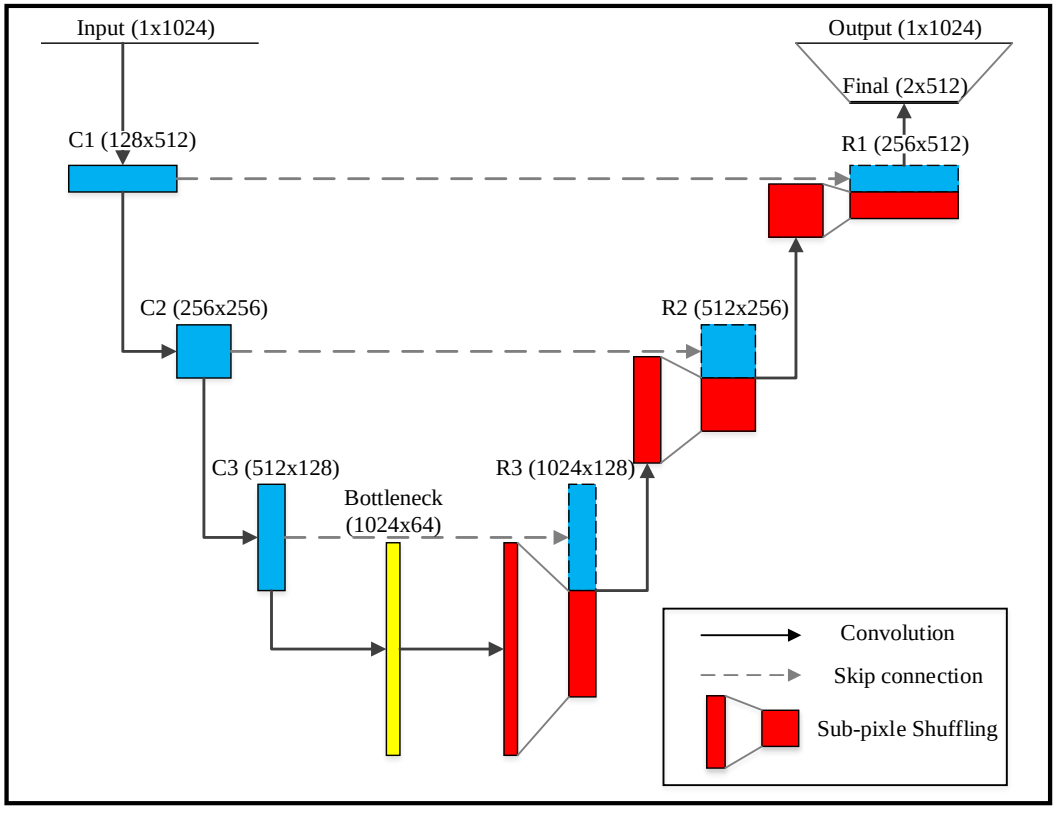


Figure 4-1. The architecture of the used ResNet with a bottleneck structure.

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 4.2.3. 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 4.2.4; 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

$$Loss = \frac{1}{N}\sum_{k=1}^{N}\left\|y_k - f_\phi(x_k)\right\|_2^2 \tag{4.1}$$

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.

4.2.2 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 (Srivastava et al., 2014). 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 (Xu et al.,

2015). Each neuron has an independent probability *p* to be dropped. *p*=0.5 is suggested by the inventors and used in this study (Srivastava et al., 2014).

4.2.3 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 (Maas et al., 2013) as an improved version of ReLU (Nair & Hinton, 2010) is an advanced activation function. Leaky ReLU has a nonlinear gradient over its whole domain as follows

$$f(x) = \begin{cases} x, & if\ x > 0 \\ 0.01x, & \text{otherwise} \end{cases} \tag{4.2}$$

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 (Maas et al., 2013).

4.2.4 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 4-1, 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 4-2 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 (Shi et al., 2016). 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 (Odena et al., 2016).

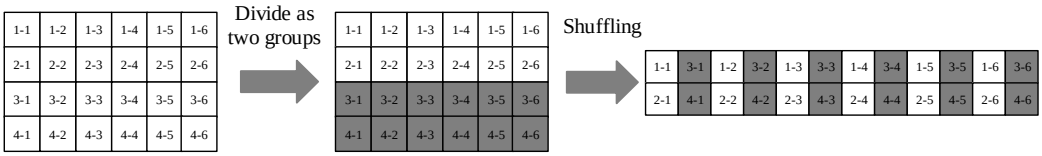


Figure 4-2. An example of the one-dimensional sub-pixel shuffling with 4 feature maps and 6 elements in each map.

4.2.5 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,

$$\mathrm{SNR} = 10 \times \log_{10} \frac{P_{signal}}{P_{noise}} = 10 \times \log_{10} \frac{\sum_1^n {y_i}^2}{\sum_1^n (y_i - y_i')^2} \tag{4.3}$$

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

$$\xi = \frac{\|f - f'\|_2^2}{\|f\|_2^2} = \frac{\sqrt{\sum_1^n (f_i - f_i')^2}}{\sqrt{\sum_1^n {f_i}^2}} \tag{4.4}$$

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) (Brincker et al., 2001), 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.

4.3. Experimental validations

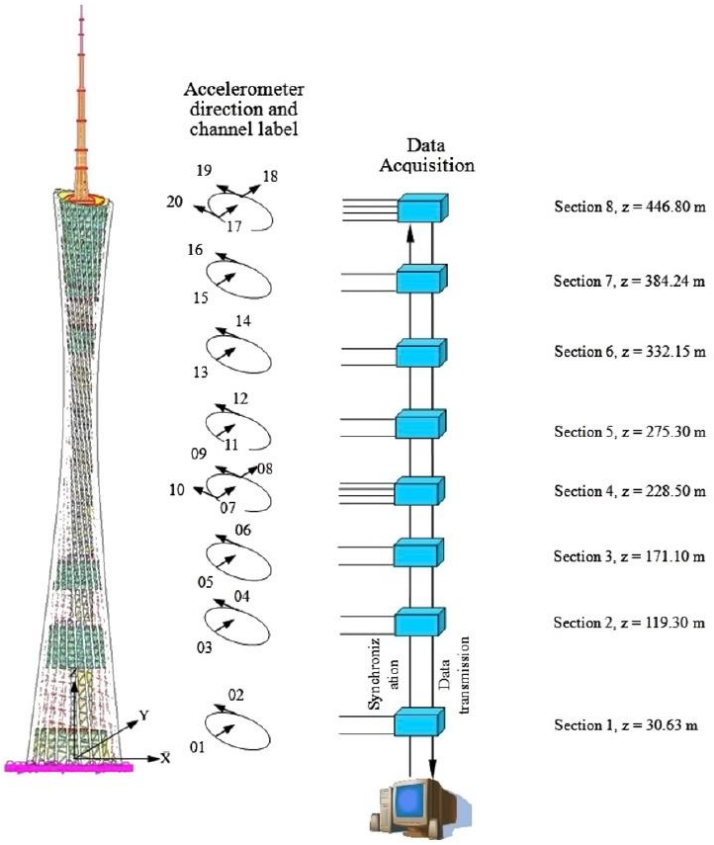
To demonstrate the effectiveness and robustness of using the proposed approach based on ResNet for vibration signal denoising, experimental studies using the in-situ acceleration data measured from a long-term SHM system on Guangzhou New TV Tower are conducted. The vibration signals of Guangzhou New 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.

4.3.1 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 4-3(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 4-3(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. (Ni et al., 2009). The previous studies have shown the effectiveness of the SHM system (Chen et al., 2011; Xia et al., 2011) and it has become a benchmark platform for high-rise structures.

(a)



(b)

Figure 4-3. (a) Guangzhou New TV Tower; (b) Sensor deployment and data acquisition (Chen et al., 2011).

4.3.2 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 (Chen et al., 2011) 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 (Zhang et al., 2016). 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

$$Signal_{noisy} = Signal_{clean} + Noise \times N_l \tag{4.5}$$

where *Noise* is a normally distributed random vector with zero mean and standard deviation of the original vibration signals (Li & Law, 2009), 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

$$accleration_{normalized} = \frac{accleration}{max\ (|accleration|)} \tag{4.6}$$

The paired normalized original signal and noisy signals are depicted in Figure 4-4. 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.

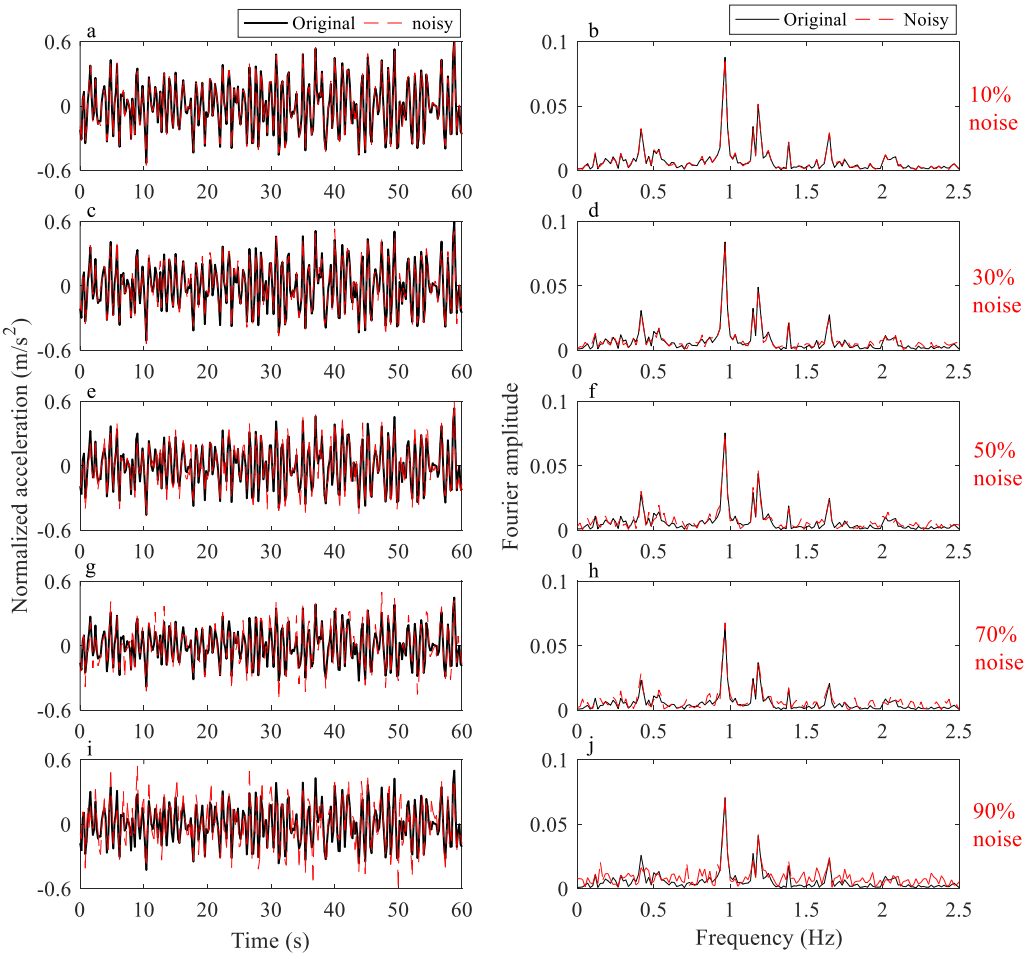


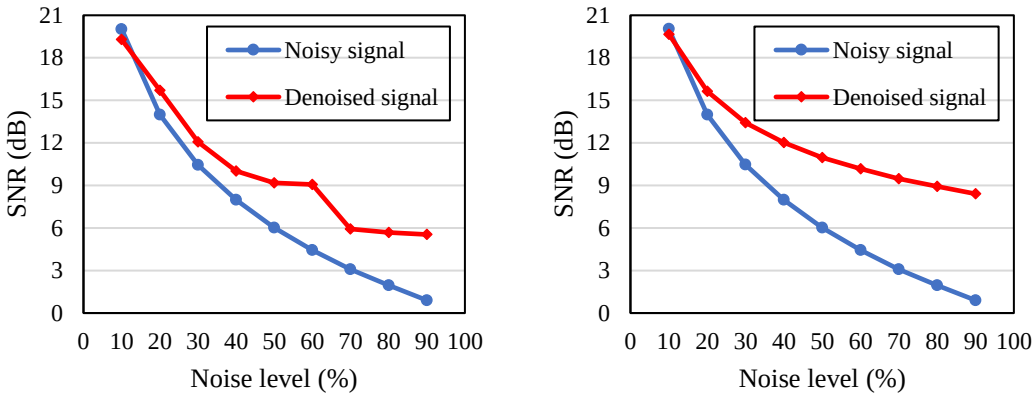
Figure 4-4. Original and noisy signals with different noise levels. In the time domain: (a) 10%, (c) 30%, (e) 50%, (g) 70% and (i) 90% and in the frequency domain: (b) 10%, (d) 30%, (f) 50%, (h) 70%, (j) 90%.

4.3.3 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.

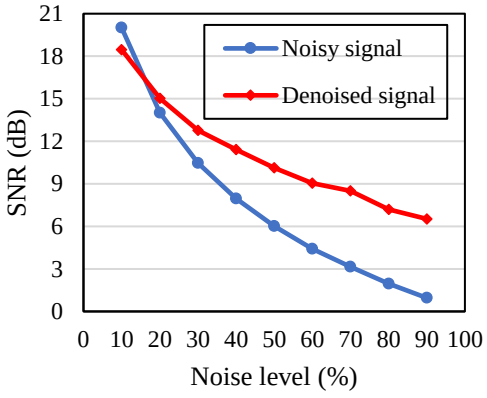
4.3.3.1 Evaluation of results in time and frequency domains

SNRs of the input noisy and output generated denoised signals are computed based on Equation (4,3). 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 4-5. 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 (4.4). The results as shown in Figure 4-6 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 4-6(a) – (c).

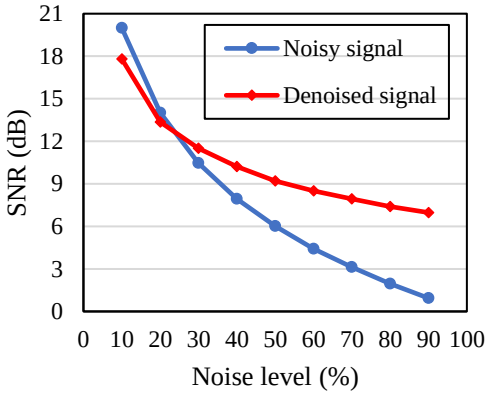


(a)

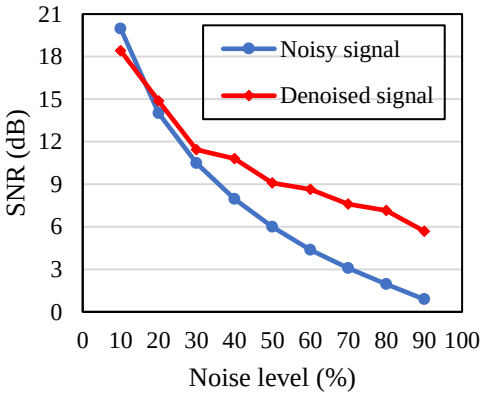
(b)



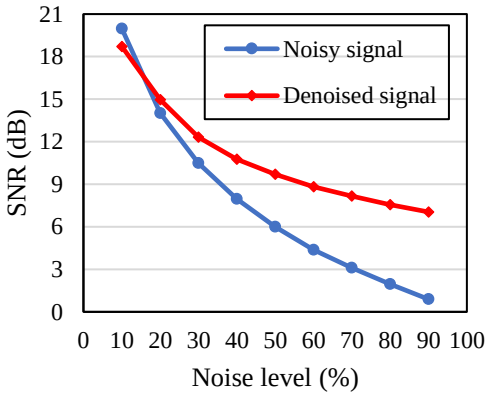
(c)



(d)



(e)



(f)

Figure 4-5. 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.

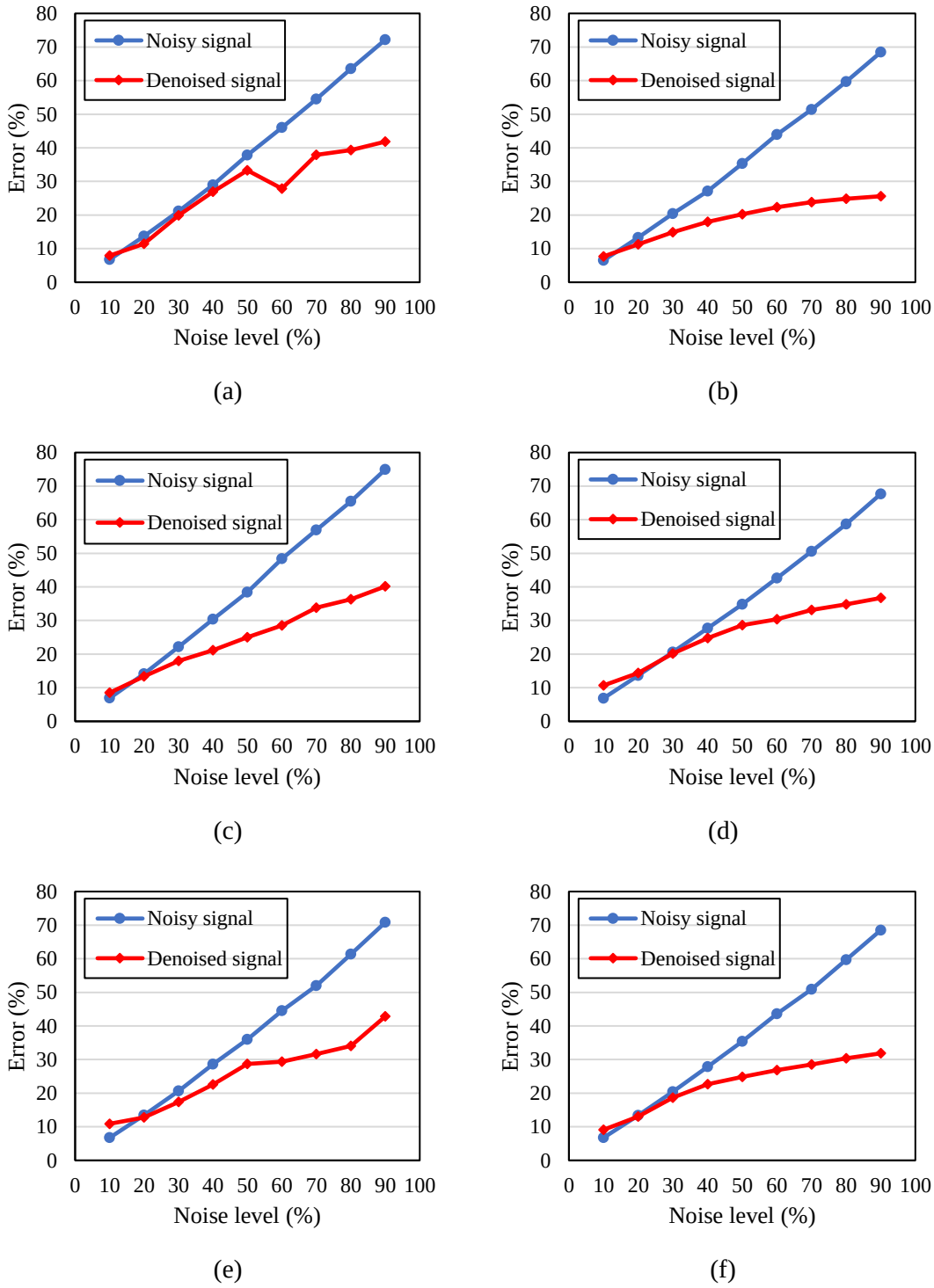


Figure 4-6. Comparison of errors of noisy and denoised signals in the frequency domain. Results for short-axis direction: (a) Testing data 1; (c) Testing data 2 and (e) Testing data 3. Long-axis direction: (b) Testing data 1; (d) Testing data 2 and (f) Testing data 3.

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 4-7. 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 (Lin et al., 2017) 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.

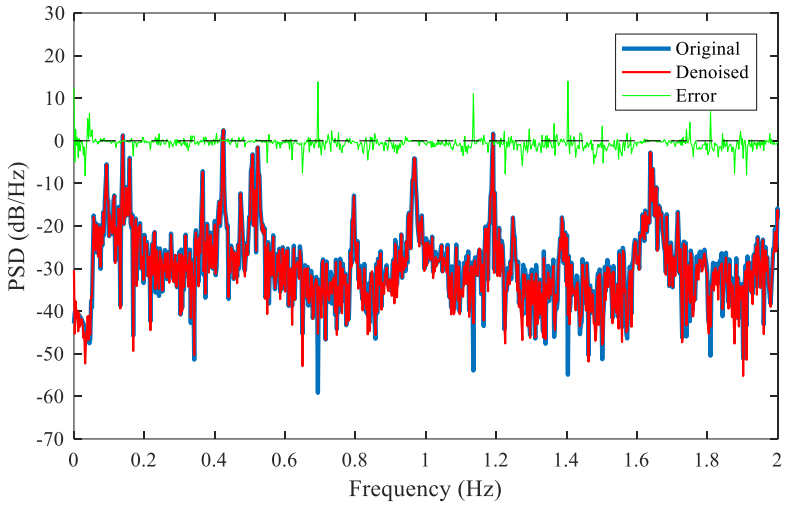


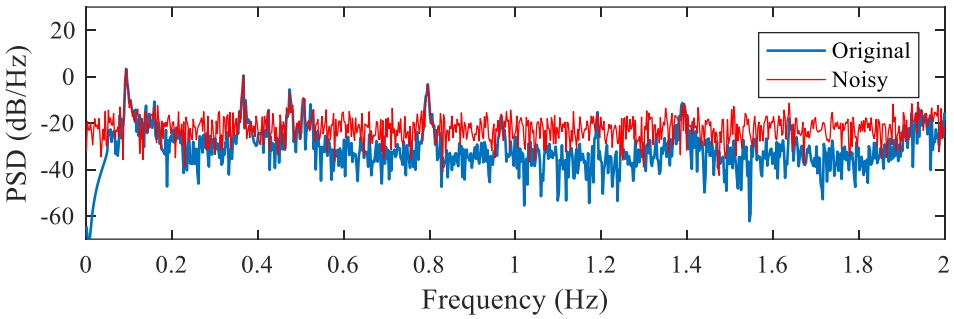
Figure 4-7. PSD spectra of the pre-processed original and denoised signals, and the errors in PSD.

4.3.3.2 Comparison of results with wavelet transform based method

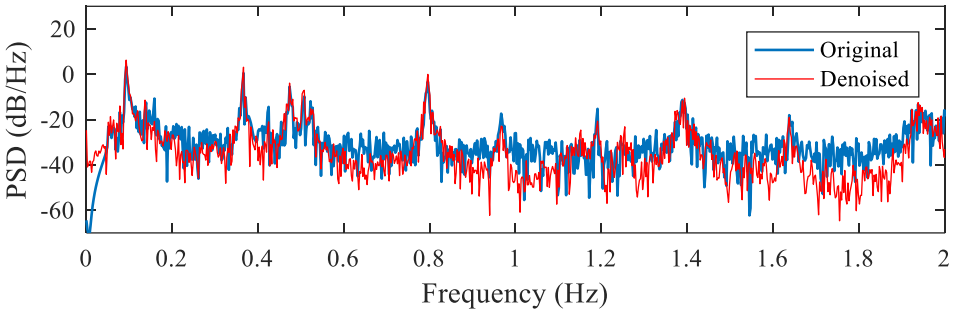
Wavelet transform based method is one of the most popular and most researched methods for vibration signal denoising in the current literature, therefore it is selected for comparing the performance with the proposed method. Wavelet transform decomposes signals into different scales by using a selected 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 (Johnstone & Silverman, 2005). 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 4-8. Figures 4-8(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 4-8(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 4-6, 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.

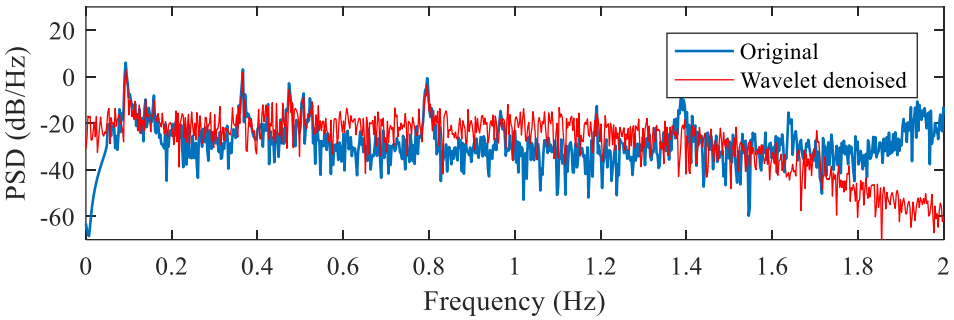
By observing the PSDs of the denoised signals by using the trained ResNet and wavelet transform based method as shown in Figures 4-8(c), (d), (e) and (f), it is obvious that the proposed approach outperforms the wavelet transform based method on denoising vibration signals. Wavelet transform has a minor 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.


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

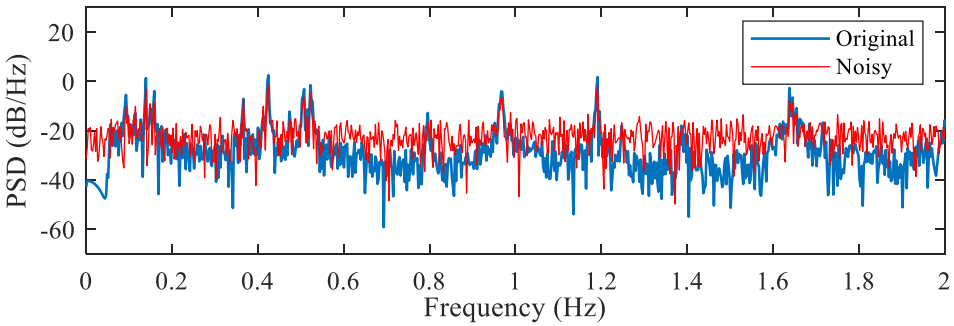
(a)


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

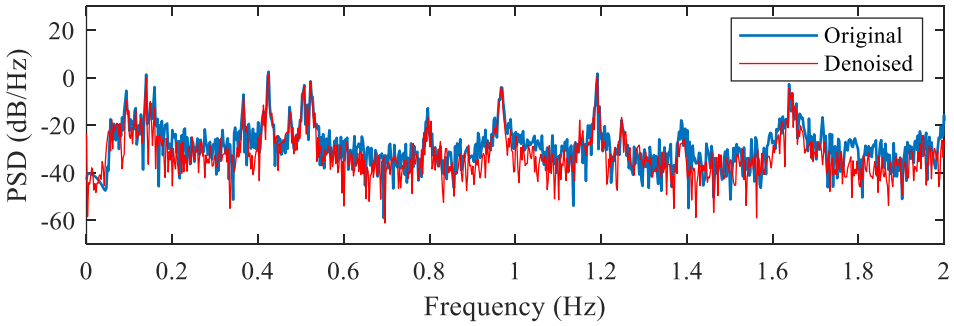
(b)


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

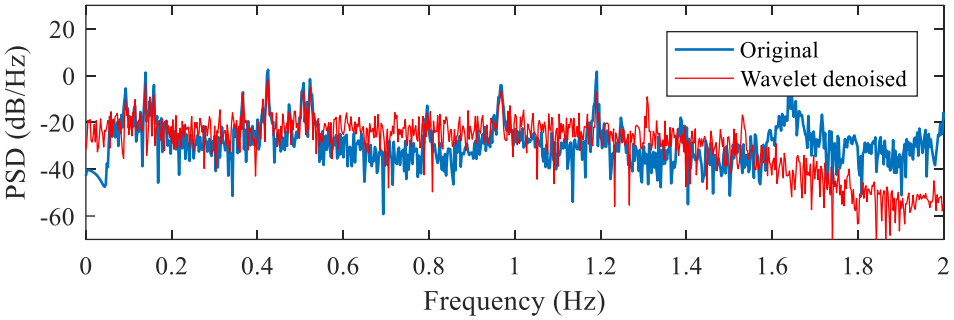
(c)


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

(d)



(e)



(f)

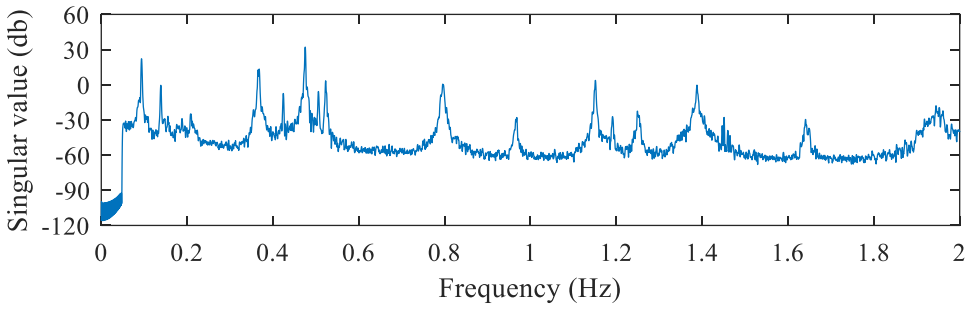
Figure 4-8. Comparison of PSD spectra of signals in Testing data 1, Short-axis direction: (a) Original and noisy signals with 90% noise; (b) Original and ResNet denoised signals; (c) Original and wavelet denoised signals. Long-axis direction: (c) Original and noisy signals with 90% noise; (d) Original and ResNet denoised signals. (f) Original and wavelet denoised signals.

4.3.4 Modal analysis of the testing results using denoised signals

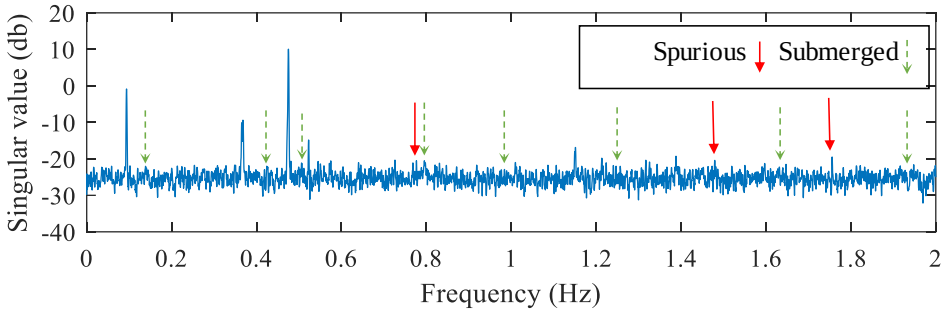
4.3.4.1 Quantitative evaluation of the denoised signals

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 4-9 and 4-10 show the identification results of original, noisy and denoised signals in the short and long-axis directions using FDD. Figures 4-9(a) and 4-10(a) demonstrate the high quality of the original signals, where all the 15 physical modes can be identified and few spurious modes are

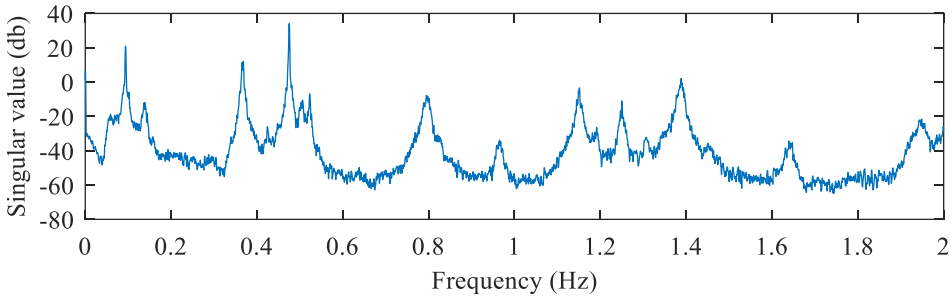
observed. Figures 4-9(b) and 10(b) show that under such strong noise, the usability of signals is significantly affected with most of the modes submerged by the noise except several strongly excited modes as marked with green dash arrows. In addition, the noise effect at some frequencies is very pronounced as marked with the red arrows in Figures 4-9(b) and 4-10(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 4-9(c) and 4-10(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 4-9(b) and (c) and Figures 4-10(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.



(a)

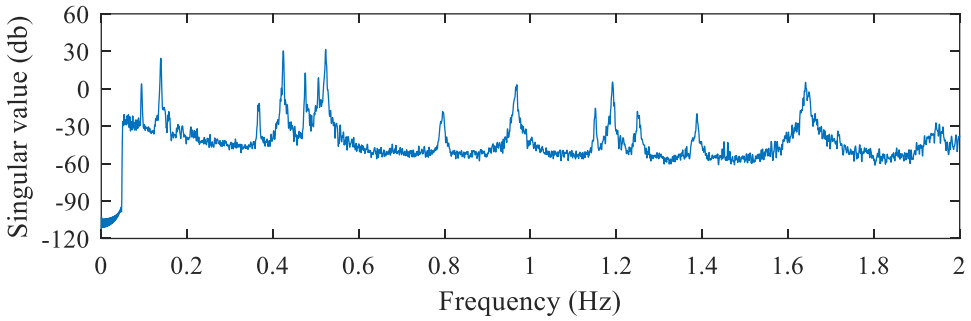


(b)

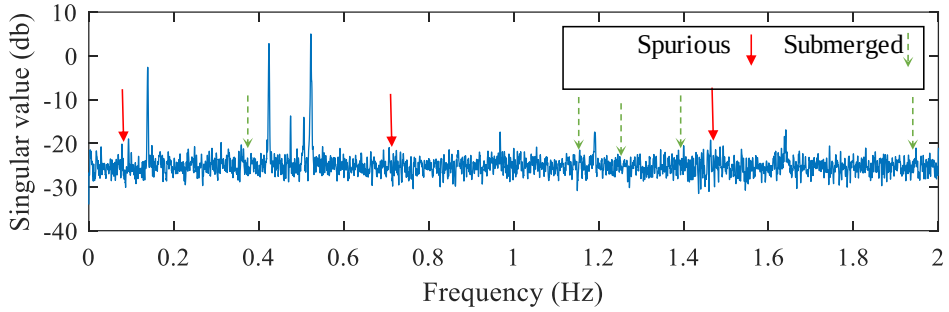


(c)

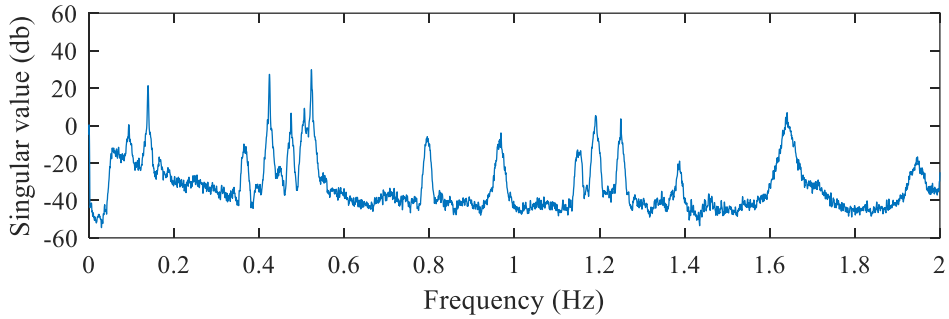
Figure 4-9. Singular value spectra of: (a) original; (b) noisy; and (c) denoised signals of Testing data 1 in the short-axis direction.



(a)



(b)



(c)

Figure 4-10. Singular value spectra of: (a) original; (b) noisy; and (c) denoised signals of Testing data 1 in the long-axis direction.

4.3.4.2 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 4-2 and 4-3 for short and long-axis directions, respectively. Modal identification results using the original signals show good consistency with those in Ref. (Zhang et al., 2016). 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 4-2 and 4-3. 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 4-2 and 4-3, 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 4-2. 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 4-3. 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 4-11 and 4-12, 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 4-11 and 4-12, 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.

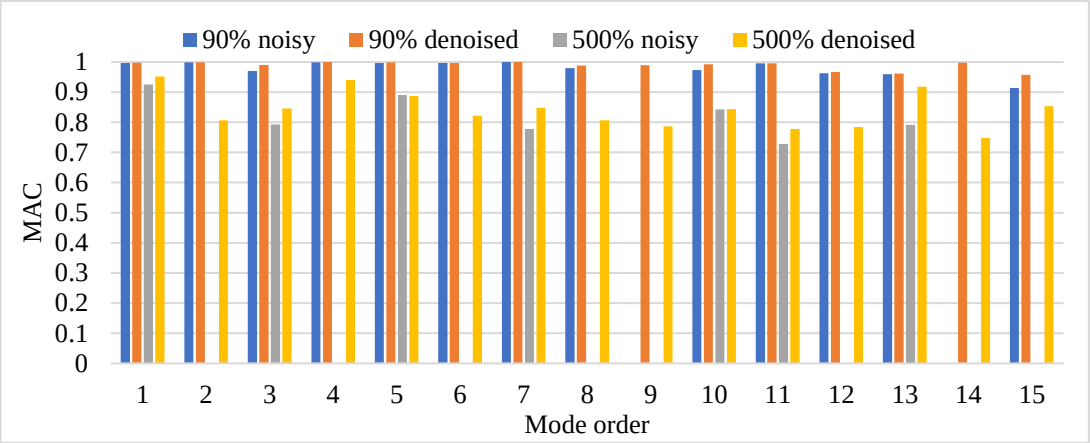


Figure 4-11. MAC values of the identified mode shapes in the short-axis direction.

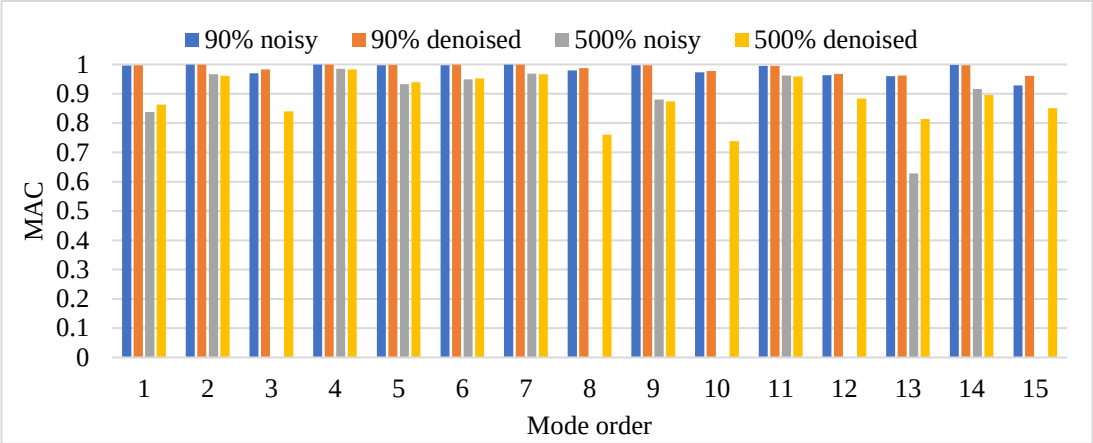


Figure 4-12. MAC values of the identified mode shapes in the long-axis direction.

4.4 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,

$$PSD_{pink\ noise} \propto \frac{1}{f} \tag{4.7}$$

The means to generate a noisy signal with pink noise is similar to that of white noise using Equation (4.5). The difference is that *Noise* vector becomes a time sequence of pink noise with zero mean and standard deviation of the original signal.

Figure 4-13 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 4-14. 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.

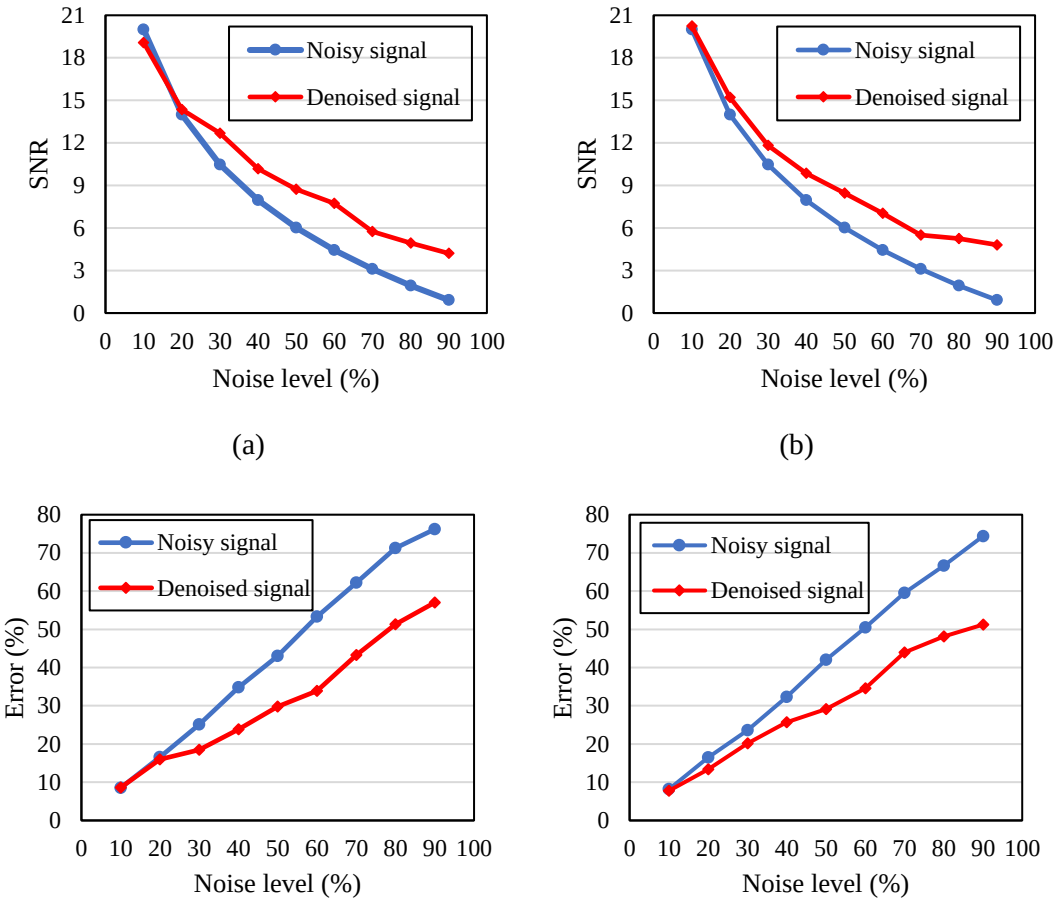
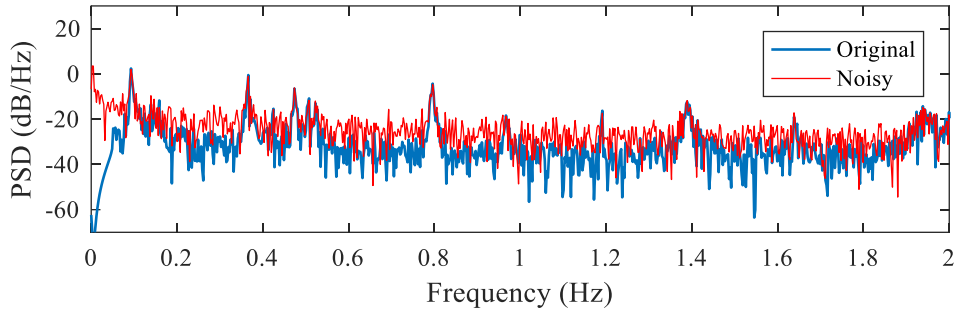
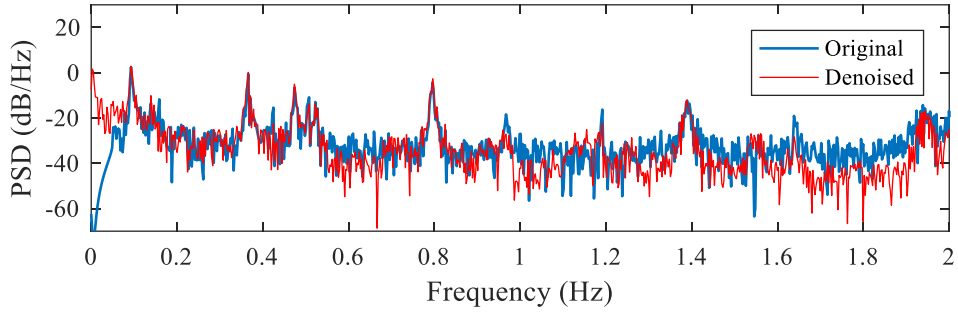


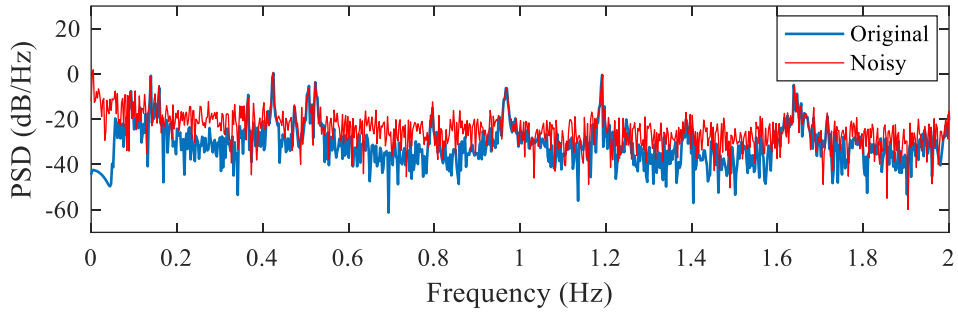
Figure 4-13. Comparison of errors of the noisy and denoised signals in the time and frequency domains. Results for short-axis direction in: (a) Time domain and (c) Frequency domain; Long-axis direction in: (b) Time domain and (d) Frequency domain.



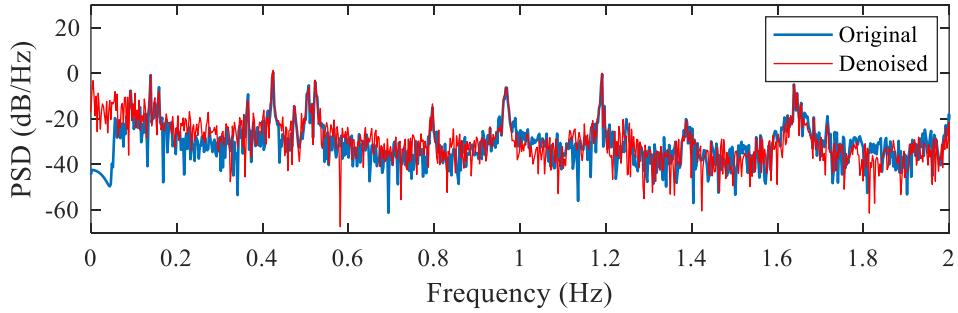
(a)



(b)



(c)



(d)

Figure 4-14. 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.

4.5 Conclusions

This chapter 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.

4.6 References

Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M., & Inman, D. J. (2017). Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration, 388,* 154-170.

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence, 39*(12), 2481-2495.

Bao, Y., Tang, Z., Li, H., & Zhang, Y. (2019). Computer vision and deep learning–based data anomaly detection method for structural health monitoring. *Structural Health Monitoring, 18*(2), 401-421.

Bi, F., Ma, T., & Wang, X. (2019). Development of a novel knock characteristic detection method for gasoline engines based on wavelet-denoising and EMD decomposition. *Mechanical Systems and Signal Processing, 117,* 517-536.

Braun, S. (2011). The synchronous (time domain) average revisited. *Mechanical Systems and Signal Processing, 25*(4), 1087-1102.

Brincker, R., Zhang, L., & Andersen, P. (2001). Modal identification of output-only systems using frequency domain decomposition. *Smart Materials and Structures, 10*(3), 441.

Cha, Y. J., Choi, W., & Büyüköztürk, O. (2017). Deep learning-based crack damage detection using convolutional neural networks. *Computer-Aided Civil and Infrastructure Engineering, 32*(5), 361-378.

Chegini, S. N., Bagheri, A., & Najafi, F. (2019). Application of a new EWT-based denoising technique in bearing fault diagnosis. *Measurement, 144*, 275-297.

Chen, W., Lu, Z., Lin, W., Chen, S., Ni, Y., Xia, Y., & Liao, W. (2011). Theoretical and experimental modal analysis of the Guangzhou New TV Tower. *Engineering Structures, 33*(12), 3628-3646.

Cross, E., Koo, K., Brownjohn, J., & Worden, K. (2013). Long-term monitoring and data analysis of the Tamar Bridge. *Mechanical Systems and Signal Processing, 35*(1-2), 16-34.

Fan, G., Li, J., & Hao, H. (2019). Lost data recovery for structural health monitoring based on convolutional neural networks. *Structural Control and Health Monitoring, 26*(10), e2433.

Guo, X., Chen, L., & Shen, C. (2016). Hierarchical adaptive deep convolution neural network and its application to bearing fault diagnosis. *Measurement, 93*, 490-502.

He, Q., Wang, X., & Zhou, Q. (2014). Vibration sensor data denoising using a time-frequency manifold for machinery fault diagnosis. *Sensors, 14*(1), 382-402.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *science, 313*(5786), 504-507.

Hou, S., Liang, M., & Li, Y. (2011). An optimal global projection denoising algorithm and its application to shaft orbit purification. *Structural Health Monitoring, 10*(6), 603-616.

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). *Image-to-image translation with conditional adversarial networks.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, USA.

Jain, V., & Seung, S. (2009). *Natural image denoising with convolutional networks.* Paper presented at the Advances in Neural Information Processing Systems, Vancouver, Canada.

Jang, K., Kim, N., & An, Y.-K. (2019). Deep learning–based autonomous concrete crack evaluation through hybrid image scanning. *Structural Health Monitoring, 18*(5-6), 1722-1737.

Jing, L., Zhao, M., Li, P., & Xu, X. (2017). A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox. *Measurement, 111*, 1-10.

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

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). *Imagenet classification with deep convolutional neural networks.* Paper presented at the Advances in neural information processing systems, Lake Tahoe, USA.

Kuleshov, V., Enam, S. Z., & Ermon, S. (2017). Audio super resolution using neural networks. *https://arxiv.org/abs/1708.00853*.

Lei, Y., Lin, J., He, Z., & Zuo, M. J. (2013). A review on empirical mode decomposition in fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing, 35*(1-2), 108-126.

Li, J., & Law, S. (2012). Damage identification of a target substructure with moving load excitation. *Mechanical Systems and Signal Processing, 30*, 78-90.

Li, J., Law, S., & Ding, Y. (2012). Substructure damage identification based on response reconstruction in frequency domain and model updating. *Engineering Structures, 41*, 270-284.

Li, X. Y., & Law, S. S. (2009). Identification of structural damping in time domain. *Journal of Sound and Vibration, 328*(1-2), 71-84.

Lin, J., & Qu, L. (2000). Feature extraction based on Morlet wavelet and its application for mechanical fault diagnosis. *Journal of Sound and Vibration, 234*(1), 135-148.

Lin, Y. Z., Nie, Z. H., & Ma, H. W. (2017). Structural damage detection with automatic feature-extraction through deep learning. *Computer-Aided Civil and Infrastructure Engineering, 32*(12), 1025-1046.

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). *Rectifier nonlinearities improve neural network acoustic models.* Paper presented at the International conference on machine learning, Atlanta, USA.

Mao, X., Shen, C., & Yang, Y. B. (2016). *Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections.* Paper presented at the Advances in neural information processing systems, Barcelona, Spain.

Nair, V., & Hinton, G. E. (2010). *Rectified linear units improve restricted boltzmann machines.* Paper presented at the Proceedings of the 27th international conference on machine learning, Haifa, Israel.

Nguyen, T., Chan, T. H., Thambiratnam, D. P., & King, L. (2015). Development of a cost-effective and flexible vibration DAQ system for long-term continuous structural health monitoring. *Mechanical Systems and Signal Processing, 64*, 313-324.

Ni, Y. Q., Xia, Y., Liao, W. Y., & Ko, J. M. (2009). Technology innovation in developing the structural health monitoring system for Guangzhou New TV Tower. *Structural Control and Health Monitoring, 16*(1), 73-98.

Odena, A., Dumoulin, V., & Olah, C. (2016). Deconvolution and checkerboard artifacts. *Distill, 1*(10), e3.

Park, S. R., & Lee, J. (2016). A fully convolutional neural network for speech enhancement. *https://arxiv.org/abs/1609.07132*.

Pathirage, C. S. N., Li, J., Li, L., Hao, H., Liu, W., & Ni, P. (2018). Structural damage identification based on autoencoder neural networks and deep learning. *Engineering Structures, 172*, 13-28.

Pathirage, C. S. N., Li, J., Li, L., Hao, H., Liu, W., & Wang, R. (2019). Development and application of a deep learning–based sparse autoencoder framework for structural damage identification. *Structural Health Monitoring, 18*(1), 103-122.

Poungponsri, S., & Yu, X.-H. (2013). An adaptive filtering approach for electrocardiogram (ECG) signal noise reduction using neural networks. *Neurocomputing, 117*, 206-213.

San Martin, G., López Droguett, E., Meruane, V., & das Chagas Moura, M. (2018). Deep variational auto-encoders: a promising tool for dimensionality reduction and ball bearing elements fault diagnosis. *Structural Health Monitoring, 18*(4), 1092-1128.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., & Wang, Z. (2016). *Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, USA.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*(1), 1929-1958.

Vincent, P., Larochelle, H., Lajoie, I., Bengio, Y., & Manzagol, P.-A. (2010). Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research, 11*(Dec), 3371-3408.

Xia, B., & Bao, C. (2013). *Speech enhancement with weighted denoising auto-encoder.* Paper presented at the Interspeech, Lyon, France.

Xia, Y., Chen, B., Zhou, X., & Xu, Y. (2013). Field monitoring and numerical analysis of Tsing Ma Suspension Bridge temperature behavior. *Structural Control and Health Monitoring, 20*(4), 560-575.

Xia, Y., Ni, Y. Q., Zhang, P., Liao, W. Y., & Ko, J. M. (2011). Stress development of a supertall structure during construction: field monitoring and numerical analysis. *Computer-Aided Civil and Infrastructure Engineering, 26*(7), 542-559.

Xie, J., Xu, L., & Chen, E. (2012). *Image denoising and inpainting with deep neural networks.* Paper presented at the Advances in Neural Information Processing Systems, Lake Tahoe, USA.

Xu, Y., Du, J., Dai, L.-R., & Lee, C.-H. (2015). A regression approach to speech enhancement based on deep neural networks. *IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), 23*(1), 7-19.

Xu, Y., Wei, S., Bao, Y., & Li, H. (2019). Automatic seismic damage identification of reinforced concrete columns from images by a region-based deep convolutional neural network. *Structural Control and Health Monitoring, 26*(3), e2313.

Zhang, F. L., Ni, Y. Q., Ni, Y. C., & Wang, Y. W. (2016). Operational modal analysis of Canton Tower by a fast frequency domain Bayesian method. *Smart Structures and Systems, 17*(2), 209-230.

Zhang, Y., Miyamori, Y., Mikami, S., & Saito, T. (2019). Vibration-based structural state identification by a 1-dimensional convolutional neural network. *Computer-Aided Civil and Infrastructure Engineering, 34*(9), 822-839.

Zhao, M., & Jia, X. (2017). 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*, 129-147.

Zhao, R., Yan, R., Chen, Z., Mao, K., Wang, P., & Gao, R. X. (2019). Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing, 115*, 213-237.

CHAPTER 5 DYNAMIC RESPONSE RECONSTRUCTION FOR STRUCTURAL HEALTH MONITORING USING DENSELY CONNECTED CONVOLUTIONAL NETWORKS

Abstract: This chapter proposes a novel dynamic response reconstruction approach for structural health monitoring by using Densely Connected Convolutional Networks (DenseNets). Skip connection and dense block techniques are carefully applied in the designed network architecture, which greatly facilitates the information flow, increases the training efficiency and accuracy of feature extraction and propagation with fewer parameters in the network. Sub-pixel shuffling and dropout techniques are used in the designed network and applied to reduce the computational demand and improve training efficiency. The network is trained in a supervised manner, where the input and output are the measurements of the available channels at response available locations and desired channels at response unavailable locations. The proposed DenseNets automatically extracts the high-level features of the input data and constructs the complicated nonlinear relationship between the responses of available and desired locations. Experimental studies are conducted using the measured acceleration responses from Guangzhou New Television Tower to investigate the effects of the locations of available responses, the numbers of available and unavailable channels, and measurement noise. The results demonstrate that the proposed approach can accurately reconstruct the responses in both time and frequency domains with strong noise immunity. The reconstructed response is further used for modal identification to demonstrate the usability and accuracy of the reconstructed responses. The applicability of the proposed approach for structural health monitoring is further proved by the highly consistent modal parameters identified from the reconstructed and true responses.

This chapter was accepted by *Structural Health Monitoring* with the full bibliographic citation as follows: Fan G, Li J, Hao H. (In press). Dynamic response reconstruction for structural health monitoring using densely connected convolutional networks. *Structural Health Monitoring*.

5.1 Introduction

Ensuring the safety and reliability of civil engineering structures is an important agenda in society. To real-time monitor and assess the structural condition and early report the anomaly in structural conditions and vibration behaviour, an increasing number of long-term structural health monitoring (SHM) systems have been installed on large-scale civil engineering structures, such as bridges and high rise slender structures (Cross et al., 2013; Nguyen et al., 2015; Ni et al., 2009; Xia et al., 2013). However, during the long-term monitoring, it is possible that measurement data from one or more sensors could be lost due to the technical issues such as the malfunction of cable connections, power supply interruption, signal transmission disturbance, sensor malfunction, or regular maintenances such as equipment check and sensor replacement. In practice, the locations of installed sensors on the monitored structures are selected carefully, and the number of sensors is much smaller compared to the total number of degrees of freedom (DOFs) of structures because of the limitations on the budget and available channels of data acquisition systems (Law et al., 2011), and the inaccessibility of some locations for measurement during operations (Kammer, 1997). Credible damage detection and condition assessment techniques assume that all the sensors are functioning in perfect condition (Ma et al., 2018). The incomplete acquired measurements that miss signals of one or more channels will greatly affect the performance of structural condition monitoring since some important local information of structures could be lost. For example, under the extreme events, SHM data could be lost due to the malfunction of systems and the lack of power supply of some sensors and data acquisition systems. However, the measurement data under those events are important for assessing structural conditions and evaluating the damage severities. As a result, structural dynamic response reconstruction of the lost data for some channels becomes an important research topic in the field of SHM.

In the past decades, several types of methods regarding the reconstruction of dynamic responses have been proposed utilizing the finite element model (FEM) based techniques. Transmissibility concept (Ribeiro et al., 2000) based structural response reconstruction methods were studied in both the frequency domain (Law et al., 2011) and wavelet domain

(Li & Law, 2011). The responses at the desired locations are reconstructed using the responses at other measured locations by using the transmissibility matrix. The limitation of these methods is that the information on the locations of the input excitations is necessary, and an accurate FEM is required. Deriving the transformation matrix utilizing the FEM of the structure to obtain the response at unavailable through the available response is another typical method (Kammer, 1997). This research is further combined with empirical mode decomposition (EMD), which decomposes the measurable time domain data into several Intrinsic Mode Function (IMFs), and reconstructs the responses at unavailable DOFs using the independent transfer equation of each IMF (He et al., 2012; Wan et al., 2014). The selection of IMFs is manually conducted, leading to the performance of the methods highly dependent on empirical experiences. An alternative structural response reconstruction method was developed based on optimal multi-type sensors placement and Kalman filter with unknown excitations (Xu et al., 2016; Zhang & Xu, 2016). Minimum-variance unbiased estimates of the generalized state of the structure and the external excitations were computed based on the measurements of the carefully selected locations. The generalized state of the structure and the external excitations were then used to reconstruct the response at the desired DOFs that are not installed with sensors.

These FEM based methods demand an accurate FEM of the target structure to calculate the transmissibility matrix, identify the mode shapes or determine the optimal sensor locations. However, developing an accurate FEM for large and complex civil engineering structures with a large number of elements and DOFs is time consuming and challenging. Furthermore, system and geometry properties, and boundary conditions of the structures are changing with the variations of the environmental and operational conditions, which leads to the establishment of an accurate FEM representing the in-service structures much more difficult. On the other hand, with the rapid development and application of data analysis techniques, realizing the response reconstruction by the data-driven methods are attracting significant attention. The majority of these data based methods focus on reconstructing the strain data (Huang et al., 2011; Zhang & Luo, 2017) or the strain distribution (Chen et al., 2018) using the inter-channel correlation. Compared to the strain data, acceleration response is more

sensitive to the structural vibration and accumulated damage in structures. However, the nonlinear inter-channel relationships are more complicated. Studies on the reconstruction of vibration acceleration data are mainly conducted to reconstruct the randomly missing data or short-term continuous lost data in a specific channel. For example, Bao et al. (2013) recovered the randomly lost acceleration data by leveraging the compressive sensing and further improved the developed method with group spare optimization (Bao et al., 2018) and machine learning techniques (Bao et al., 2020). Yang and Nagarajaiah (2016) also addressed the random data missing issue with sparse representation for inter-channel data reconstruction of low rank structures. Wan and Ni (2019) proposed a Bayesian multi-task learning method with multi-dimensional Gaussian process to predict the short-term continuous lost data using the measurements from the same sensor ahead of the forecasting. However, when a long-term successive acceleration data loss occurs or acceleration responses at critical but inaccessible locations are desirable, these methods may not be capable of accurately reconstructing the required responses.

With the advance of big data analytics in computer science and the tremendous growth of computing power, deep learning has gained much attention in broad research areas. Deep learning is capable of automatically identifying the pattern and mining the hidden highly abstract features from a large volume of data. SHM systems continuously measure the structural vibration responses, which can be used as training data and serve as a platform for using deep learning techniques to extract the abstract features and train the complex nonlinear relationships between input and output for dynamic response reconstruction. Developing and applying the deep learning techniques for SHM has gained significant research attention recently. The existing development and application of deep learning techniques in SHM mainly focus on the visual inspection and damage detection of structures by using the crack images (Cha et al., 2018; Gao & Mosalam, 2018; Xu et al., 2019), vibration responses (Abdeljaber et al., 2017; Lin et al., 2017; Zhang et al., 2019) and the pre-extracted vibration characteristics (Pathirage et al., 2018; Pathirage et al., 2019) as the input to the networks. A comprehensive review on deep learning-based SHM techniques is provided in Ref.(Ye et al., 2019). Through training the networks with a massive amount of datasets, the high-level

representations of the input data are automatically extracted by the deep learning models and then used to construct the nonlinear relationships with the output, which can be the structural crack locations and widths, damage location and severity. More applications in computer vision have been reported, i.e. the pixel-to-pixel recovery of images and audio super resolution (Dong et al., 2015; Kuleshov et al., 2017) and image inpainting (Mao et al., 2016; Pathak et al., 2016; Yeh et al., 2017). Considering that the concept of structural response reconstruction is similar to one-dimensional image recovery, deep learning technique has high potential to be leveraged to conduct the response reconstruction for SHM. Fan et al. (2019) developed the vibration lost data recovery approach by using convolutional neural networks (CNN) for recovering the lost sensor data, by using the measurement data from the same sensor set under various loading conditions. However, the reconstruction of vibration responses from instrumented sensor locations to other locations with sensor data lost or without sensors, by using the artificial intelligence techniques, has rarely been studied.

This chapter proposes a novel structural response reconstruction approach based on Densely Connected Convolutional Networks (DenseNets), for inter-channel vibration response reconstruction. Skip connection, dense connection, sub-pixel shuffling and dropout techniques are described and applied in the designed DenseNets. The effectiveness and accuracy of the proposed approach for the reconstruction of the vibration acceleration responses are investigated with the in-field testing data on Guangzhou New Television Tower (GNTT). The effects of the locations of available sensors and malfunction sensors, the number of available sensors and measurement noise are studied. The response reconstruction accuracy is evaluated in both time and frequency domains. Modal identification by using the reconstructed responses is also conducted to validate the accuracy of reconstructed responses qualitatively and quantitatively.

5.2 Methodology

In this study, the designed deep learning model, namely DenseNets, for structural dynamic response reconstruction is a one-dimensional fully CNN with densely connected layers. The DenseNets learns the inter-sensor relationships by a supervised manner using the

training datasets generated from measured vibration data before the sensors at desired locations become unavailable (faulty or inaccessible). The trained model can then be used to reconstruct the acceleration responses at desired locations leveraging the measurements from other available sensors. The idea of the architecture of this designed network is inspired from U-net (Ronneberger et al., 2015) (also known as convolutional encoder-decoder (Mao et al., 2016)), which consists of the compression path to capture the high-level features of the input and the reconstruction path to gradually expand the features to reconstruct the output. During the compression path, the feature maps are shrunken to facilitate the network to learn robust features of the input data while sifting the noise with uncertain patterns. Deep network with such architecture is proved to be efficient for conducting the pixel-to-pixel reconstruction tasks (Ronneberger et al., 2015). In addition, skip connection, dense connection, sub-pixel shuffling and dropout techniques are carefully applied to improve the performance of the developed DenseNets. Compared with the typical U-net, this newly designed network uses dense connection and skip connection instead of the successively connected convolutional layers to boost the information flow among the networks. The use of dense connection and skip connection strongly increases the training efficiency and accuracy of feature extraction and propagation, substantially reduce the number of parameters and alleviate the vanishing-gradient problem. In this case, the demanded volume of training data is significantly reduced, which makes the preparation of training data simpler. Fewer training data also bring efficient training compared with traditional deep neural networks. Moreover, optimized information flow also fastens the convergence speed which reduces the computational demand and further reduces the training time. Sub-pixel shuffling is a recently attractive method to replace the deconvolution which realizes the feature upscaling in a computationally efficient manner. Dropout technique is applied in the convolutional layers to mitigate the overfitting. The details and benefits of these techniques are elaborated in the following sections.

5.2.1 Configurations of the proposed DenseNets

The elaborated architecture of the proposed DenseNets is shown in Figure 5-1 and the detailed configurations of the proposed network are summarized in Table 5-1. Figure 5-1

shows an example of the networks, by using some functional SHM channel responses to reconstruct the response of a specific channel that is faulty or inaccessible.

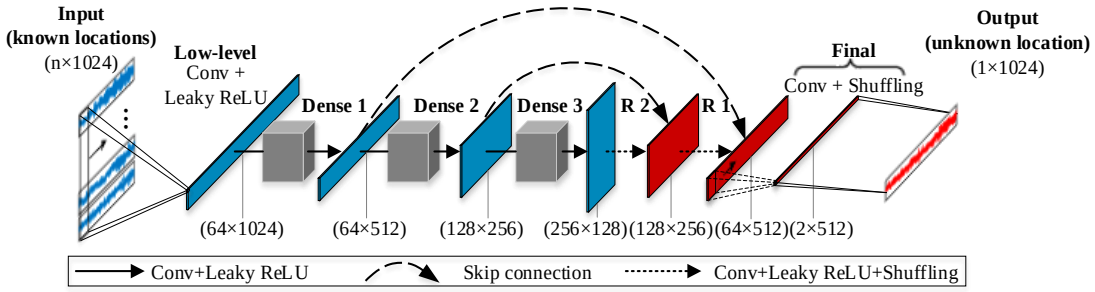


Figure 5-1. The architecture of the proposed DenseNets.

Table 5-1. The detailed configurations of the proposed DenseNets.

Layer	Kernel number	Kernel size	Stride	Padding	Input shape	Output shape	Leaky ReLU	sub-pixel
Input	-	-	-	-	(n, 1024)	(n, 1024)	-	-
Low level	64	64	1	Same	(1, 1024)	(128, 1024)	Y	N
Dense 1	64	64	-	Same	(64, 1024)	(64, 512)	Y	N
Dense 2	128	32	-	Same	(64, 512)	(128, 256)	Y	N
Dense 3	256	16	-	Same	(128, 256)	(256, 128)	Y	N
R2	256	16	1	Same	(256, 128)	(128, 256)	Y	Y
R1	128	32	1	Same	(256, 256)	(64, 512)	Y	Y
Final	2	64	1	Same	(128, 512)	(1, 1024)	N	Y
Output	-	-	-	-	(1, 1024)	(1, 1024)	-	-

The input and output of the DenseNets are the acceleration responses from *n* available locations and an unavailable location. The kernel number and size of each layer are given in Table 5-1. The first low-level feature extraction layer (Low-level as shown in Figure 5-1) aims to enrich the shallow features from the original input data with limited available channels by applying the convolution operation. 'Same' padding in Table 5-1 means zero-padding equally at the beginning and end of the input to allow the layers generating output with the same length as input when the stride is 1 and with a half-length of the input when the stride is 2. From the shallow features, the higher representative features of the input are gradually extracted using three successive dense blocks (Dense 1 to 3 as shown in Figure 5-1). After each dense block, the length of features is halved to facilitate the network for refining the higher-level

representative features and eliminating the noise with no certain features. Meanwhile, the number of feature maps is doubled to store more information. The output of the last dense block (Dense 3) has the highest-level abstraction of input and the largest number of feature maps as denoted in Figure 5-1. The reconstruction layers (R1 and R2 as shown in Figure 5-1) follow the last dense block to expand the feature maps by using the sub-pixel shuffling. The sub-pixel shuffling used here doubles the length and halves the number of feature maps. After two reconstruction layers, the output feature maps have half of the length of the input data. The final convolutional layer (Final as shown in Figure 5-1) finally reconstructs the structural response data at desired unavailable locations by convolutional and sub-pixel shuffling operations without the nonlinear transfer of the output (Shi et al., 2016). It should be noted that the output features at the symmetric levels have the same length, and the output features of the dense block are transported to the symmetric location by the skip connection. The shuttled features are concatenated with the output features of the reconstruction layers as the input of the next layer. It should be noted that the input and output sizes in Figure 5-1 and Table 5-1 are obtained, assuming that the length of data points in input and output sensor responses is 1024 time steps. Since the input data are halved three times by the dense blocks, the input length should be a multiple of 8 (2^3) to guarantee the length of the smallest feature maps is an integer. It is recommended to tailor the input response longer than the kernel size of the low-level feature extraction layer to fully utilize its capacity.

The proposed DenseNets for dynamic response reconstruction is based on supervised learning. For performing the structural response reconstruction at unavailable locations, the training datasets are generated using the measurement data with the functional sensors placed on these available and unavailable locations. The datasets consist of paired input data from available sensor locations and output data from unavailable locations. When using the training datasets with a massive number of samples, inputting all the samples to the networks at the same time will occupy a huge amount of memory and cause a large computational demand. Alternatively, the samples are loaded in batches to tune the parameters, i.e. weights and biases of the proposed network. The network parameters are tuned using the back-propagation algorithm. During the training process, the input data flow through the DenseNets to

reconstruct the responses, and the target of the training process is to minimize the difference between the predicted responses and true measured responses as output labels. The way to compute the loss between the predictions and labels is using the mean of the normalized L2-norm error of each sample in a batch as

$$Loss = \frac{1}{N}\sum_{k=1}^{N}\frac{\left\|y_k - C_\phi(x_k)\right\|_2^2}{\|y_k\|_2^2} \tag{5.1}$$

where *N* is the number of samples in each batch, y_k and x_k denote the input and output of the *k*th sample in this mini batch. The proposed network is simply depicted as a nonlinear model $C_\phi(x)$ 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. To fully train the networks, the training datasets will be generally loaded by more than 100 epochs considering the data complexity.

5.2.2 One-dimensional convolution

Compared to the traditional deep artificial neural network (ANN) that fully connects the neurons between adjacent layers, CNN is a class of deep neural networks consisting of one or more convolutional layers. CNN is a powerful deep learning architecture that can extract the hierarchical pattern in data by stacked convolutional layers with fewer parameters. A convolutional layer contains several convolutional kernels with each consisted of a weight matrix and a bias number to extract the feature of input. Through sliding the kernel across the input data with a specific stride, convolution results are obtained as the dot product of the weight matrix with the scanned segment of the input data plus the bias number. After the kernel reaching the end of input, the results of all positions are concatenated together with the sequence to form the output feature map of this kernel. The output feature maps from all kernels will be subsequently used as the input for the successive layer. Generally, as the complex relationship between the input and output cannot be simply represented by linear functions, the output of convolution will be nonlinearly transferred by the activation function which can be expressed as

$$O = H(I, W, b) = \Phi(W * I + b) \tag{5.2}$$

where *O* and *I* are the output and input of the convolutional layer, respectively. *W*, and *b* denote the learnable weights and bias of the kernel, *H* and Φ represent the convolutional layer and the activation function, respectively. The activation function used in this study is an advanced activation function named Leaky Rectified Linear Unit (*Leaky ReLU*) (Maas et al., 2013), which is developed from Rectified Linear Unit (*ReLU)* (Nair & Hinton, 2010). Compared with the *ReLU* which defines the output as zero when the input is less or equal to zero, and *'sigmoid'* or *'tanh'* activation functions which squash the input into a very narrow output range ([0 -1] for *'sigmoid'*, and [-1 1] for *'tanh'*), *Leaky ReLU* has a nonlinear gradient over its whole domain as follows

$$Leaky\ ReLU(x) = \begin{cases} x, & if\ x > 0 \\ 0.01x, & \text{otherwise} \end{cases} \tag{5.3}$$

Leaky ReLU has a broader range of output which mitigates the gradient vanishing. Meanwhile, it allows a non-zero small gradient to overcome the defect of *ReLU*, that is, the neuron is potentially never be activated once a large gradient passes it (Maas et al., 2013). Acceleration data are considered as time-dependent one-dimensional signals. Therefore, the one-dimensional convolution is selected among all the layers in this application for structural response reconstruction. One-dimensional convolution can be considered as a special case of the two-dimensional convolution, where the convolutional kernel slides along only one axis of the input. To adopt this method for the case with multichannel signals, the input acceleration data from multiple channels are parallelly arranged. Meanwhile, the kernel height is shaped as the same as the number of channels.

5.2.3 Skip connection

It should be mentioned that when tuning the networks with deep architectures of a layer-to-layer sequential connection by using the gradient and back-propagation, the gradient vanishing in the deeper layers is a crucial problem which significantly influences the extraction of higher-level hidden features (Huang et al., 2017). Skip connection directly shuttles the low-

level features extracted in the bottom layers to the top layers and allows the features to be back-propagated to the bottom layers directly, which enhances the information flow and alleviates the gradient vanishing issue (Isola et al., 2017). Moreover, it can also pass the low-level details that are lost during the convolution to the bottom layers (Mao et al., 2016). Networks with skip connections are also known as Residual Neural Networks (ResNet). The skip connection technique in the proposed DenseNets is used in both shuttling the features from the bottom layer to the top layer and embedding in the dense block. The involvement of skip connection significantly strengthens the feature propagation and consequently boosts the convergence efficiency.

5.2.4 Dense block

The dense block is used in the proposed network to enhance the extraction of the hidden high-level features from the input. A dense block consists of multiple densely connected convolutional layers (Huang et al., 2017; Zhang et al., 2018). The dense connectivity strategy is implemented by skip connection between any layer and all subsequent layers as shown schematically in Figure 5-2. Dense connection improves the information flow from the shallow layer to the deep layers in this block and reduces the number of parameters by sharing the features of shallow layers with deep layers. The dense connection extracts the higher-level features much more efficiently with fewer parameters than the traditional layer-to-layer sequential connected convolutional layers which use a large number of kernels and parameters. Compared with traditional deep learning models using the layer-to-layer connection, networks equipped with densely connected layers is stronger in feature extraction. Networks with dense connections are more efficient in training, and the extracted features are complete and comprehensive. Traditional layer-to-layer connected CNN with feature loss is hard to achieve the same training accuracy as the proposed networks equipped with the skip and dense connections. Especially for time domain response reconstruction that belongs to pixel-to-pixel tasks which contain a vast feature extraction and propagation work, which is almost impossible by using the traditional CNN. Layer-to-layer connection for response reconstruction requires much more kernels resulting in a significant increase in the number of parameters to be trained.

In this study, to achieve the similar results as the developed DenseNet, the required number of parameters for networks with the same depth but composed of layer-to-layer connection only is around 2 times and the training time is more than 4 times because it needs more training iterations for learning response features.

The detailed configuration of these three successive convolutional layers is shown in Table 5-2. In each of the dense blocks, the numbers of kernel and kernel size keep the same. The first three convolutional layers (Conv 1-3) implement the convolution by sliding the kernels with a stride of 1 to keep the output have the same length. The last convolution layer (Conv 4) conducts the convolution utilizing the extracted feature maps from all the previous layers. Convolutional kernels slide with a stride of two, which halves the dimension of feature maps. The output high-level features are used as the input of the next dense block.

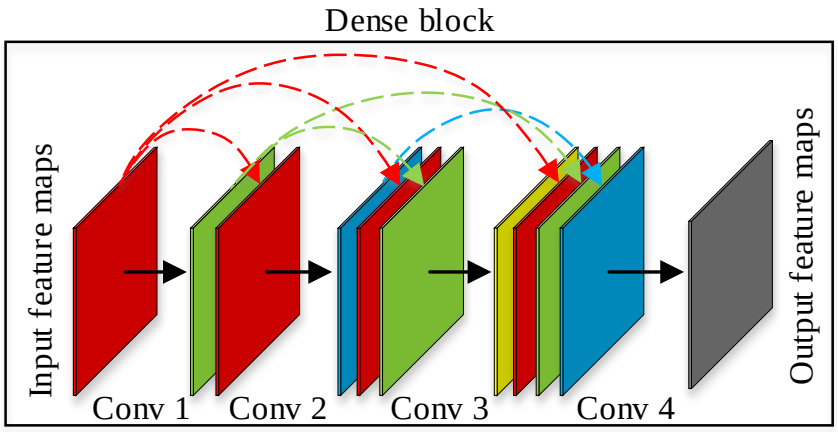


Figure 5-2. A schematic 4-layers dense block.

Table 5-2. The detailed configurations of the dense blocks.

	Layer	Kernel number	Kernel size	Stride	Padding	Input shape	Output shape	Leaky ReLU
Dense Block 1	Conv 1	64	64	1	Same	(128, 1024)	(64, 1024)	Y
	Conv 2	64	64	1	Same	(192, 1024)	(64, 1024)	Y
	Conv 3	64	64	1	Same	(256, 1024)	(64, 1024)	Y
	Conv 4	64	64	2	Same	(320, 1024)	(64, 512)	Y
Dense Block 2	Conv 1	128	32	1	Same	(64, 512)	(128, 512)	Y
	Conv 2	128	32	1	Same	(192, 512)	(128, 512)	Y
	Conv 3	128	32	1	Same	(320, 512)	(128, 512)	Y
	Conv 4	128	32	2	Same	(448, 512)	(128, 256)	Y
Dense Block 3	Conv 1	256	16	1	Same	(128, 256)	(256, 256)	Y
	Conv 2	256	16	1	Same	(384, 256)	(256, 256)	Y
	Conv 3	256	16	1	Same	(640, 256)	(256, 256)	Y
	Conv 4	256	16	2	Same	(896, 256)	(256, 128)	Y

5.2.5 Sub-pixel shuffling

To gradually expand the feature maps in the reconstruction layers of the designed networks, the sub-pixel shuffling operation is embedded in the reconstruction layers following the nonlinear activation operation to upscale the feature maps by a factor of 2. Figure 5-3 demonstrates an example of sub-pixel shuffling with four feature maps consisting of 6 features each. It divides the feature maps to 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 the sub-pixel convolution layer (Shi et al., 2016). This is an efficient operation for upscaling features, which costs less computational demand than the general deconvolution. It has also been attested to have strong workability that introduces less artefact in the output (Odena et al., 2016).

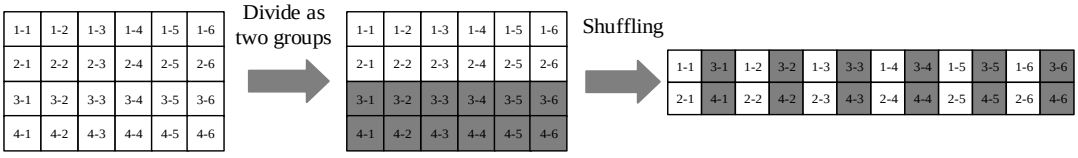


Figure 5-3. The schematic procedure of one-dimensional sub-pixel shuffling with 4 feature maps and 6 elements in each map.

5.2.6 Dropout

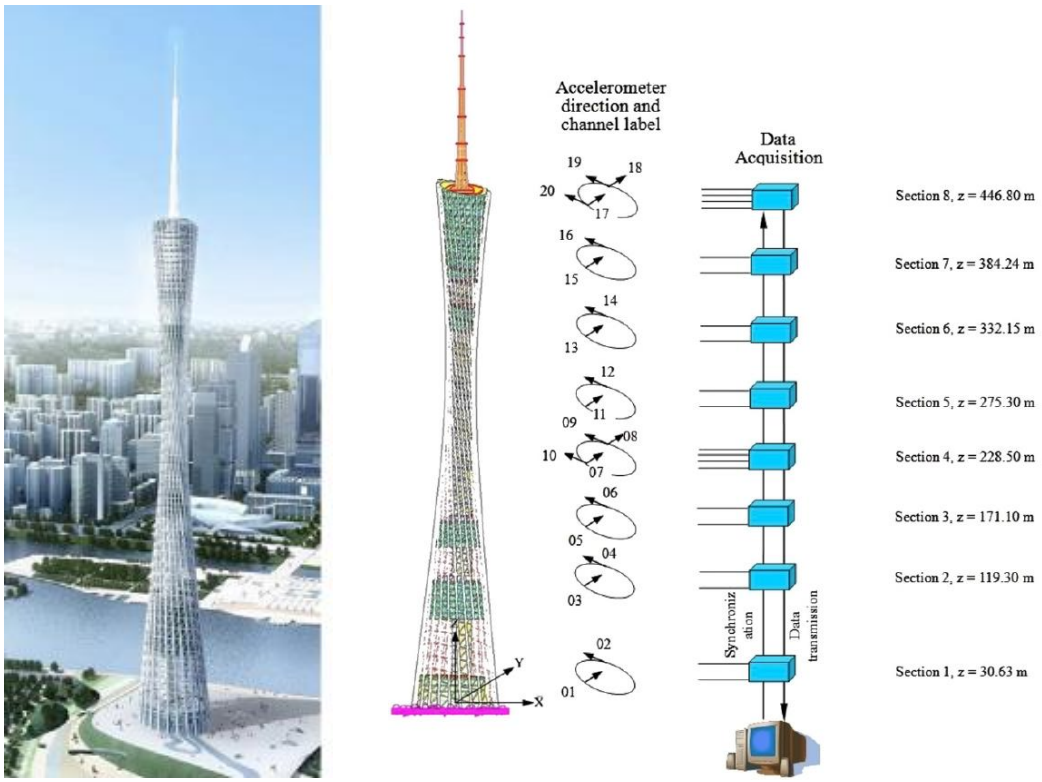
Overfitting is a common issue in the training of networks with deep architectures. It happens when the network model overfits the training datasets and loses the generalization. Instead of spending vast time and computational resources on training several networks with random initialization and averaging their outputs, dropout technique is an emerging tool for addressing the overfitting issue. Dropout technique de-activates a certain proportion of neurons and disconnects these neurons with the adjacent input and output layers to break up the co-adapted sets of neurons (Srivastava et al., 2014). The selection of dropping out neurons is random when training with a different batch of samples. Those remained neurons are trained more robustly, and the generalization capacity of the networks is enhanced. Each neuron has an independent probability p to be dropped, where p=0.5 is suggested by the inventors (Srivastava et al., 2014) and used in the training of the proposed DenseNets.

5.3 Experimental studies

5.3.1 GNTT and its SHM system

GNTT is a 610 meters supertall slender structure. It has a tube-in-tube structural design including a reinforced concrete inner tube and a steel outer tube, which consists of 24 concrete-filled-tube columns and 46 steel ring beams and bracings. 24 columns are uniformly spaced in an oval shape and twisted in the vertical direction as shown in Figure 5-4 (a). The size of the oval changes from 50m×80m at the ground level to 20.65 m×27.5 m at the height of 280 m and further to 41m×55 m at the top of the tube with a height of 454 m. 37 floors connecting the inner tube and the outer tube by beams, are used for offices, entertainment, catering and emission of the television signals (Ni et al., 2009). A sophisticatedly designed SHM system including more than 600 sensors is installed on the GNTT for monitoring the structural vibration behaviour and ambient environmental conditions in both the construction and in-service stages. The in-service acceleration measurement data are selected as the training and testing of the proposed method. The location placement of the accelerometers is shown in Figure 5-4(b). Totally 20 uni-axial accelerometers are installed at the assigned heights in both

the long-axis and short-axis directions. The sampling rate of the acceleration measurement is 50 Hz. The raw measurement is processed by a high-pass filter with a 0.05Hz cut-off frequency to eliminate the shift at the zero frequency. The previous studies (Chen et al., 2011; Xia et al., 2011) have validated the effectiveness of the SHM system and extended GNTT as a benchmark platform for high-rise structures. The vibration measurement data will be used to validate the effectiveness and accuracy of the proposed DenseNets for dynamic structural response reconstruction.



(a) (b)

Figure 5-4. (a) GNTT, (b) The deployment of accelerometers and sensor numbers.

In this study, 24 hours measurement data along the short-axis direction including channels No. 1, 3, 5, 7, 11, 13, 15 and 17 are processed as the datasets for evaluating the accuracy of using the proposed approach for response reconstruction. Two case studies are conducted. The first case study is conducted to investigate the effectiveness of reconstructing the response of one unavailable channel by using the responses of different numbers of available channels. The second case study is conducted to evaluate the performance of the proposed method, when data from more than one channels are lost and are reconstructed,

especially when the most correlated channel to the desired channel is not available. The noise immunity of the proposed method will also be investigated in the second case study by using measurement data contaminated with noise effect.

5.3.2 Data pre-processing

The previous studies (Chen et al., 2011; Zhang et al., 2016) have conducted the modal identification of the GNTT and reported that the first 15 vibration modes are within 2 Hz. Therefore, the frequency of interest of this study is selected as 0 to 2 Hz. The original measurements sampled at 50 Hz are pre-processed by using the low-pass filtering with a cut-off frequency of 5Hz and then downsampled to 10 Hz. The filtering processing effectively eliminates the redundant information contained in the signals and consequently reduces the number of features to be learned by the network. Meanwhile, downsampling the acceleration response by a factor of five significantly reduces the dataset size and greatly boost the training efficiency. The pre-processed acceleration data are then scaled to a range between -1 and 1 without affecting its statistical distribution by

$$A'_{i,j} = \frac{A_{i,j}}{max\ (|A|)} \tag{5.4}$$

where A is the acceleration response vector after downsampling, which consists of the measurements from totally k involved channels with n sampling points, $A_{i,j}$ denotes the ith sampling point of the jth channel, where i ranged in [1, 2…n] and j ranged in [2, 3…k], $A'_{i,j}$ is the normalized data after scaling.

The normalized data A' are then used to generate the training, validation and testing datasets. For training the proposed DenseNets by supervised training, the input and output are the measurements from available sensor locations and the desired unavailable locations. To facilitate the network to learn robust features from comprehensive datasets, the normalized acceleration responses in 24 hours are divided as 96 segmentations, where each segmentation contains 15 minutes measurement data. The last 10% of segmentations are used as the testing data to simulate the situation that the measurement data of the desired channels are unavailable. This will be used to test the accuracy of the network for dynamic response reconstruction. The

remaining segmentations are randomly grouped as the training and validation datasets containing 80% and 10% of the total number of segmentations, respectively, to construct and validate the relationship between input and output. As mentioned in Section 5.2.1, utilizing all the training data to tune the parameters of deep networks at the same time is impractical. Due to the changing of load and vibration characteristics with the variation of ambient conditions, inputting the segmentations one by one will lead to the model unstable and hard to reach the optimal parameters. Alternatively, the segmentations are cut as small samples with 1024 data points each, by a sliding window which is 1024 data points long with a stride of 512. Those samples are shuffled and inputted to the DenseNets in batches for both training and validation. The batch size is chosen as 32 for the following studies. The acceleration data with measurements from available channels are directly inputted to the trained networks and the measurements of the desired channel are defined as the labelled output. For testing the network, the reconstructed responses are compared with the original true measurement data from the channels at unavailable locations to validate the accuracy of the proposed approach.

5.3.3 The effect of input channels

Selecting proper available channels as the input of the DenseNet is important for an effective response reconstruction. In the first case study, channel 17 is assumed to be faulty after a period of service for making reliable measurements. The effect of using different locations and numbers of available channels as the input of the DenseNets to conduct the dynamic response reconstruction is investigated. Firstly, the selection of the most correlated input channels for reconstructing channel 17 is studied. For response reconstruction by deep learning models, a channel containing substantial overlapped natural frequencies and vibration modes with the desired channel will have a higher correlation. Traditional calculation of correlation function of two response series in time domain may not be directly useful since the similarity in learned features, i.e. vibration modal information between the recorded and desired channels is more important for response reconstruction. On the other hand, manual selection of correlated channels may introduce large errors due to noise and measurement errors. Therefore, the correlations between channels are evaluated by one-to-one channel

reconstruction. Recorded channels used for response reconstruction producing lower reconstruction errors means the responses from these channels can provide more useful information for feature learning and dynamic response reconstruction of desired channels. Seven groups of training, validation and testing datasets containing the available measurement data from one of the channels 1, 3, 5, 7, 11, 13 and 15, respectively, and the output from channel 17 are generated following the procedure introduced in Section 5.3.2. One to one sensor response reconstruction is conducted. Using these datasets, seven networks are trained and tested. The relative reconstruction error is quantified by computing and normalizing the L2-norm of the discrepancy between the reconstructed acceleration response A_R and original true acceleration response A_T as:

$$\varepsilon = \frac{\|A_R - A_T\|_2^2}{\|A_T\|_2^2} = \frac{\sqrt{\sum_{i=1}^{n}(A_R^i - A_T^i)^2}}{\sqrt{\sum_{i=1}^{n} {A_T^i}^2}} \tag{5.5}$$

The relative error is the summation of differences at all the points considered. When evaluating the error of response reconstruction in the frequency domain, A_R and A_T are the acceleration responses in the frequency domain obtained by Fast Fourier Transfer (FFT).

The coding of the DenseNets is compiled with Keras under TensorFlow backend. The computer used for training is built with a GTX2080TI GPU, an i7-6700K CPU and 16GB memory. Each of the above-mentioned networks is trained by the training dataset for 150 epochs. The training of each network spends around 10 minutes and the testing can be implemented in almost real time. The training and validation loss of the DenseNets using input data from channel 15 after each epoch of training is illustrated in Figure 5-5. It can be observed that both the training and validation losses decrease sharply in the first several epochs because of the involved dense connection and skip connection techniques which improve the convergence speed effectively. The training loss continues to decrease with increased training epoch. While the validation loss stops decreasing around the 110 epochs and finally exceeds the training loss when 125 training epochs are finished, indicating that the network starts overfitting the training data. After the training process, a minor discrepancy between the

training and validation errors is found, which means that the developed DenseNets learns robust features from the training data and also fits the validation data well. Selecting 150 epochs for training is practical, which ensures the proposed networks to be fully trained without overfitting.

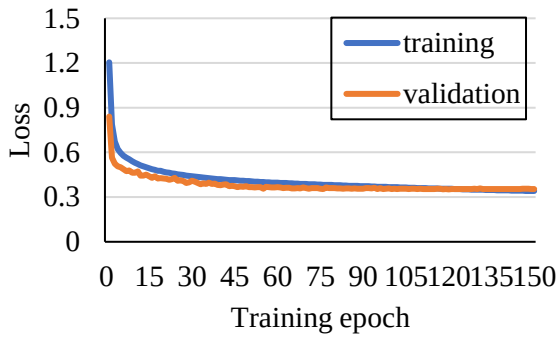


Figure 5-5. The convergence curve of DenseNets with input measurement from channel 15.

The reconstruction errors for each network with different training dataset in both time and frequency domains are shown in Figure 5-6. The errors are computed using all the testing data in 150mins to represent more reliable results. The reconstruction errors in the time and frequency domains show similar trends. Selecting channel 1 as the input brings the largest response reconstruction error, while using channel 15 as input leads to a much smaller reconstruction error. Figures 5-7(a) and 8(a) demonstrate the comparison between the original true and reconstructed responses of the unavailable location at channel 17 in the time domain within an arbitrarily selected 60s, by using responses from channels 1 and 15, respectively. Figures 5-7(b) and 8(b) show the true and reconstructed responses of channel 17 in the frequency domain. The spectrum of the response at unavailable locations, i.e., the response used for constructing the unavailable response, in the frequency domain is also shown. It should be noted that the accelerometer of channel 1 is installed at a height of 30.63 meters of the GNTT. Compared to the total height of the structure which is 610 meters, the vibration response of the structure under ambient excitation at the location of channel 1 is very small and most of the vibration modes are weakly excited as shown in Figure 5-7(b). This leads to a low signal to noise ratio (SNR) and results in the reconstruction difficult based on the low quality information. This can also explain that except reconstruction by using channel 7, the reconstruction errors decrease with the rising height of the available channels with the

increasing quality of the input information. Evidenced by Figures 5-7 and 5-8, it is noted that for an effective response reconstruction, the responses from the channels at available locations should contain the information of the significant vibration modes of the responses at desired locations. This idea is similar to optimal sensor placement for modal identification, which ensures that the maximum modal information is obtained and the mode shapes are highly correlated. Locations that are weakly participated in a certain vibration mode will not contribute too much valuable information for response reconstruction at the corresponding modal frequency. As a result, the measurement data from these locations may not provide essential information for reconstructing the responses related to this particular mode at other locations. In this study, for the case with 15 modes, it is impractical that all the involved modes of the desired locations are well excited at any one single node, which is far away from the desired locations. Therefore, multiple sensors are necessary for accurate dynamic response reconstruction, especially when considering a large number of vibration modes. By conducting the modal analysis of the measurement data, acceleration response from channel 7 is found to have the minimum number of overlapped modes with channel 17, which causes a relatively large reconstruction error, as observed in Figure 5-6. Consequently, the reconstruction performance is strongly correlated with the response information and the overlapped modes between the available channels and desired channels.

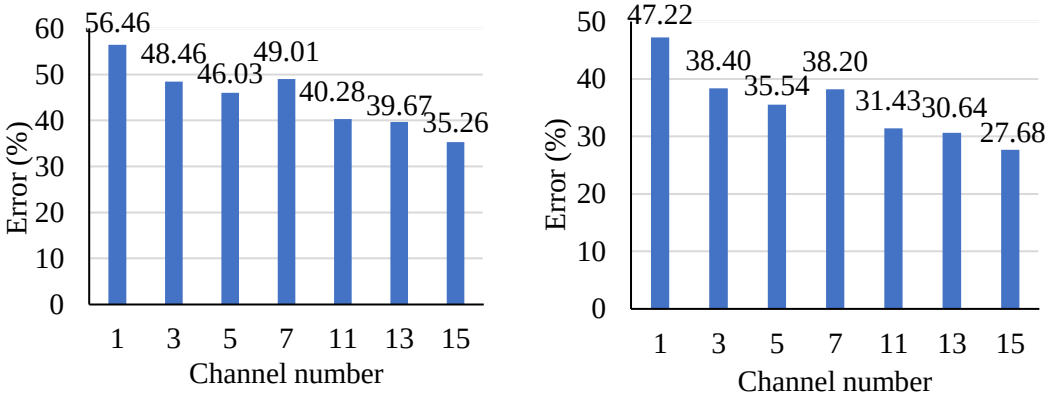
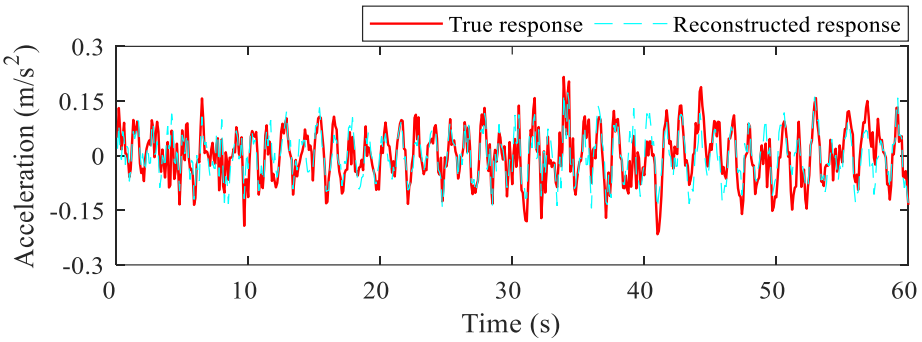
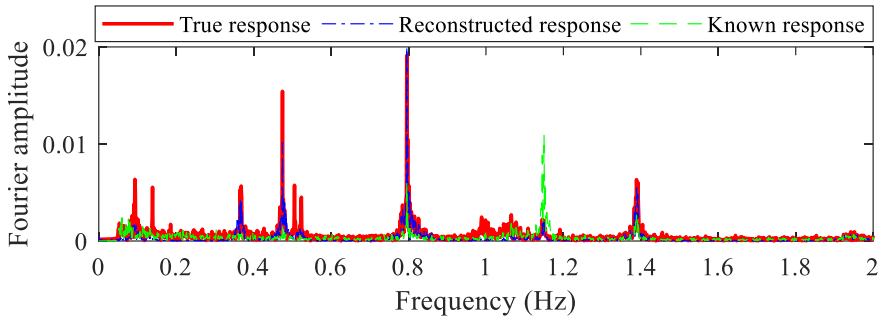


Figure 5-6. One-to-one response reconstruction error by using input measurement from different channels. (a) Time domain. (b) Frequency domain.

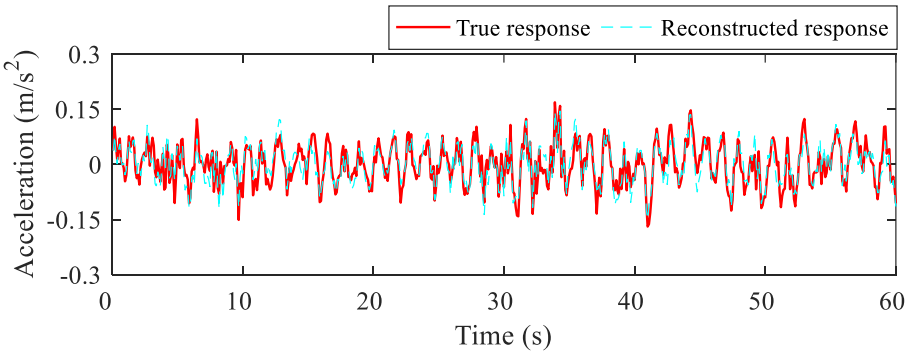


(a)

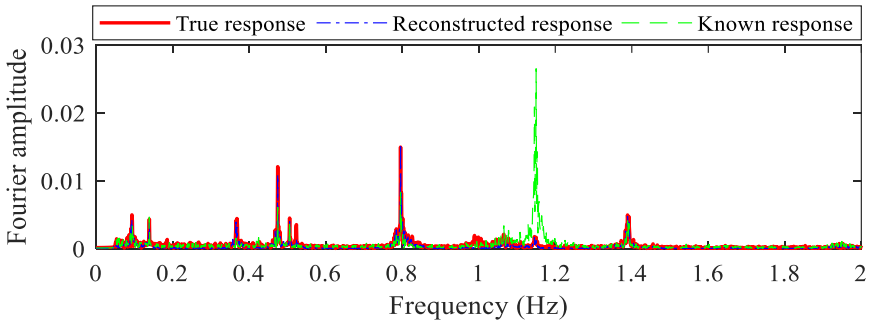


(b)

Figure 5-7. Response reconstruction of channel 17 within an arbitrarily selected 60s by using channel 1: (a) Time domain; (b) Frequency domain.



(a)



(b)

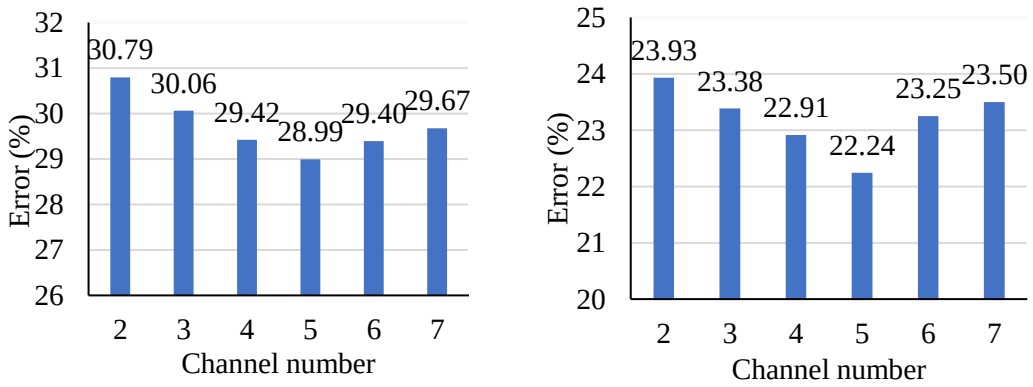
Figure 5-8. Response reconstruction of channel 17 within an arbitrarily selected 60s by using channel 15: (a) Time domain and (b) Frequency domain.

On the contrary, using input from multiple channels simplifies this issue. In this section, structural response reconstruction with multiple channels input is also conducted to reconstruct the responses at the desired locations. To test the reconstruction performance of using every single and multi-channels to reconstruct the response of channel 17, the most correlated two, three, and until all the available channels are selected and processed as the input to train different networks. The number of input channels and the corresponding channels is summarized in Table 5-3. The training time of multiple channels is also around 10 minutes for each network, since the architecture of the network and the number of training and validation samples keep the same. Because the used dense connection and skip connection make the feature extraction and propagation very efficient, the small size of training datasets is required for constructing the nonlinear relationship. Owning to the high convergence efficiency, the training of the networks is only about 10 minutes, which is much shorter compared with training a traditional deep network that usually takes hours or even days. The reconstruction errors of using each trained network in time and frequency domains are shown in Figure 5-9. As shown, the reconstruction error decreases with the increase of available channels, and achieves the minimum when data from five channels are used in the reconstruction of the response of channel 17. After that, the error increases with the number of channels used for reconstruction. It should be noted that when data from channel 1 and channel 7 are used, the reconstruction error is higher. This is because channel 1 is farthest from channel 17 and data from channel 7 has the minimum overlapped modes with channel 17. As discussed above, the measurement from these two channels may not provide valid information for reconstruction but slightly affect the effective training and fitting of the network model. From the deep learning aspect, guaranteeing the quality of training datasets is of great importance for promoting the network to learn effective and abstract features. For example, when conducting the image classification using CNN, mixing low-resolution or damaged photos in the training datasets will bring serious negative impact on the classification accuracy. Figures 5-10(a) and (b) show the reconstructed responses by using the best 5 channels as input to demonstrate the effectiveness of the proposed method. The responses are accurately reconstructed in the time domain with minor discrepancies, and the vibration modes are matched well in the frequency

domain. The measurement data from multiple channels can provide more comprehensive information of responses and vibration modes, however, the use of low quality data may even affect the accuracy. Nonetheless, the response reconstruction accuracy of using measurement data from multiple channels as input is much better than using a single channel, even when some interferential information from the farthest channels are involved in the input data. When selecting the optimal locations and numbers of sensors from available measurement channels for structural response reconstruction, one-to-one channel response reconstruction is conducted and sorted from the most to least correlated. The number of available channels is increased based on the order of the above obtained sequence from results on one-to-one channel reconstruction to serve as input for multiple channel response reconstruction to choose the combination with the best performance. It should be noted that this is not the most effective way to select correct channels but is effective for most cases with a limited number of installed sensors.

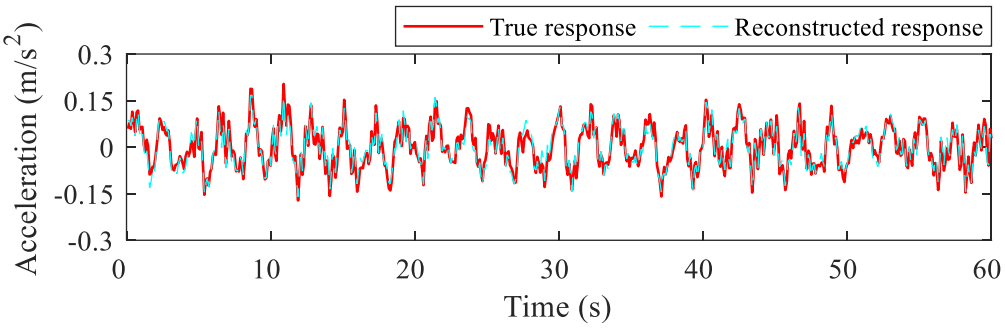
Table 5-3. The number of input channels and the corresponding channel numbers.

No. of input channels	Channel numbers
2	13, 15
3	11, 13, 15
4	5, 11, 13, 15
5	3, 5, 11, 13, 15
6	3, 5, 7, 11, 13, 15
7	1, 3, 5, 7, 11, 13, 15

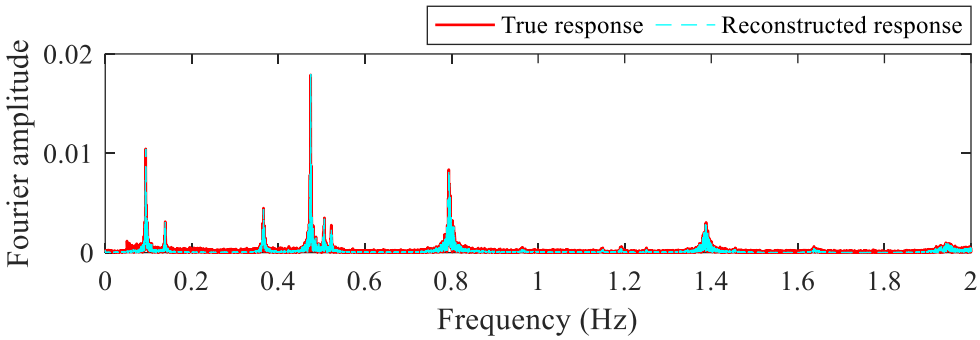


(a) (b)

Figure 5-9. The reconstruction errors of using an increasing number of input channels. (a) Time domain. (b) Frequency domain.



(a)



(b)

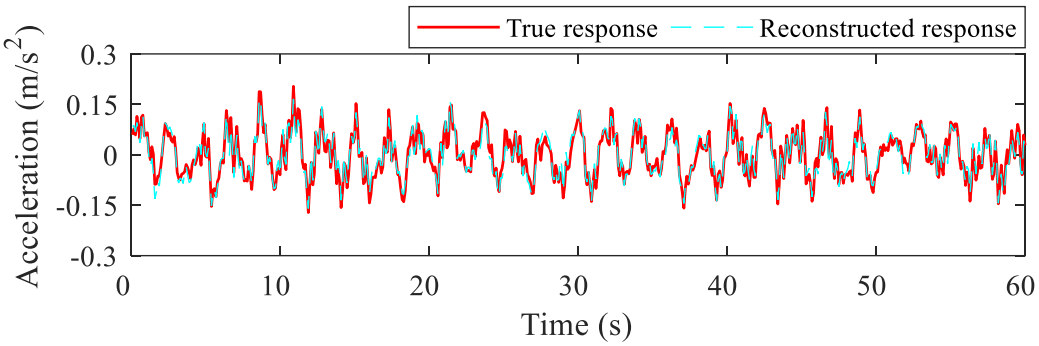
Figure 5-10. Reconstruction of response at channel 17 by using 5 channels. Comparison in (a) time domain and (b) frequency domain.

5.3.4 Most correlated channel is unavailable

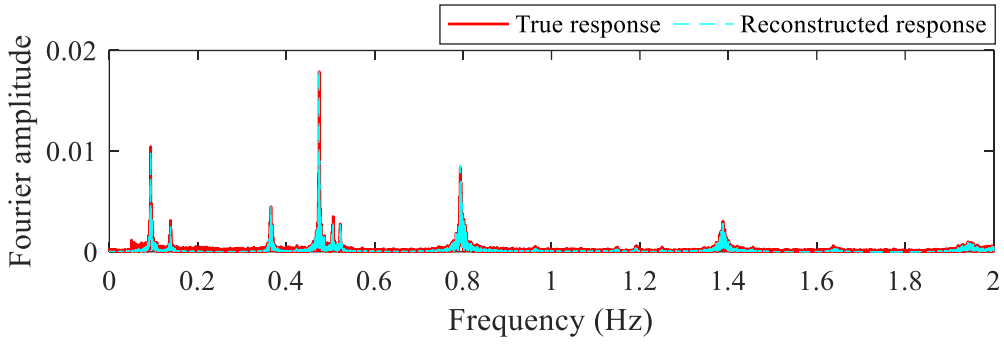
As discussed above, the correlation between responses of available and unavailable locations is very important for structural response reconstruction. One rigorous case is when the measurement from the most correlated channel to the desired channels is not available. From Figure 5-6, the one-to-one reconstruction results indicate that channel 15 is the most correlated channel with channel 17 and provides the best one-to-one reconstruction accuracy. From the results as shown in Figure 5-9, using the measurement data from channels 1 and 7 will induce negative effects on the response reconstruction of channel 17. Therefore, the measurement data of the rest channels from No. 3, 5, 11 and 13 are used as the input. The same training and testing procedures are followed as described in Section 5.3.2. To clearly show the discrepancy, a 60 seconds segment of the true and reconstructed responses in the time domain are plotted in Figure 5-11(a). The true and reconstructed responses are then transferred to the frequency domain and illustrated in Figure 5-11(b). The reconstructed response in the time domain shows a good agreement with the true response in both amplitudes and waveforms.

Confirmed by the response in the frequency domain, the trained DenseNets accurately reconstructs all the involved vibration modes of channel 17 in both resonant frequencies and amplitudes. That indicates the developed DenseNets successfully constructs the complex nonlinear relationships between the input and output through the training process. Robust high-level features are extracted from the testing data, which is then used to accurately reconstruct the response at the desired location. By observing Figure 5-11(b), it can be found that the reconstructed response has a higher SNR compared to the raw measurement. This proves the advantage of using the bottleneck structure where noise components are eliminated when extracting the high level of features by shrinking the feature maps. The reconstruction error is 30.11% and 23.51% for the reconstructed response in the time domain and frequency domain, respectively. The error is mainly contributed by the noise components submerged in the true response in the region of the frequency band without natural frequencies with significant energy. In contrast with the reconstruction using the measurement from the best 4 channels as listed in Table 5-3, the error is slightly higher, but the desired response can still be reconstructed with a high degree of accuracy and reliability.

The case study to reconstruct the responses of two channels by one network is also conducted. A minor adjustment of the architecture of DenseNets is implemented, which only increases the feature maps of the final layer from 2 to 4. As a result, the output of DenseNets becomes the measurement of two channels. The training data is generated where the input data contain measured responses from channels 3, 5, 11 and 13 and the output data consist of reconstructed responses from channels 15 and 17. Reconstructed responses for these two channels are evaluated together, where the errors for response reconstruction in time and frequency domains are 31.98% and 25.37%, respectively. With the increasing complexity of the relationships between input and output, the reconstruction error is slightly higher than reconstructing a single channel. Compared with preparing two datasets and training two networks individually, the developed network can also conduct the response reconstruction of multiple channels simultaneously which promotes the reconstruction efficiency but slightly sacrifices the accuracy.



(a)



(b)

Figure 5-11. Reconstruction response of channel 17 when the most correlated channel is unavailable. Comparison in (a) time domain and (b) frequency domain.

5.3.5 The noise effect

Noise is a critical issue that influences the effectiveness and accuracy of condition assessment and damage detection. The noise immunity of the proposed method is investigated by adding a certain level of white Gaussian noise to the input measurement and evaluating the accuracy of the reconstructed response. The white noise is simulated as a random vector N with the same length and sampling rate as the input response. The random vector follows a distribution with zero mean and unit standard deviation. The added noise is proportional to the root mean square (RMS) of the response A, the noisy acceleration response A_{noisy} is generated as

$$A_{noisy} = A + N * RMS(A) * l \tag{5.6}$$

where l represents the noise level. In this study, the responses from channels 3, 5, 11, 13 and 15 are selected as the input, and the response of channel 17 is selected as the output to be reconstructed. The network is trained with the raw measurement and tested by the noisy

response with 20% RMS noise added to the input of the testing data. An example of a segmental original and noisy acceleration responses is shown in Figure 5-12. The noisy response is then inputted to the trained DenseNets, and the reconstructed response of channel 17 in time and frequency domains are demonstrated in Figures 5-13(a) and (b), respectively. The reconstructed response in the time domain shows a good agreement with the true one. From the frequency domain, the modal frequencies are accurately reconstructed excepting some minor differences in the amplitudes of a few modes. The errors of the reconstructed responses in time and frequency domains are quantified as 31.34% and 25.62%, respectively, compared with the true responses. Referring to the results in Section 5.3.3, the errors are 2.35% and 3.38% higher than using measured response without noise. The increases of reconstruction errors are moderate in term of the severity of the injected noise, demonstrating the superiority of using the proposed approach in immunizing the noise effect. The proposed DenseNets eliminates most of the noise in the higher level features which can be attributed to the specially designed network architecture.

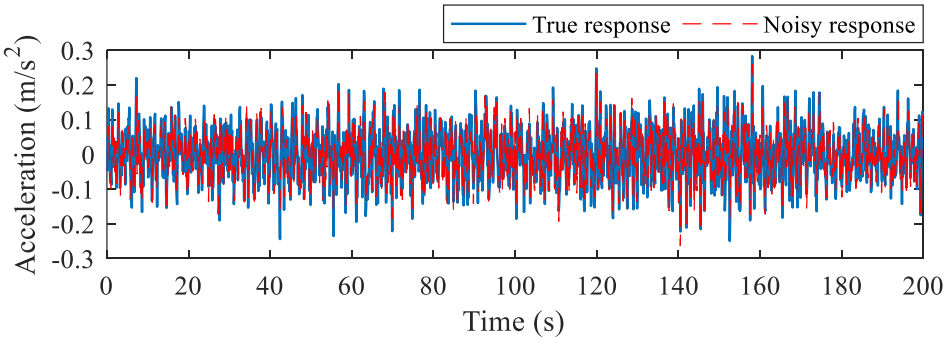


Figure 5-12. Comparison between the true response and noisy response with 20% noise.

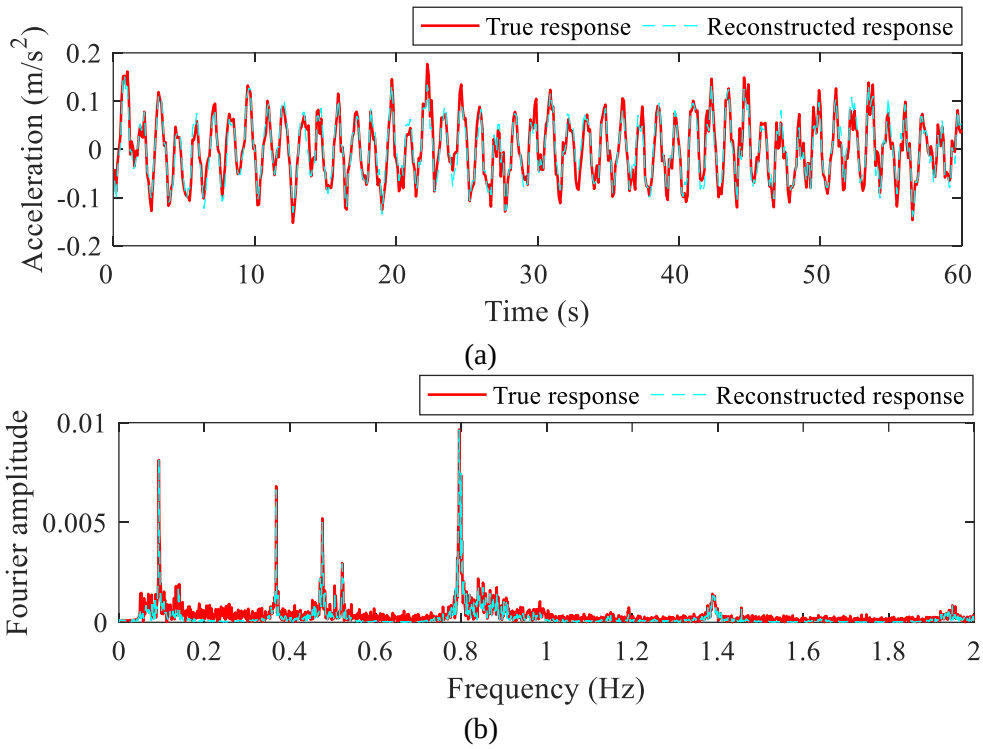


Figure 5-13. Comparison of true and reconstructed responses in the (a) time and (b) frequency domain with noisy input.

5.4 Modal identification using the reconstructed response

Modal parameters including natural frequency, damping ratio and mode shape which reflect the dynamic vibration characteristics of the structure, are widely used to develop the damage index for condition assessment and damage detection. Accurate identification of modal parameters is of significant importance for effective condition monitoring, since the identification error may mask the variation of modal parameters induced by structure condition changes. Thus, modal analysis by using the true and reconstructed responses is further conducted to evaluate the applicability of the proposed method. The responses of channel 17 are reconstructed using responses of available channels No. 3, 5, 11, 13 and 15. A classic operational modal identification method named Frequency Domain Decomposition (FDD) (Brincker et al., 2001) is selected for modal parameters identification using original true and reconstructed responses from channel 17. FDD method decomposes the spectral density function matrix of the responses by singular value decomposition (SVD) and estimates the vibration modes by peak picking. The following sections will evaluate the feasibility of using

reconstructed responses for modal identification, by comparing with the FDD results of true responses from qualitative and quantitative aspects.

5.4.1 Qualitative analysis of the reconstructed response

For providing the reference modal information, the modal parameters of GNTT are firstly identified by using Stochastic Subspace Identification (SSI) and FDD methods with measurements taken from all the channels as shown in Figure 5-14. 15 modes within 2 Hz are accurately identified, which agree very well with the results reported in previous studies (Chen et al., 2011; Zhang et al., 2016). By conducting the modal parameters identification with true and reconstructed responses from channel 17, FDD outputs of these two responses are shown in Figures 5-15(a) and (b), respectively. It can be observed from Figure 5-15 that modes No. 1, 3, 5, 8, 11, 13 and 15 are significantly excited, and are accurately identified with the true and reconstruction responses of channel 17. Those modes are selected by peak picking and marked by pink points with the corresponding system mode orders. Picking the peak values from the plot of singular values is based on the engineering judgement. Consequently, the effectiveness of modal identification using FDD is dependent on the quality of responses where signals with low SNR can result in a confusing output. Observing the FDD results of using the reconstructed responses as shown in Figure 5-15(b), all the excited modes are effectively and clearly reconstructed without adding any extra spurious modes. It should be highlighted that the power of the response components outside the bandwidths of natural frequencies are much lower than that of natural frequencies, which reveals a high signal quality. Comparing Figure 5-15(b) with Figure 5-15(a), the SNR of the reconstructed response is better than the true response. The results demonstrate that the proposed DenseNets can effectively reconstruct the responses which are not measured or are lost, and eliminate the noise components simultaneously owning to its strong capacity and specifically designed architecture. In addition, it is worth mentioning that vibration modes of GNTT are very closely spaced, which leads to the response reconstruction task more difficult for the traditional methods especially for defining the thresholds or parameters to split closely spaced modes. In contrast, the proposed method can automatically extract the hidden features from the input

data and learn the complicated relationships between the input and output accurately. By analyzing Figure 5-15(b) where the vibration modes are well reconstructed and the noise is eliminated, those learned features are very likely to be the vibrational characteristics. The results are consistent with the finding reported in an existing study (Lin et al., 2017) where the extracted high level features from vibration responses represent the natural frequencies and mode shapes of the structure.

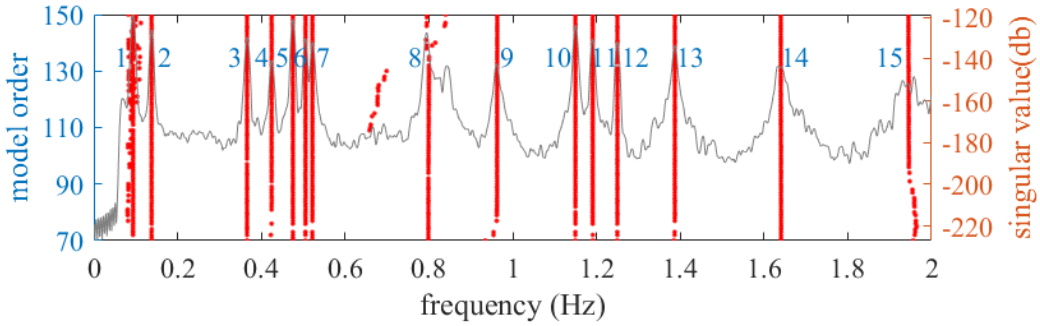


Figure 5-14. SSI and FDD outputs of using measurements from all channels.

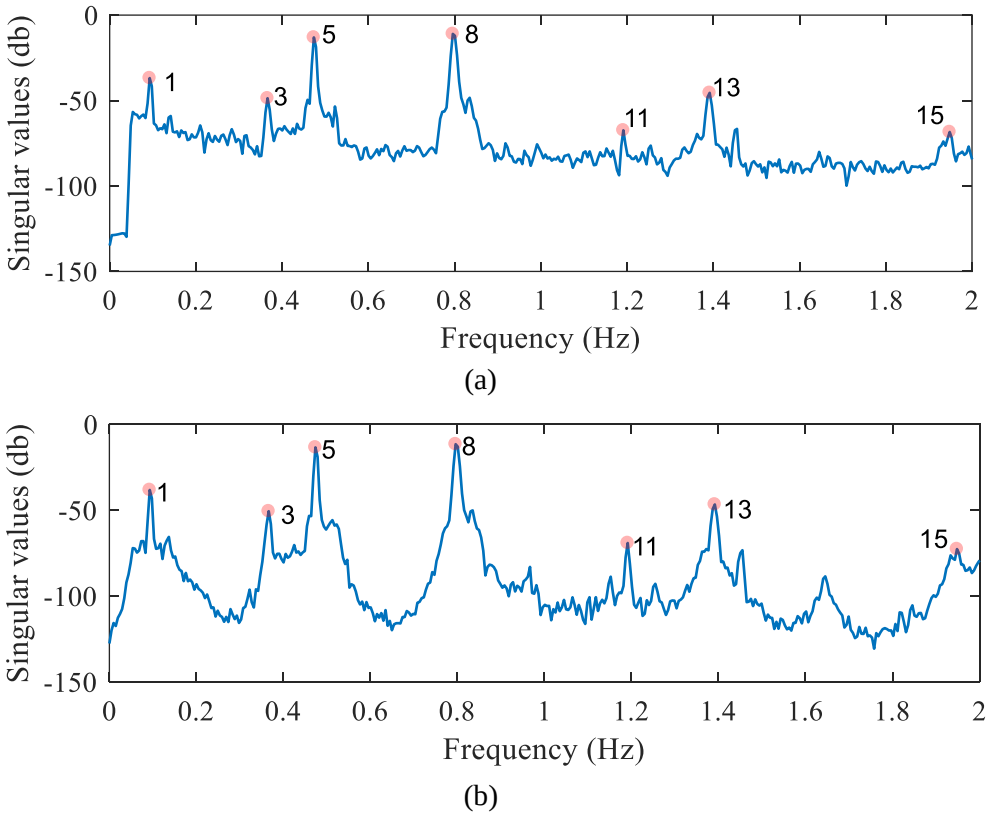


Figure 5-15. Modal identification results of using: (a) True response and (b) Reconstructed response.

5.4.2 Quantitative analysis of the modal parameters

The effectiveness of the proposed method is further verified by comparing the identified modal parameters from the reconstructed responses and those from the true responses. The natural frequencies and damping ratios are identified using true and reconstructed responses, respectively. As mode shapes cannot be identified using only one channel, evaluation of the mode shapes is realized by considering the correlation between channels of true and reconstructed responses. The correlation accuracy evaluation of mode shapes is performed by using the Co-ordinate Modal Assurance Criterion (COMAC), which is superior in assessing the measurement from a single DOF (Allemang, 2003). The COMAC of a single DOF is computed as

$$COMAC = \frac{(\sum_{k=1}^{m} |\phi_k^T \phi_k^R|)^2}{\sum_{k=1}^{m} {\phi_k^T}^2 \sum_{k=1}^{m} {\phi_k^R}^2} \tag{5.7}$$

where m is the total number of modes, ϕ_k^T and ϕ_k^R are real values representing the mode shape values of the kth mode estimated from the true and reconstructed responses. The identified modal parameters of those 7 modes observed in Figure 5-15 and the corresponding errors are summarized in Table 5-4. The errors in the identified natural frequencies and mode shapes are computed as the relative errors, while the errors in the damping ratios are represented as the absolute errors since the oscillation in identified damping results is usually bigger. The identified natural frequencies from the reconstructed response are completely consistent with the ones from the true response, which confirms the observation in Figure 5-15. Results of damping ratios show relatively large errors. Considering that the damping ratios are very sensitive to the measurement noise and modal identification method, the accuracy of identified damping ratios with the reconstructed responses is satisfactory. The mode shapes of these excited modes are also very close between the ones from true and reconstructed responses. The maximum relative error is 2.03% for mode 15, which is a high order mode. As shown in Figure 5-14, this mode is relatively unstable comparing to other modes even using all the sensors, indicating this mode is weakly excited in this duration. In addition, when evaluating the mode shapes correlation of this DOF using the widely accepted criterion, the

COMAC value reaches 99.99% which shows a very high correlation. All the modal parameters of these observed modes are accurately identified from the reconstructed response, demonstrating the applicability of the proposed method to large civil structures.

Table 5-4. Comparison of the identified modal parameters from true and reconstructed responses.

Mode	f_{true} (Hz)	$f_{reconstructed}$ (Hz)	f error (%)	ξ_{true} (%)	$\xi_{reconstructed}$ (%)	ξ error (%)	ϕ error (%)
1	0.093	0.093	0.00	2.13	2.05	0.08	0.28
3	0.366	0.366	0.00	0.84	0.94	0.10	0.81
5	0.474	0.474	0.00	0.26	0.24	0.02	1.42
8	0.796	0.796	0.00	0.36	0.37	0.01	1.37
11	1.191	1.191	0.00	0.13	0.17	0.04	1.88
13	1.392	1.392	0.00	0.82	0.79	0.03	0.90
15	1.948	1.948	0.00	0.43	0.38	0.05	2.03
						COMAC: 99.99%	

Note: f: frequency; ξ: damping ratio; ϕ: mode shape.

5.5 Conclusions

This chapter proposes a novel approach based on DenseNets for reconstructing the acceleration responses at locations without recorded data, by using recorded responses at other locations. The detailed design of the DenseNets for facilitating the reconstruction of vibration response is elaborated. The developed DenseNets can alleviate the vanishing-gradient issue, strengthen feature extraction and propagation, and substantially reduce the number of parameters, which significantly increases the training efficiency and accuracy. The effectiveness and robustness of the proposed method are demonstrated by experimental studies with real measurement data taken from GNTT. The response reconstruction is accurately performed in both time and frequency domains. Meanwhile, the identified modal parameters from the true and reconstructed responses show a good agreement. The results demonstrate that the proposed method can extract the high-level features from the input data and develop the nonlinear relationship between these features and the output responses to be reconstructed. The proposed method is also proved to have a strong noise immunity, which is practical for civil engineering applications. Further study can be conducted to investigate this topic to select

the optimal sensor numbers and locations for response reconstruction by using deep learning techniques.

5.6 References

Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M., & Inman, D. J. (2017). Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration, 388*, 154-170.

Allemang, R. J. (2003). The modal assurance criterion–twenty years of use and abuse. *Sound and vibration, 37*(8), 14-23.

Bao, Y., Li, H., Sun, X., Yu, Y., & Ou, J. (2013). Compressive sampling–based data loss recovery for wireless sensor networks used in civil structural health monitoring. *Structural Health Monitoring, 12*(1), 78-95.

Bao, Y., Shi, Z., Wang, X., & Li, H. (2018). Compressive sensing of wireless sensors based on group sparse optimization for structural health monitoring. *Structural Health Monitoring, 17*(4), 823-836.

Bao, Y., Tang, Z., & Li, H. (2020). Compressive-sensing data reconstruction for structural health monitoring: a machine-learning approach. *Structural Health Monitoring, 19*(1), 293-304.

Brincker, R., Zhang, L., & Andersen, P. (2001). Modal identification of output-only systems using frequency domain decomposition. *Smart Materials and Structures, 10*(3), 441.

Cha, Y. J., Choi, W., Suh, G., Mahmoudkhani, S., & Büyüköztürk, O. (2018). Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. *Computer-Aided Civil and Infrastructure Engineering, 33*(9), 731-747.

Chen, W., Lu, Z., Lin, W., Chen, S., Ni, Y., Xia, Y., & Liao, W. (2011). Theoretical and experimental modal analysis of the Guangzhou New TV Tower. *Engineering Structures, 33*(12), 3628-3646.

Chen, Z., Bao, Y., Li, H., & Spencer Jr, B. F. (2018). A novel distribution regression approach for data loss compensation in structural health monitoring. *Structural Health Monitoring, 17*(6), 1473-1490.

Cross, E., Koo, K., Brownjohn, J., & Worden, K. (2013). Long-term monitoring and data analysis of the Tamar Bridge. *Mechanical Systems and Signal Processing, 35*(1-2), 16-34.

Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence, 38*(2), 295-307.

Fan, G., Li, J., & Hao, H. (2019). Lost data recovery for structural health monitoring based on convolutional neural networks. *Structural Control and Health Monitoring, 26*(10), e2433.

Gao, Y., & Mosalam, K. M. (2018). Deep transfer learning for image-based structural damage recognition. *Computer-Aided Civil and Infrastructure Engineering, 33*(9), 748-768.

He, J., Guan, X., & Liu, Y. (2012). Structural response reconstruction based on empirical mode decomposition in time domain. *Mechanical Systems and Signal Processing, 28*, 348-366.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). *Densely connected convolutional networks.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, USA.

Huang, Y.-W., Wu, D.-G., & Li, J. (2011). Structural healthy monitoring data recovery based on extreme learning machine. *Computer Engineering, 16*, 241-243.

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). *Image-to-image translation with conditional adversarial networks.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, USA.

Kammer, D. C. (1997). Estimation of structural response using remote sensor locations. *Journal of Guidance, Control, and Dynamics, 20*(3), 501-508.

Kuleshov, V., Enam, S. Z., & Ermon, S. (2017). Audio super resolution using neural networks. *https://arxiv.org/abs/1708.00853*.

Law, S., Li, J., & Ding, Y. (2011). Structural response reconstruction with transmissibility concept in frequency domain. *Mechanical Systems and Signal Processing, 25*(3), 952-968.

Li, J., & Law, S. (2011). Substructural response reconstruction in wavelet domain. *Journal of Applied Mechanics, 78*(4), 041010.

Lin, Y. Z., Nie, Z. H., & Ma, H. W. (2017). Structural damage detection with automatic feature-extraction through deep learning. *Computer-Aided Civil and Infrastructure Engineering, 32*(12), 1025-1046.

Ma, S., Li, J., Hao, H., & Jiang, S. (2018). Structural response recovery based on improved multi-scale principal component analysis considering sensor performance degradation. *Advances in Structural Engineering, 21*(2), 241-255.

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). *Rectifier nonlinearities improve neural network acoustic models.* Paper presented at the International conference on machine learning, Atlanta, USA.

Mao, X., Shen, C., & Yang, Y. B. (2016). *Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections.* Paper presented at the Advances in neural information processing systems, Barcelona, Spain.

Nair, V., & Hinton, G. E. (2010). *Rectified linear units improve restricted boltzmann machines.* Paper presented at the Proceedings of the 27th international conference on machine learning, Haifa, Israel.

Nguyen, T., Chan, T. H. T., Thambiratnam, D. P., & King, L. (2015). Development of a cost-effective and flexible vibration DAQ system for long-term continuous structural health monitoring. *Mechanical Systems and Signal Processing, 64-65*, 313-324.

Ni, Y. Q., Xia, Y., Liao, W. Y., & Ko, J. M. (2009). Technology innovation in developing the structural health monitoring system for Guangzhou New TV Tower. *Structural Control and Health Monitoring, 16*(1), 73-98.

Odena, A., Dumoulin, V., & Olah, C. (2016). Deconvolution and checkerboard artifacts. *Distill, 1*(10), e3.

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). *Context encoders: feature learning by inpainting.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, USA.

Pathirage, C. S. N., Li, J., Li, L., Hao, H., Liu, W., & Ni, P. (2018). Structural damage identification based on autoencoder neural networks and deep learning. *Engineering Structures, 172*, 13-28.

Pathirage, C. S. N., Li, J., Li, L., Hao, H., Liu, W., & Wang, R. (2019). Development and application of a deep learning–based sparse autoencoder framework for structural damage identification. *Structural Health Monitoring, 18*(1), 103-122.

Ribeiro, A., Silva, J., & Maia, N. (2000). On the generalisation of the transmissibility concept. *Mechanical Systems and Signal Processing, 14*(1), 29-35.

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-net: convolutional networks for biomedical image segmentation.* Paper presented at the International conference on medical image computing and computer-assisted intervention, Munich, Germany.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., & Wang, Z. (2016). *Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, USA.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*(1), 1929-1958.

Wan, H.-P., & Ni, Y.-Q. (2019). Bayesian multi-task learning methodology for reconstruction of structural health monitoring data. *Structural Health Monitoring, 18*(4), 1282-1309.

Wan, Z., Li, S., Huang, Q., & Wang, T. (2014). Structural response reconstruction based on the modal superposition method in the presence of closely spaced modes. *Mechanical Systems and Signal Processing, 42*(1-2), 14-30.

Xia, Y., Chen, B., Zhou, X., & Xu, Y. (2013). Field monitoring and numerical analysis of Tsing Ma Suspension Bridge temperature behavior. *Structural Control and Health Monitoring, 20*(4), 560-575.

Xia, Y., Ni, Y. Q., Zhang, P., Liao, W. Y., & Ko, J. M. (2011). Stress development of a supertall structure during construction: field monitoring and numerical analysis. *Computer-Aided Civil and Infrastructure Engineering, 26*(7), 542-559.

Xu, Y., Wei, S., Bao, Y., & Li, H. (2019). Automatic seismic damage identification of reinforced concrete columns from images by a region-based deep convolutional neural network. *Structural Control and Health Monitoring, 26*(3), e2313.

Xu, Y., Zhang, X., Zhu, S., & Zhan, S. (2016). Multi-type sensor placement and response reconstruction for structural health monitoring of long-span suspension bridges. *Science bulletin, 61*(4), 313-329.

Yang, Y., & Nagarajaiah, S. (2016). Harnessing data structure for recovery of randomly missing structural vibration responses time history: Sparse representation versus low-rank structure. *Mechanical Systems and Signal Processing, 74*, 165-182.

Ye, X., Jin, T., & Yun, C. (2019). A review on deep learning-based structural health monitoring of civil infrastructures. *Smart Structures and Systems, 24*(5), 567-585.

Yeh, R. A., Chen, C., Lim, T.-Y., Schwing, A. G., Hasegawa-Johnson, M., & Do, M. N. (2017). *Semantic image inpainting with deep generative models.* Paper presented at the The IEEE conference on computer vision and pattern recognition, Honolulu, USA.

Zhang, C. D., & Xu, Y. (2016). Optimal multi-type sensor placement for response and excitation reconstruction. *Journal of Sound and Vibration, 360*, 112-128.

Zhang, F. L., Ni, Y. Q., Ni, Y. C., & Wang, Y. W. (2016). Operational modal analysis of Canton Tower by a fast frequency domain Bayesian method. *Smart Structures and Systems, 17*(2), 209-230.

Zhang, Y., Miyamori, Y., Mikami, S., & Saito, T. (2019). Vibration-based structural state identification by a 1-dimensional convolutional neural network. *Computer-Aided Civil and Infrastructure Engineering, 34*(9), 822-839.

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018). *Residual dense network for image super-resolution.* Paper presented at the Proceedings of the IEEEconference on computer vision and pattern recognition, Salt Lake City, USA.

Zhang, Z., & Luo, Y. (2017). Restoring method for missing data of spatial structural stress monitoring based on correlation. *Mechanical Systems and Signal Processing, 91*, 266-277.

CHAPTER 6 DYNAMIC RESPONSE RECONSTRUCTION UNDER EXTREME LOADING CONDITIONS USING SEGMENT BASED GENERATIVE ADVERSARIAL NETWORK

Abstract: Reconstruction of lost response under extreme loading conditions is important for structural health monitoring to evaluate the safety of structures. This chapter proposes a Segment based Conditional Generative Adversarial Network (SegGAN), which is a powerful deep learning model for solving pixel-to-pixel tasks, to conduct structural dynamic response reconstruction under extreme loading conditions. The proposed network consists of a bottlenecked generator and a segment based discriminator. Generator features skip and dense connections to improve feature extraction, and segment based discriminator uses conditioned input to facilitate generator to learn detailed and robust features. Numerical studies on a steel frame structure are conducted to evaluate the accuracy and noise immunity of using SegGAN for structural response reconstruction. Numerical study on a nonlinear building model is performed to validate the capability and accuracy of using the proposed approach in nonlinear response features extraction and response reconstruction. Experimental studies on a laboratory steel frame structure are conducted to investigate the effectiveness and robustness of the proposed SegGAN for response reconstruction with limited measurement data from noisy and operational conditions for training and validation, and a few sample data from earthquake loads for testing of the developed networks. The reconstruction results of using SegGAN are compared with those generated by a densely connected network (DenseNet) with the same configuration as the generator and a traditional convolutional neural network (CNN). Responses from available and unavailable sensors are selected as input and output of these three networks, respectively. SegGAN outperforms the other two networks and produces outstanding reconstruction results in both time and frequency domains.

6.1 Introduction

SHM has attracted significant research interest in recent decades. Long-term SHM systems measure structural vibration responses continuously, assist engineers for real-time assessing the structural condition and early warn the anomaly to prevent calamity. Owning to the significant benefits to society and public safety, a rapidly growing number of long-term SHM systems have been deployed in large-scale civil structures (Chen et al., 2011; Cross et al., 2013; Nguyen et al., 2015; Peeters & De Roeck, 2001; Xia et al., 2013; Zhang et al., 2012). However, restricted by construction budget, available channels of data acquisition systems (Law et al., 2011), and the inaccessibility of locations for measurement during operations (Kammer, 1997), SHM systems typically contain a small number of sensors compared to the total number of DOFs of structures. Locations of the limited number of sensors are therefore carefully selected with a few redundancies, leading the reliability of following analyses heavily dependent on measurements from all sensors. Guaranteeing SHM systems to work properly all-weather under all kinds of circumstances to steadily record the vibration data (Kullaa, 2010) is vital in this case. Unfortunately, it is possible that data from one or a few channels are lost or corrupted due to technical issues such as malfunction of cable connections, power supply interruption, signal transmission disturbance or sensor fault, especially when structures are under extreme loading conditions, e.g. earthquakes. Since the degradation or damage of structures is often caused by extreme loading, vibration responses measured from structures under such circumstances are more precious for determining structural conditions. Therefore, reconstruction of responses of structures at locations with data loss under extreme loading conditions is significant for SHM.

Structural response reconstruction methods can be categorized into two main streams, namely, FEM based methods and data driven methods. Among the FEM based methods, Li and Law (2011) proposed an approach based on transmissibility concept (Ribeiro et al., 2000) to reconstruct the vibration data at desired locations with no data available using available response data at other locations. Transmissibility function derived from the frequency response function (FRF) was utilized. Responses were transferred to the wavelet domain for

reconstruction in Li and Law's study and further extended to the frequency domain by Law et al. (2011). It should be noted that both methods require the locations of excitations. Using transformation matrix derived from mode shapes of structures for response reconstruction is an alternative FEM based method (Kammer, 1997). He et al. (2012) combined the transformation matrix with empirical mode decomposition (EMD) technique to reconstruct responses at locations without properly measured data. Time domain vibration measurements are decomposed by using EMD into several Intrinsic Mode Function (IMFs), then responses of unavailable DOFs are reconstructed by leveraging independent transfer equations of IMFs. Wan et al. (2014) developed a method for solving the difficulty of distinguishing closely spaced modes. However, the choice of IMFs is highly dependent on engineers' empirical experiences, resulting in that the reconstructed performance lacks stability. Another kind of method was developed based on optimal multi-type sensors placement and Kalman filter that conducts the excitation and response reconstruction with unknown excitation (Xu et al., 2016; Zhang & Xu, 2016). For those FEM based methods, an accurate FEM of the investigated structure is essential for deriving the transmissibility function, transformation function or selecting the optimal sensors locations. The performances of those methods are closely related to the accuracy of the built FEM, which could be a major limitation in the applications of these methods to the large and complex civil structures with many inevitable uncertainties in modelling. Besides the inevitable modelling error, varying operational and environmental conditions introduce the change in material properties and boundary conditions of structure. These indicate that establishing an accurate deterministic FEM that represents the in-service vibration behaviour of structures under varying environmental and operational conditions becomes time consuming and challenging.

On the other hand, data driven methods along with the advance in computational power and data analysis techniques that directly extract the hidden relationships of vibration responses between available and unavailable locations from measured data, have been vigorously explored recently. Many researches have been conducted to reconstruct the strain (Huang et al., 2011; Oh et al., 2020; Zhang & Luo, 2017) or strain distribution (Chen et al., 2018) of bridges or spatial structures. They employed intra-channel correlations of stresses,

owing to the plane assumption of stresses at different locations. Acceleration responses have been widely measured and applied in SHM since they are convenient to measure and sensitive to structural conditions. However, reconstruction of acceleration responses has not been widely studied due to the complicated intra-channel relationships. Bao et al. (2013) developed the compressive sampling based method to recover the random lost acceleration data. The performance is further improved when involving group sparse optimization (Bao et al., 2018) and machine learning (Bao et al., 2020) techniques. Wan and Ni (2019) proposed a Bayesian multi-task learning method with multi-dimensional Gaussian process to recover short-term continuous lost acceleration data. These methods used measurements from the same sensor for recovering lost data of this particular sensor, but will lose effectiveness when a long-term successive data loss occurs. Deep learning as an emerging branch of machine learning technique, has achieved an outstanding progress in multiple disciplines such as image processing, medical diagnosis, autopilot etc. Development and application of deep learning techniques in the field of SHM are recently a hot topic, and mainly focus on structural damage detection based on the crack images (Cha et al., 2018; Gao & Mosalam, 2018; Xu et al., 2019), vibration responses (Abdeljaber et al., 2017; Lin et al., 2017; Zhang et al., 2019; Zhang et al., 2019) or damage sensitive features (Pathirage et al., 2018; Pathirage et al., 2019). The applications have also been extended to signal processing, such as detecting the anomaly of measurements by a deep neural network (Bao et al., 2019). For dynamic response reconstruction, Zhang et al. (2019) proposed a long-short term memory network that used the measured excitation as input to predict structural seismic response. Dozens of records under earthquakes must be provided for training, which might not be practical for those structures with short service time or located in areas with few earthquakes. Fan et al. (2019) developed a ResNet to reconstruct the randomly lost acceleration responses. Reconstructing the long-term successive data loss needs to be further investigated. Sun et al. (2017) applied the traditional CNN for intra-channel acceleration data reconstruction. With the simple network structure, the performance is still good for an experimental simply-supported beam under ambient excitation. However, a simple network may not be applicable for real structures with a large number of DOFs, uncertainties and different kinds of loadings. Chapter 5 proposed a

DenseNet for reconstructing response of a real building. The performance is remarkable when the building is under ambient excitations.

Generative adversarial network (GAN) (Goodfellow et al., 2014) is one class of the state-of-the-art deep learning frameworks consisting of a generative model and a discriminative model, which are named as generator and discriminator, respectively. GAN is trained through an adversarial process. The generator learns the distribution of real data and generates samples to confuse the discriminator, which makes it believe that the generated samples are from the real data distribution. On the other hand, the discriminator learns the features of the generated and real data to distinguish the source of samples. With continuous adversarial training, the generator distribution keeps approaching the data distribution by using feedbacks (gradients) provided by discriminator, and the discriminator learns more robust features of generated and real data from more and more real generated samples. After training, generator can produce samples that are very close to real data. Discriminator works as a self-learned objective function of generator to tune its parameters. Owning to the framework of GAN, it has exhibited a powerful capability on the pixel-to-pixel tasks such as image super-resolution (Ledig et al., 2017), style translation (Zhu et al., 2017) and inpainting (Isola et al., 2017; Pathak et al., 2016; Yeh et al., 2017). The generated images display clear outlines (low-frequency structure of data) and even sharp texture details (high-frequency structure of data), which networks only utilizing a human defined loss function may miss. Dynamic response reconstruction can also be considered as a pixel-to-pixel task, where the dynamic response is a one-dimensional signal with hidden relationships between DOFs of a structure. Vibration characteristics such as natural frequencies and mode shapes can be the low-frequency data structure of vibration response since they have a small perturbation crossing the time history, while the amplitude of each individual time step or instantaneous frequency change can be considered as the high-frequency data structure. GAN captures both low-frequency and high-frequency data features, is therefore very promising to be implemented for response reconstruction of structures under extreme loading.

This chapter proposes a specially designed segment-based conditional GAN (SegGAN) leveraging the advantages of GAN to reconstruct the dynamic responses of structures under

ambient excitations or extreme loadings using the intra-channel correlations. When applying GAN for structural response reconstruction, generator is used to learn the mapping between available and desired responses. Discriminator learns to evaluate the accuracy of reconstructed responses, and provides feedback to generator to facilitate the learning of more detailed features. SegGAN features conditioned input, skip connection, dense connection, and dropout techniques to improve the reconstruction. The remainder of this chapter is organized as follows. Section 6.2 introduces the detailed configurations and describes the adversarial training process of the proposed SegGAN. Section 6.3 validates noise immunity of the proposed method by numerical studies on a FEM of a seven-storey steel frame. Networks are trained by using simulated ambient responses and tested by using a response with a high level of noise under earthquake. Section 6.4 introduces a numerical study on a nonlinear building model, which is conducted to assess the effectiveness of nonlinear response feature extraction. The training data is extremely limited in this study. Section 6.5 presents experimental studies with measured responses of a seven-storey steel frame. The limited ambient responses are used for training and the measured responses under a ground motion excitation are employed for testing. The performance of the proposed SegGAN is compared with that using a DenseNet (Chapter 5) and a traditional CNN. DenseNet with the same configuration as the generator is proposed to validate the advantage of the proposed approach involving discriminator. Section 6.6 concludes the results and findings of this study.

6.2 Methodology

The proposed SegGAN for structural response reconstruction is inspired by Isola et al. (2017). It consists of a generator and followed by a segment based discriminator with conditioned input. It is trained in a supervised manner based on an adversarial training strategy. Measured raw acceleration responses in time domain from available and desired locations are processed to train SegGAN. The specially designed generator is a fully CNN, which has a bottleneck structure with skip connection and dense blocks to advance information flow among layers, so as to improve the training efficiency and feature extraction capability. The input to the generator is the measured response from selected sensors that continuously work

properly, and the output is the reconstructed response from desired locations that are unavailable after a period of normal recording. The generator is proposed to learn the mapping between the input and output responses by using a combined objective function including a defined objective function and a self-learned objective function provided by the discriminator. For the case without the discriminator, the generator can be categorized as a type of network using designed objective functions such as mean absolute error (MAE) or mean squared error (MSE) to learn mappings, that tends to produce blurry prediction in pixel-to-pixel tasks (Pathak et al., 2016). The reason behind this is that the objective functions are minimised by averaging output values, which nonetheless capture the low-frequency structure of data such as frequencies and mode shapes of structures but lose some local details such as fast shifts of amplitudes and instantaneous frequency changes. The well trained generator of SegGAN will be then used to reconstruct structural responses of unavailable locations under both ambient and extreme loading conditions. It should be noticed that response reconstruction is fully automatic that only uses raw time domain data without any human selected features. The segment based discriminator is a fully stacked CNN that aims to provide additional loss gradients to the generator for reinforcing high-frequency data structure learning. The proposed discriminator has a conditioned input which is the available and reconstructed responses (input and output of generator) or available and true responses (input and ground truth of output of generator). A condition is given to the input so that the discriminator can consider both the authenticity of input and the consistency with the condition when discriminating the input. For response reconstruction, it means that rather than learning features from reconstructed and true responses, the conditioned input adds intra-channel relationships as the features, that brings more physics based significance to the reconstructed responses. The output of segment based discriminator is an array which contains scalars ranging from 0 to 1, indicating the ‘truth’ of each mapped input segment. Compared to the original GAN (Goodfellow et al., 2014) that evaluates the truth of full input data, segment based discriminator evaluates input segment-by-segment to learn the mapping of the high-frequency structure of data more effectively.

6.2.1 Generator with skip and dense connections

The generator is designed to reconstruct responses by training with a defined objective function and discriminator. The specially designed generator is a fully CNN with a bottleneck structure that is typically known as ‘U-net’ (Ronneberger et al., 2015) or convolutional encoder-decoder (Mao et al., 2016). Networks structured like a ‘bottleneck’ consist of compressive layers, bottleneck layer and reconstruction layers, which is developed to gradually extract the higher level of data features by the compressive layers and bottleneck layer, and then reconstruct the output using the most representative data features extracted by bottleneck layer through the stacked reconstruction layers. The compressive layers and bottleneck layer gradually shrink the size of feature maps to facilitate the network to learn robust features containing useful information of data while eliminating noise with no certain pattern. Skip connection is embedded among the network layers, and densely connected blocks are involved to substitute traditional compressive layers that are constituted of only a single convolution layer. Here, the generator with skip connection and densely connected blocks is named as DenseNet. These two techniques skip connect or densely connect convolutional layers to greatly improve the information flow among networks. As a result, the training data size and training parameters (including weights and bias of convolutional layers) can be reduced, and the gradient vanishing issue can be mitigated, resulting in the extraction and propagation of the higher level features more efficient and accurate. In addition, the convergence speed is also significantly boosted by reducing the computational demand and training iterations.

6.2.1.1 Configurations of generator

The architecture of the generator is demonstrated in Figure 6-1 with responses from multiple available locations served as input and one unavailable location as output. The low-level features of input data are first enriched by the *Low-level* convolutional layer. Three successive dense blocks (*Dense 1* to *3* as shown in Figure 6-1) containing densely connected convolutional layers are herein substituted by the layer-by-layer connected convolutional layers to form the compressive path for extracting high level of latent features. Data features

passing through each dense block will be halved in its length and doubled in the number of feature maps. The last convolutional layers in dense block shrink the length of feature maps by sliding the convolutional kernels with a stride of 2. The extracted features by *Dense 3* have the smallest dimension but the highest level of abstraction, which are used for reconstructing unavailable responses. Two successive reconstruction layers *R2* and *R1* and a final reshaping layer *Final*, as shown in Figure 6-1, are designed to reconstruct the unavailable responses using the features extracted by *Dense 3* and skip connections. Skip connections denoted by dashed arrows shuttle features of the shallow layers to the deeper layers. *R2* and *R1* upscale the length of feature maps by a factor of 2 using a one-dimensional sub-pixel shuffling operation. Figure 6-2 shows an example of the sub-pixel shuffling with four feature maps consisting of six 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. As a one-dimensional case presented in a previous study (Shi et al., 2016), this is an efficient operation for upscaling features, which costs less computational demand than the general deconvolution and has strong workability with less artifact introduced in the output (Odena et al., 2016).

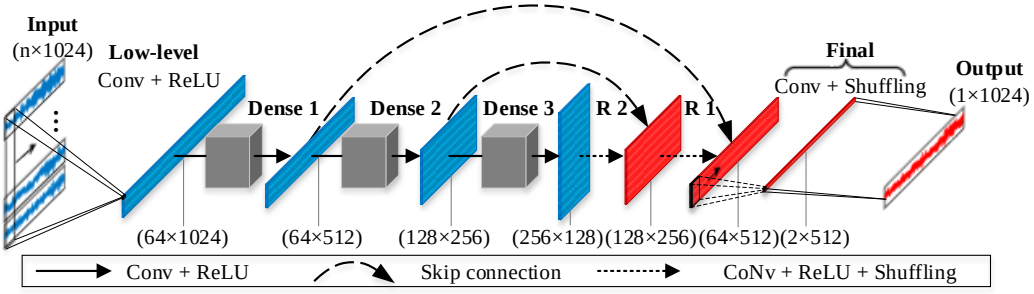


Figure 6-1. The architecture of the generator.

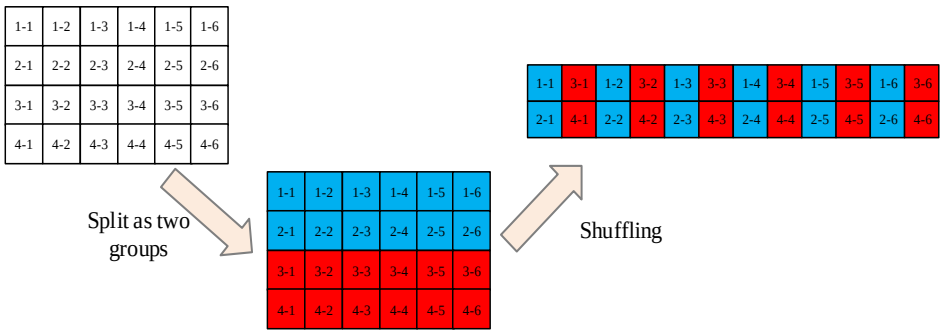


Figure 6-2. The schematic procedure of one-dimensional sub-pixel shuffling with 4 feature maps and 6 elements in each map.

The detailed configurations of the proposed generator are summarized in Table 6-1. The depth of DenseNet is determined by testing the performance on structural response reconstruction with 1 to 6 pairs of dense block and reconstruction layer. 2 pairs are selected since the performance stops gaining at this formation, while the computational burden continues increasing. Other model hyper-parameters including kernel number and kernel size, are determined by trial and error with an initial value derived from (Kuleshov et al., 2017). The mapping relationship between input and output is complicated so that a linear function cannot represent accurately. Leaky rectified linear unit, namely, *Leaky ReLU* (Maas et al., 2013) as an advanced activation function, is selected to transfer the mapping from linear to nonlinear (simplified as *ReLU* in Figure 6-1). The function of *Leaky ReLU* can be expressed as

$$Leaky\ ReLU(x) = \begin{cases} x, & if\ x > 0 \\ 0.01x, & \text{otherwise} \end{cases} \tag{6.1}$$

Leaky ReLU has a broader range of output which mitigates the gradient vanishing. Meanwhile, it allows a non-zero small gradient to overcome the defect of using traditional *ReLU*, that is, the neuron is potentially never activated when a large gradient passes it (Maas et al., 2013). Input and output shapes in Figure 6-1 and Table 6-1 are determined by considering the input having responses form *n* locations with a length of 1024 data points. Since there is no fully connected layer which needs a constant input to design the number of neurons, fully CNNs such as the proposed generator and discriminator accept input data with a flexible length.

Table 6-1. Detail configurations of the generator.

Layer	Kernel number	Kernel size	Stride	Input shape	Output shape	Leaky ReLU	Shuffling	Dropout
Input	-	-	-	(n, 1024)	(n, 1024)	-	-	N
Low level	64	64	1	(1, 1024)	(128, 1024)	Y	N	Y
Dense 1	64	64	-	(64, 1024)	(64, 512)	Y	N	Y
Dense 2	128	32	-	(64, 512)	(128, 256)	Y	N	Y
Dense 3	256	16	-	(128, 256)	(256, 128)	Y	N	Y
R2	256	16	1	(256, 128)	(128, 256)	Y	Y	Y
R1	128	32	1	(256, 256)	(64, 512)	Y	Y	Y
Final	2	64	1	(128, 512)	(1, 1024)	N	Y	N
Output	-	-	-	(1, 1024)	(1, 1024)	-	-	N

6.2.2.2 Skip connection

Training a deep neural network with a layer-to-layer sequential connection by using the gradient and back-propagation strategy always faces the gradient vanishing issue in deeper layers, which could significantly influence the effectiveness of feature extraction and lead the deep architecture to fail in training (Huang et al., 2017). Skip connection as shown in Figure 6-1 shuttles features extracted by shallow layers to deep layers to pass the low-level details that are lost during the convolution to the deep layers (Mao et al., 2016), which alleviates the gradient vanishing issue (Isola et al., 2017). Skip connection also allows the gradient to be back-propagated to the shallow layers directly, which further enhances the information flow and improves the convergence efficiency. Skip connection with these advantages is then embedded in the densely connected blocks to advance the feature extraction.

6.2.2.3 Dense block

Data feature extraction as a core part of the proposed DenseNet is emphasized by using dense block. A dense block consists of several convolutional layers, where any extracted feature maps are skip connected with all the successive ones (Huang et al., 2017; Zhang et al., 2018). Information flow among a dense block is strongly reinforced. Networks composed by dense blocks are much more efficient in extracting high-level features than the layer-by-layer sequential connected networks. Furthermore, instead of using a large single layer containing considerable parameters, dense block sharing parameters crossing the embedded layers can reduce the required parameters. The proposed dense blocks in this study have 4 densely connected convolutional layers, namely, *Conv 1* to *Conv 4*, as schematically shown in Figure 6-3. Detailed configurations of all these dense blocks are summarized in Table 6-2. Convolution layers inside a dense block sharing the same number of kernel size and kernel number. *Conv 1 to Conv 3* are three successive convolutional layers with a stride of 1 to keep the output size consistent. *Conv 4* shrinks the dimension of the features by a convolution operation with a stride of 2. Consequently, the sequential dense blocks perform as compressive layers in the bottleneck network.

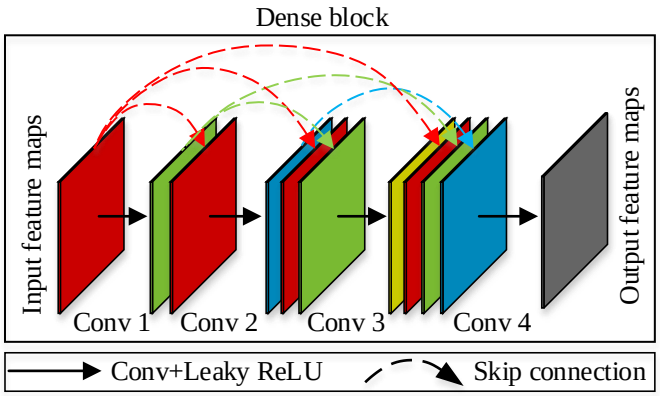


Figure 6-3. A dense block with four convolutional layers.

Table 6-2. The detailed configurations of the dense blocks.

	Layer	Kernel number	Kernel size	Stride	Input shape	Output shape	Leaky ReLU
Dense Block 1	*Conv1*	64	64	1	(128, 1024)	(64, 1024)	Y
	Conv2	64	64	1	(192, 1024)	(64, 1024)	Y
	Conv3	64	64	1	(256, 1024)	(64, 1024)	Y
	Conv4	64	64	2	(320, 1024)	(64, 512)	Y
Dense Block 2	*Conv1*	128	32	1	(64, 512)	(128, 512)	Y
	Conv2	128	32	1	(192, 512)	(128, 512)	Y
	Conv3	128	32	1	(320, 512)	(128, 512)	Y
	Conv4	128	32	2	(448, 512)	(128, 256)	Y
Dense Block 3	*Conv1*	256	16	1	(128, 256)	(256, 256)	Y
	Conv2	256	16	1	(384, 256)	(256, 256)	Y
	Conv3	256	16	1	(640, 256)	(256, 256)	Y
	Conv4	256	16	2	(896, 256)	(256, 128)	Y

6.2.2 Segment based discriminator

The segment based discriminator as shown in Figure 6-4 is proposed to learn the discrimination between true and reconstructed response. For response reconstruction, it can also be considered as establishing an objective function for training the generator on mapping details of response data. As denoted schematically by the specific segment component (denoted in purple colour) in Figure 6-4, a prediction unit is mapping from a segment of input data. The segment based discriminator determines the truth of input segment-by-segment, which learns more detail information from training data. This is the major difference between

SegGAN and the original GAN (Goodfellow et al., 2014) where the output of original GAN is a single scalar that signifies ‘real’ or ‘fake’ of full input data. Furthermore, as introduced in Section 6.2, the input of the discriminator is conditioned by the input of the generator which brings more physics based significance for the learned objective function for the generator that allows generator to produce more reasonable responses. As a result, SegGAN is enhanced to not only reconstruct response as being close to ‘truth’ as possible, but also keep consistency with responses from other channels. Configurations of the discriminator are demonstrated in Figure 6-4, where *k*, *ks* and *s* represent the kernel number, kernel size and convolution stride, respectively. High-level features are progressively extracted by stacked convolutional layers from *C1* to *C5*. It is worth noting that the last convolutional layer *C5* uses a *Sigmoid* activation function, which transfers the output to a range from 0 to 1 as a logistic likelihood indicating the truth of prediction from false to truth. The expression of *Sigmoid* activation function is expressed as

$$Sigmoid(x) = \frac{e^x}{e^x + 1} \tag{6.2}$$

To enforce the discriminator to learn the discrepancy between reconstructed and true responses in training, the output labels for the reconstructed and true responses are set as a series of 0 and 1, respectively, to compute gradients.

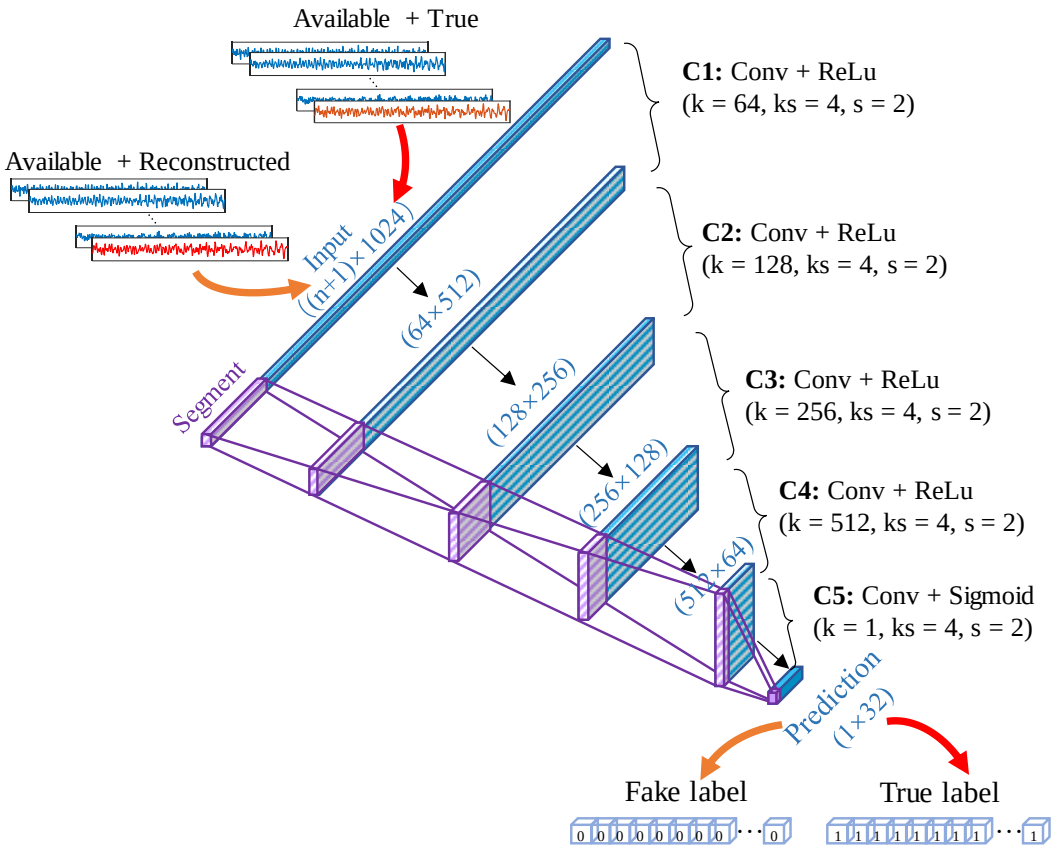


Figure 6-4. The architecture of the proposed segment based discriminator.

6.2.3 Objective function of SegGAN

For response reconstruction, the objective function of training SegGAN is to maximize the 'truth' of reconstructed responses from generator, which is determined by discriminator through an adversarial framework. During the training process, the generator tries to confuse the discriminator by learning the mapping between responses from available and desired locations and generates desired responses that appear as true. The discriminator learns the features of generated and real data and tries to allocate the maximum possibility of truth to true response and maximum possibility of fake to reconstructed response. The objective function of SegGAN for response reconstruction with conditioned input is expressed as

$$L_{SegGAN} = \arg\min_{G}\max_{D} \mathbb{E}_{a,t}[logD(a,t)] + \mathbb{E}_{a}[\log(1 - D(a, G(a)))] \tag{6.3}$$

where G and D are the generator and discriminator of SegGAN, $\mathbb{E}$ means the expected output, a and t denote the available and true responses, respectively. Here, G tries to minimize this

objective function against an adversarial *D* that tries to maximize it. Generator and discriminator play a two-player minimax game following Equation (6.3) to generate cross-entropy loss for tuning parameters. It is reported that MAE and MSE based losses that evaluate the distance between data have strong capability on learning low-frequency data features (Radford et al., 2015). To facilitate the learning of low-frequency data features, the loss of generator is herein designed to be supplied by the discriminator and a distance sense objective function. As a result, the objective of the generator becomes producing responses to cheat discriminator, meanwhile, approaching the true responses in a distance sense. Generator and discriminator will be trained simultaneously, but only the generator will remain for response reconstruction after training. The final objective function for the generator is expressed as

$$L_G = L_{SegGAN} + \lambda \cdot Loss_{distance} \tag{6.4}$$

where λ is the weight for scaling $Loss_{distance}$. The reason for applying a loss weight is because generally, L_{SegGAN} has a much larger scale than $Loss_{distance}$ which will influence the effectiveness of training. Here, Normalized MSE (N-MSE) loss function that has been validated to be effective in response reconstruction as introduced in Chapter 5 is recommended as the $Loss_{distance}$. N-MSE is defined as

$$L_{N-MSE} = \sum_{k=1}^{n} \frac{(t_k - G(a_k))^2}{{t_k}^2} \tag{6.5}$$

where *n* is the total number of sampling points of a training sample. A loss weight λ of 1000 is recommended based on Ref. (Pathak et al., 2016).

6.2.4 Adversarial training procedure of SegGAN

Training of the proposed SegGAN is realized in a supervised manner as shown in Figure 6-5. The label for generator is the corresponding true response of generated response, and the label for discriminator are 0 and 1 indicating the truth of data corresponding to reconstructed and true responses. Before starting the training, all trainable parameters of both the generator and discriminator are initialized with a normal distribution. The training of SegGAN consists of two adversarial stages. The first stage of training will freeze parameters of the generator

and train discriminator by using training data loaded in batches. Available responses from sensor measurements flow through the generator to generate reconstructed responses and then mix with true responses to form two sets of conditioned input for training the discriminator. The discriminator is then trained to distinguish generated and true responses by using computed cross entropy loss. Gradients are derived and back-propagated to each layer of the discriminator. Once a loop of training data finishes, the parameters of the discriminator will be fixed, and then parameters of the generator will be turned to trainable. Available responses are again used as input to the generator for producing the generated responses. Combined with the available responses, the generated responses are served as input to the discriminator for computing a cross entropy loss. On the other hand, N-MSE loss of generated reconstructed response is computed with the true response to be reconstructed. Two losses are summed with assigned loss weight. Gradients are back-propagated to the parameters of the generator. The training data grouped as batches are used in turn to fine-tune generator. These two training steps compose of an iteration of training. Generally, the required iteration for training a network depends on the volume of training data. With this specially designed architecture, training a SegGAN with competitive feature extraction capability can take much fewer iterations, compared with the traditional CNN.

To overcome the overfitting issue, dropout technique (Srivastava et al., 2014) is applied in each layer of the generator which randomly de-actives neurons of the convolutional kernel to break up the co-adapted sets of neurons. Remaining neurons are forced to learn data features to eliminate noise effect, which makes the trained neurons more robust. In this study, every neuron has a 50% possibility of being dropped when confronting a different batch of training data.

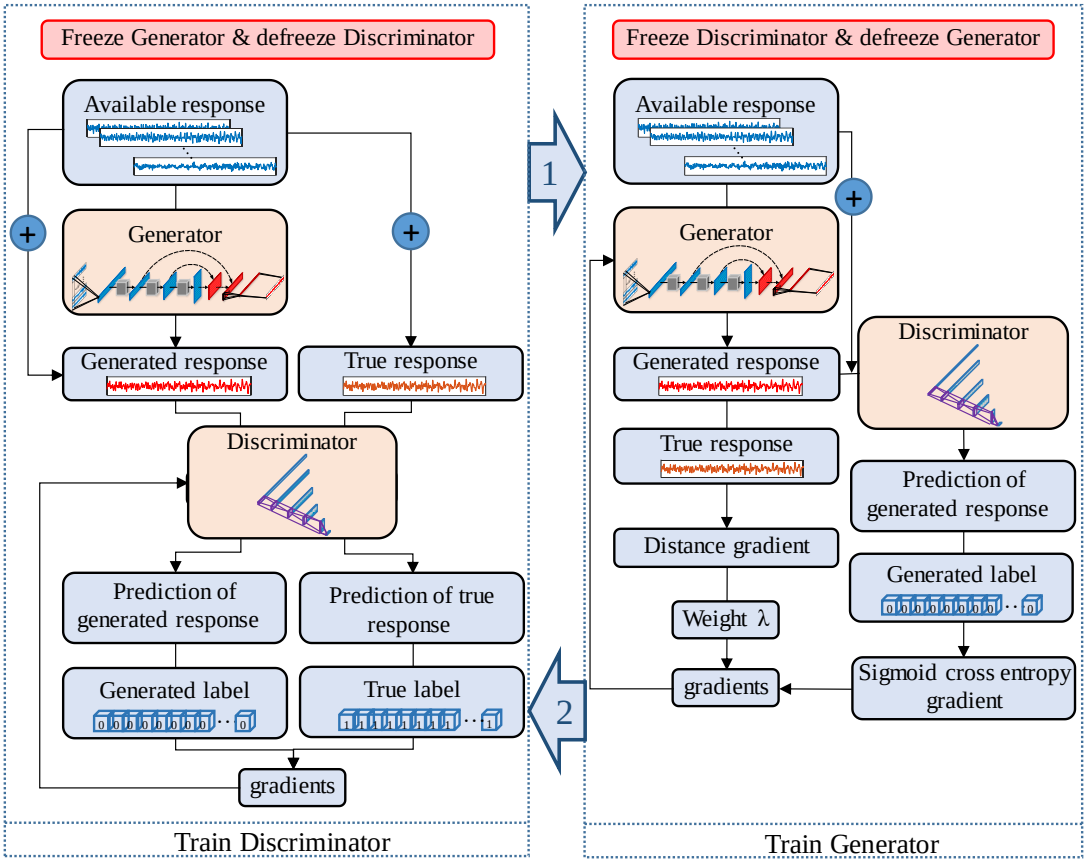


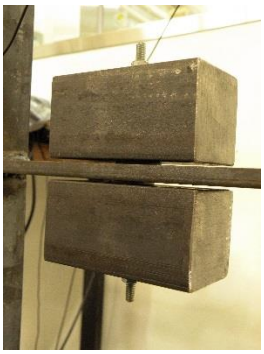
Figure 6-5. Adversarial training procedure of the proposed SegGAN.

6.3 Numerical study on a linear steel frame model

To validate the effectiveness and accuracy of the proposed SegGAN for structural response reconstruction, numerical studies are conducted based on a FEM of an experimental seven-storey steel frame structure as shown in Figure 6-6. In this steel frame structure model, each floor height is 0.3m with a total model height of 2.1m. The width is 0.5m. The cross sections of the column and beam elements are 50mm × 5mm and 50mm × 9 mm, respectively. The mass density and elastic modulus of material are 7850kg/m^3 and 210GPa, respectively. A 4 kg lumped mass is added on the corresponding DOFs at a quarter and three-quarter locations of each floor to simulate the mass on the floor of the building. An initial FEM model of the frame structure is built with planar beam elements. Each node in the FEM consists of three DOFs, which are the translational displacements in the *x* and *y* directions and the rotational displacement in the *x*-*y* plane. The FEM of the frame structure includes 70 planar elements and 65 nodes as shown in Figure 6-7. With 3 DOFs at each node, the model has 195 DOFs in

total. Two fixed supports at node 1 and 65 are modelled by restraining the corresponding DOFs. The first two modal damping ratios of the experimental structure are computed as 1.1% and 0.4% for the first two modes with the sensitivity based model updating method (Li et al., 2017). Rayleigh damping model is assumed for this structure. The stiffness matrix and boundary conditions of the initial FEM are then updated using the measured responses from the experimental structure as introduced in Ref. (Li et al., 2012). It should be noted that the updated model is used as a baseline in numerical studies for generating the training, validation and testing data.

(a)



(b)

(c)

Figure 6-6. Experimental seven-storey frame structure. (a) Frame structure; (b) Mass blocks; (c) Fixed supports.

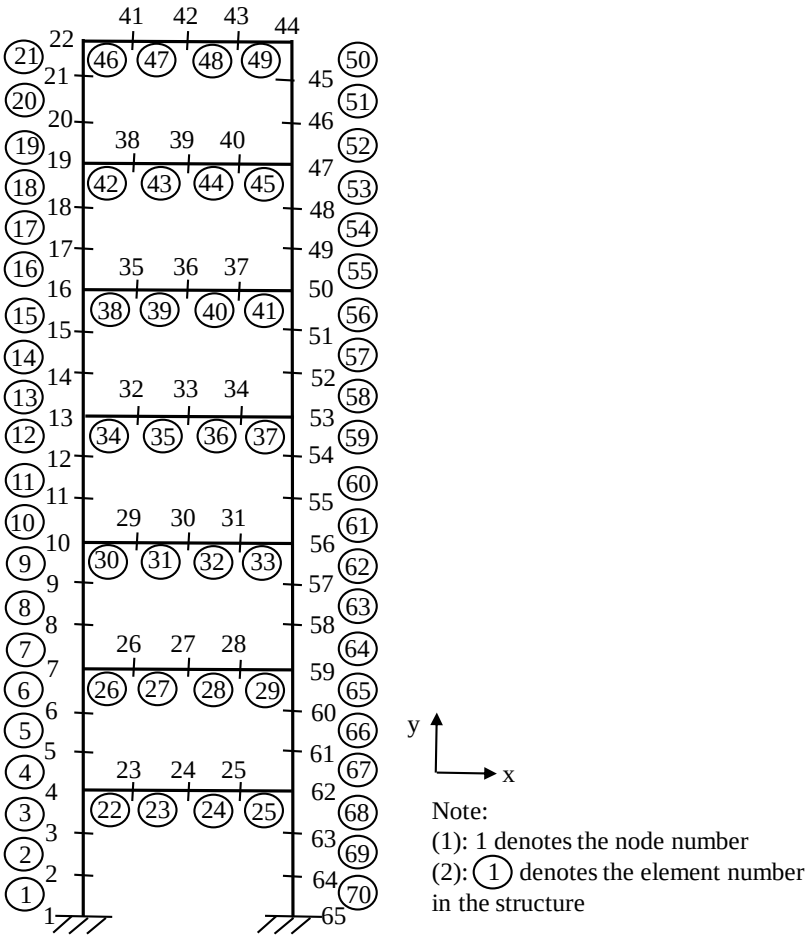


Figure 6-7. FEM of the steel frame structure.

6.3.1 Data preparation

Training and validation data are generated as ten-hour dynamic responses of this frame structure under ambient excitations along *x* direction. Dynamic responses of the structure are computed by using Newmark-beta method. The 'measured' acceleration responses in *x* direction from nodes 4, 7, 10, 13, 16 and 19 are considered as available responses and responses at node 22 in *x* direction is assumed to be unavailable after a period of normal recording. It should be noticed that for response reconstruction, the available responses may need to contain all the significant vibrational modes excited at the desired location when using methods based on intra-channel correlations for accurate reconstruction. Using available responses from multiple DOFs can fulfil the requirement. The sampling rate of acceleration data is set as 100Hz. Original measurements from all the selected channels are altogether scaled to a range between -1 and 1 for training stability while retaining mode shapes

information. These ten-hour data are processed as training and validation samples and each sample has a length of 1024 sampling points. A window with a length of 1024 that slides along the time steps of acceleration data with a stride of 512, is applied to generate samples. Generated samples are split as 80% and 20% for training and real-time validating the convergence of networks, respectively. Since validation data are never used to tune parameters during training, they are used for both validation and testing networks to allow more data for training. Testing data are prepared including validation data under ambient excitation and extra datasets, that is, structural responses under earthquake excitations. 20% root mean square Gaussian noise is added to the testing data for simulating measurement noise. The earthquake applied is the recorded acceleration of the El Centro earthquake as shown in Figure 6-8. The condition is harsh for response reconstruction for this scenario since the features of testing data is different from training data, because the excitations are different. Vibration responses under ambient excitations are used for training, but the responses under earthquake excitations are used for testing the proposed approach.

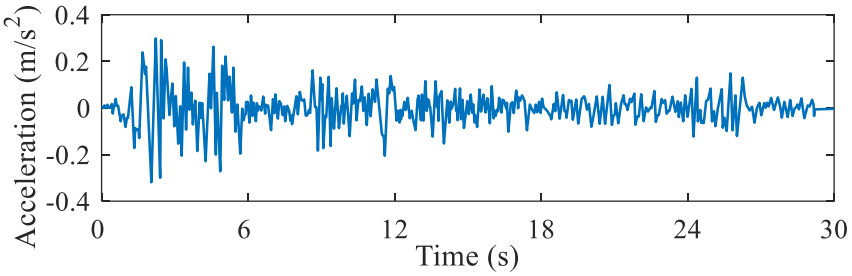


Figure 6-8. Input El Centro earthquake.

6.3.2 Selection of objective function for generator

A suitable objective function is crucial for different concerned tasks. In this study, the objective function of the generator is carefully selected by testing the performance of using several defined objective functions with different loss weights to work together with the adversarial loss. The combination of losses is expressed in Equation (6.4). Four losses with two commonly used losses, i.e., MAE and MSE, and two designed losses, namely, normalized MAE (N-MAE) and N-MSE, are tested. The objective function of N-MSE is introduced in Equation (6.5). N-MAE can be expressed as

$$L_{N-MAE} = \sum_{k=1}^{n} \frac{|t_k - G(a_k)|}{|t_k|} \tag{6.6}$$

The reconstructed response is evaluated with the corresponding true response. The reconstruction error is quantified as a relative error with the true response using L2-nrom which can be expressed as

$$\varepsilon = \frac{\|t - G(a)\|_2^2}{\|t\|_2^2} = \frac{\sqrt{\sum_{i=1}^{n}(t_i - G(a)_i)^2}}{\sqrt{\sum_{i=1}^{n} {t_i}^2}} \tag{6.7}$$

Frequency domain responses representing vibration characteristics in the frequency domain sometimes are important for SHM. Therefore, this study also evaluates the reconstructed response in the frequency domain through Fast Fourier Transfer (FFT). When calculating the error in the frequency domain, *t* and *G(a)* become the spectra of true and reconstructed responses. SegGANs with different loss function combinations are trained individually using a computer with an i7-6700k CPU, RTX 2080TI GPU and 16GB memory. The training iteration is 150 epochs and training time is half an hour for each one. The validation errors of selected objective function and different loss weights combinations are shown in Figure 6-9. The errors are small when cross entropy loss L_{SegGAN} is not involved in the generator. However, without using L_{SegGAN}, the detail information loss problem will exist. Those results win quantitatively but lose in the quality. More explanation will be provided in the following sections. Errors of these four loss functions are large when the loss weight is 10, since L_{SegGAN} is supposed to facilitate networks to learn detail features. Assigning a large contribution will influence the learning of the low-frequency structure of data. Except for MSE, errors drop sharply when the loss weight reaches 100 and slightly fluctuate with continually increasing loss weight. The smallest error of 8.67% is achieved by using N-MSE with a loss weight of 1000, which is the same as our recommendation mentioned in Section 6.2.3 and consistent with the loss function used in Chapter 5 for response reconstruction and loss weight selected in Ref. (Pathak et al., 2016) for image inpainting. As a result, this combination is selected for the rest of studies.

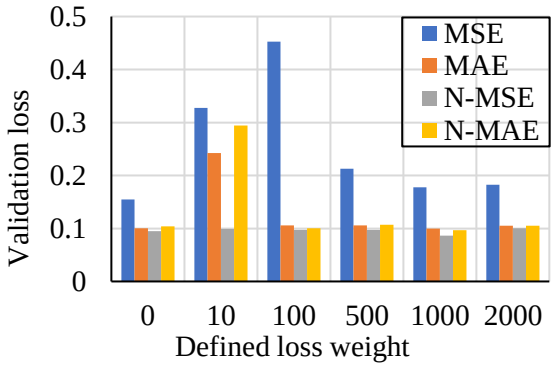
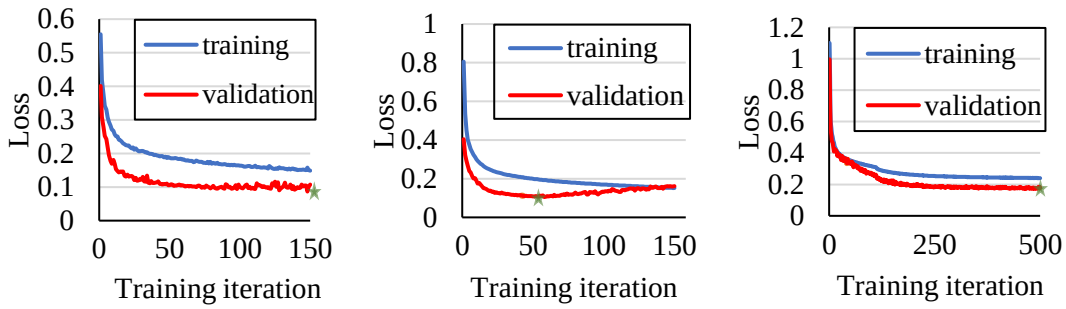


Figure 6-9. Validation errors of selected objective functions with different loss weights.

6.3.3 Results evaluation

The trained generator of SegGAN with best combination is then validated by testing data. DenseNet, and a traditional CNN with the same architecture as DenseNet but without skip connection and dense blocks, are applied for comparison. Trainings of SegGAN, DenseNet and CNN are conducted by using 150, 150 and 500 epochs and take around 0.6, 0.5 and 1.5 hours, respectively. The convergence processes for these three networks are shown in Figure 6-10. During training, parameters of networks with the best performance are kept for testing. The best points are marked as green stars in Figure 6-10. Training and validation losses of SegGAN and DenseNet drop suddenly in the first several training iterations because of the involved skip connection and dense blocks which effectively improve the feature extraction. As shown in Figure 6-10(a), the validation loss of SegGAN decreases slowly after 50 iterations and reaches the best result of 8.67% at the 148th training iteration. The validation error is always below training error, which means that no overfitting issue exists in training SegGAN. On the other hand, as shown in Figure 6-10(b), DenseNet achieves the smallest loss of 10.54% and then the loss starts raising after 60 training iterations. This phenomenon indicates that the networks overfit the training data after 60 iterations and adjustments should be made on the networks for training. On contrary, SegGAN considering both N-MSE and self-learned losses is more robust to overfitting, which represents a stable training process. Figure 6-10(c) shows the training process of CNN for 500 epochs, and CNN reaches the smallest validation loss of 17.14% at the 488th iteration. Compared to SegGAN and DenseNet, the training efficiency is much lower, and the smallest loss is also much larger than those obtained from SegGAN and

DenseNet. The comparison results further demonstrate the effectiveness and improvement of using the skip connection and dense blocks in feature extraction. When comparing the performances of SegGAN and DenseNet, training of SegGAN is more stable and the results are quantitatively better with smaller loss values.



(a) (b) (c)

Figure 6-10. Training processes of (a) SegGAN; (b) DenseNet; and (c) CNN.

The testing data are then used as input to the trained networks for evaluating the effectiveness on reconstructing responses under ambient and extreme loading conditions. Figures 6-11 and 6-12 show the comparison of reconstructed and true responses under ambient excitations and earthquake loads in both time and frequency domains. The relative errors in response reconstruction by applying these three trained networks are shown in Figures 6-13(a) and (b) under ambient and earthquake excitations, respectively. Overall, as can be observed in Figures 6-11(a), (c) and (e), all these three networks effectively reconstruct the ambient responses in time domain. Though SegGAN slightly outperforms DenseNet, they are both outstanding for response reconstruction under ambient excitations. CNN may be not as accurate as SegGAN and DenseNet in aligning the amplitudes of time steps due to feature loss during convolution. On the other hand, by observing Figures 6-11(b) and (d), natural frequencies as important information of structural response are accurately reconstructed in terms of the frequency and amplitude values. SegGAN with a self-learned loss function is stronger than DenseNet in learning and reconstructing spectrum amplitude which is correlated to the mode shape of the structure as a vital vibrational characteristic for SHM. CNN produces good results in reconstructing several well-excited modes at this DOF as shown in Figure 6-11(f). However, CNN loses information of responses in higher frequency bandwidths due to

the simple layer-to-layer connected structure which might not be effective in higher frequency feature extraction. The observed results are consistent with the quantitative errors as shown in Figure 6-13(a). Reconstructed response by using SegGAN maintains the smallest relative errors while CNN has the largest errors in both the time and frequency domains.

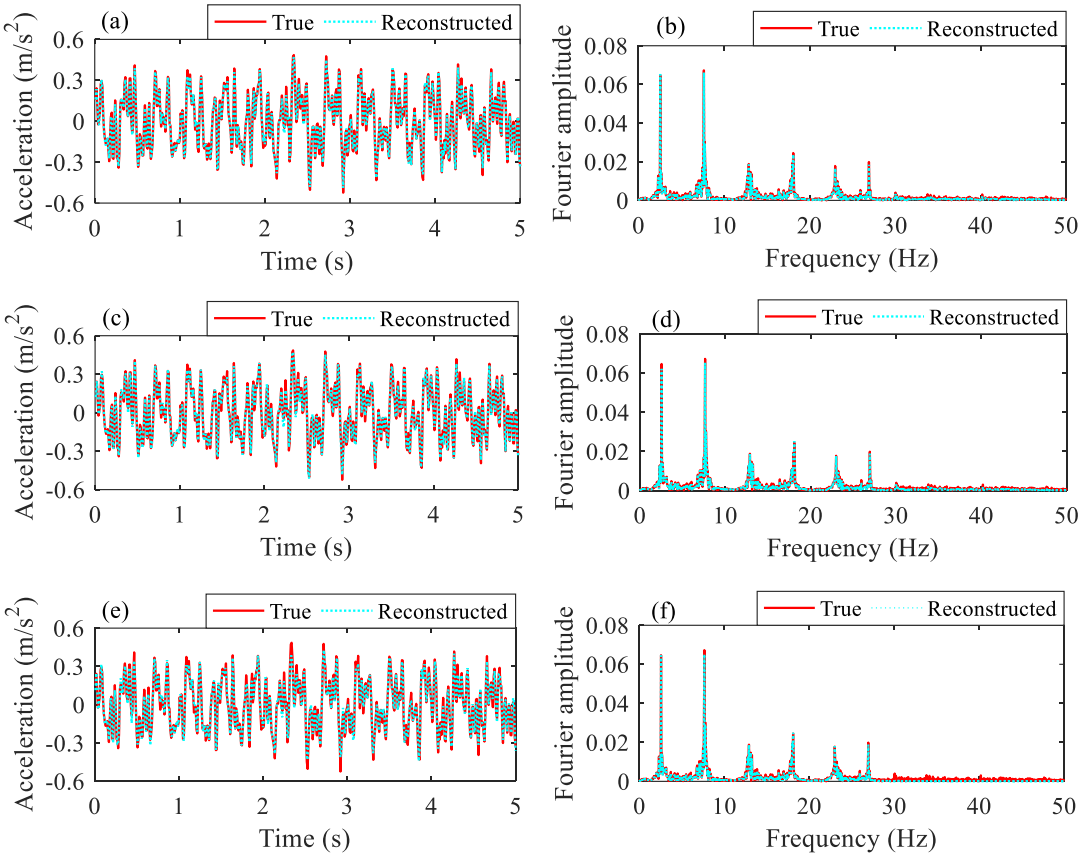


Figure 6-11. Comparison of 5 seconds true and reconstructed ambient responses in time and frequency domains under ambient excitations using: SegGAN (a) and (b); DenseNet (c) and (d) and CNN (e) and (f).

Reconstruction of earthquake responses with high measurement noise is much more challenging. As shown in Figures 6-12(a) and (b), SegGAN demonstrates a strong capability in reconstructing response details because of the adversarial framework. Variation in system input will result in differences in the features of system responses. Trained by ambient responses, SegGAN effectively reconstructs responses under earthquake excitations with dominating power concentrated in a narrow frequency bandwidth. Observed from Figure 6-12(c), time domain response reconstructed by DenseNet shows a large discrepancy in the maximum amplitude of responses which is important for evaluating earthquake induced

damage. The reconstruction performance in time domain of CNN as shown in Figure 6-12(e) is not as good as SegGAN. The reconstructed response exhibits a large discrepancy in the last 3 seconds. Both DenseNet and CNN using a pre-designed objective function alone lose some detail information of responses in different levels. SegGAN self-learns an objective function through adversarial training process, which can greatly improve the detail feature extractions. By comparing Figures 6-12(b), (d) and (f), there is no large difference between reconstructed responses in the frequency domain. The reason behind this is that the frequency spectrum obtained by using FFT cannot present the time domain response details but provide modal information only. The observations are consistent with the quantitative evaluation of reconstructed responses as shown in Figure 6-13(b) where the reconstruction errors in frequency domain are generally lower than that in time domain. From this study, it can be concluded that even CNN can automatically extract linear response features which are natural frequencies and mode shapes of a structure from civil engineers' perspective. DenseNet and SegGAN are more effective in response features extraction than CNN, especially SegGAN, which also captures local feature variations. In addition, a high level of noise is injected in the testing data but introduces minor effect on reconstruction. This indicates that networks with bottlenecked architecture have strong immunity to noise effect and can provide an accurate response reconstruction in both time and frequency domains.

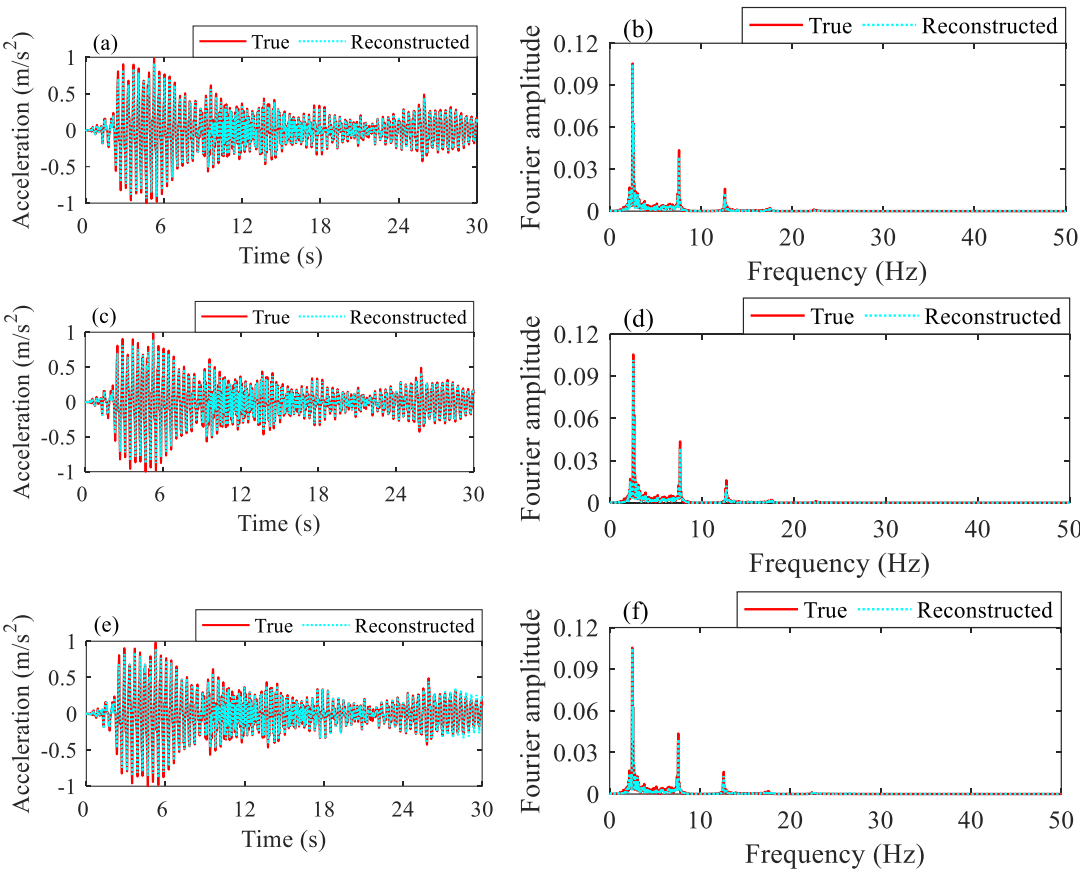
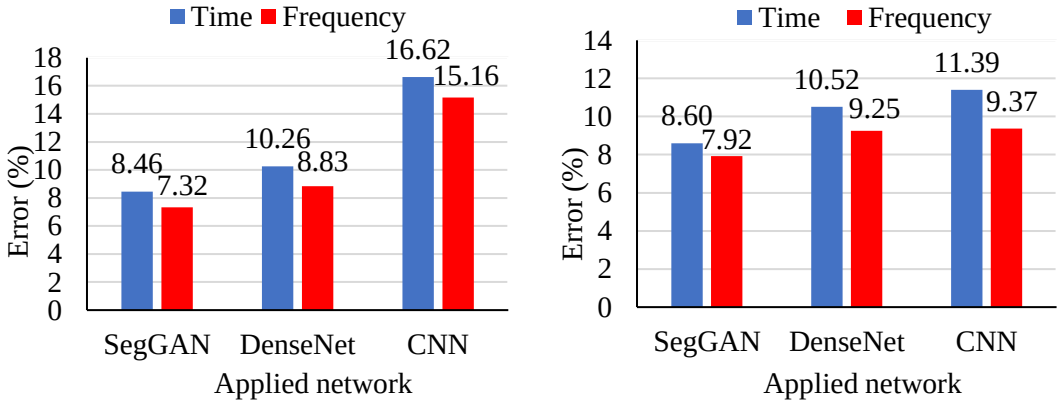


Figure 6-12. Comparison of true and reconstructed earthquake responses in time and frequency domains under earthquake loads using: SegGAN (a) and (b); DenseNet (c) and (d); and CNN (e) and (f).



(a) (b)

Figure 6-13. Relative errors in response reconstruction under: (a) ambient excitations; (b) earthquake loads.

6.4 Numerical study on a nonlinear building model

A nonlinear building model is used for evaluating the accuracy of using the proposed approach in capturing the nonlinear response features for structural response reconstruction, which exist widely in structures under extreme loading conditions. The nonlinear FEM of the building is a shear model consisting of 7 DOFs. As shown in Figure 6-14, the first storey of the building is 3.5m in height and the rest stories are 3.0m high. The first storey and the top storey are assigned with a mass of 1.6×10^4kg and 1.25×10^4kg, respectively. The masses of other stories are defined as 1.4×10^4kg. The 2nd to 7th stories use linear elastic elements with a cross section of 0.2m×0.2m and an elastic modulus of 52GPa. The first storey is assumed to have nonlinear behaviour following a Bouc-Wen hysteretic model (Wang et al., 2015). The equation of motion of the nonlinear building model can be expressed as

$$m\ddot{u}(t) + c\dot{u}(t) + f_{NL}[u(t), t] = f(t) \tag{6.8}$$

where $f_{NL}[u(t), t]$ is the nonlinear restoring force of the system which can be expressed as

$$f_{NL}[u(t), t] = \alpha k_0 u(t) + (1 - a)kz \tag{6.9}$$

where α is the rigidity ratio which is set as 0.05 in this study, k_0 as the initial tangent stiffness is defined as 8.24×10^7 kN, z denotes the hysteretic displacement response which can be further computed as

$$\dot{z} = \frac{A\dot{u} - \upsilon(\beta\dot{u}|z|^n + \gamma|\dot{u}||z|^{n-1}z)}{\eta}, \tag{6.10}$$

with

$$\nu = 1.0 + \delta_\nu \varepsilon,$$
$$\eta = 1.0 + \delta_\eta \varepsilon,$$
$$\varepsilon = \int_0^t z\dot{u}(\tau)\, d\tau$$

where β and γ are the parameters defining the hysteretic behaviour of the nonlinear Bouc-Wen model; A is the coefficient of tangent stiffness k_0, n determines the transition from the linear to nonlinear ranges, η and υ define the stiffness and strength degradation, respectively. In this study, the parameters of the Bouc-Wen model are selected as $\alpha = 0.05$, $\beta = \gamma = 2000$, n=1, $\eta = 1.0$, $\nu = 1.0$, respectively.

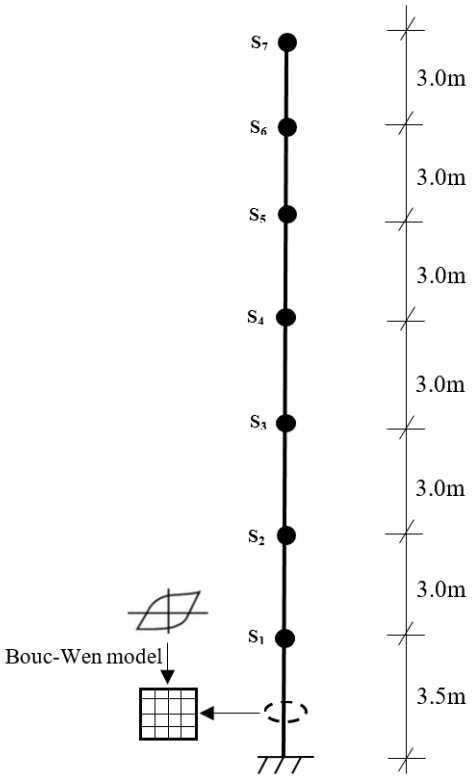


Figure 6-14. FEM of the nonlinear building model.

6.4.1 Networks training using nonlinear response data

The building model is excited by four different earthquake ground motions records from low to high intensities of ground motions. Four sets of responses recorded at the 1st, 2nd and 4th ground motions are selected for training and the 3rd one is kept for testing. The sampling rate of acceleration response data is set as 50Hz. The training samples are generated by a window with a length of 256 and sliding along the responses with a stride length of 128. There are totally 32 samples with a length of 256 in the training datasets. In this study, the response of the top storey sensor S_7 is assumed to be lost during the third earthquake excitation and needs to be reconstructed. The measured responses from sensors *S1* to *S6* are selected as the available responses to serve as the input, and the response from S_7 becomes the output of networks for response reconstruction. Responses from all the sensors are altogether scaled to a range between -1 and 1 for improving the stability of training, while retaining mode shape information. The scaled data are processed as training and validation samples by using a sliding window with a length of 256 and a stride of 128. Testing data is also scaled but kept as the original size. After reconstruction, the output data can be rescaled back to real scale by

dividing the same scale factor with the input. Three networks, namely, the proposed SegGAN, DenseNet and CNN, are assigned the same parameters as the numerical studies on the frame structure. With very limited data, these three networks are trained by 1000, 1000 and 5000 iterations for feature extraction. It takes 4min, 3min and 8min for these three networks to fully converge.

6.4.2 Testing trained networks

After training, the three trained networks are tested by using the third earthquake response that is not involved in training. The time records of true and reconstructed responses are shown in Figure 6-15, and the quantified relative errors of reconstructed responses are shown in Figure 6-16. As shown in Figures 6-15(a) and (b), responses reconstructed by using the proposed SegGAN accurately match with the corresponding true responses with minor differences in a few amplitudes in both time and frequency domains. The observations are confirmed by Figure 6-16, which shows that the relative errors in time and frequency domains are around 22% and 16%, respectively. The results demonstrate that the designed SegGAN can effectively capture both high-frequency and low-frequency features of earthquake responses. On contrary, by observing Figures 6-15(c) and (d), there are many differences in the amplitudes in the reconstructed response time history, because very limited training data are available. DenseNet assigns higher weights to low frequency components that have more energy under earthquake excitations. DenseNet does not perform well to detect response details and capture local changes, which means that it only learns low-frequency features of responses but loses time-variant features induced by the nonlinearity and high frequency components. If more datasets are available, the accuracy might be better. Furthermore, as shown in Figures 6-15(e) and (f), CNN with simple architectures and trained with such limited samples is ineffective in extracting both low and high-frequency responses. Therefore larger relative errors are obtained. In this study, SegGAN trained by adversarial strategy outperforms the other two networks on nonlinear response feature extraction under extreme loading conditions. With an outstanding capability on nonlinear response feature extraction with

limited training samples, SegGAN is of great potential for practical applications where limited data under extreme loading conditions are available.

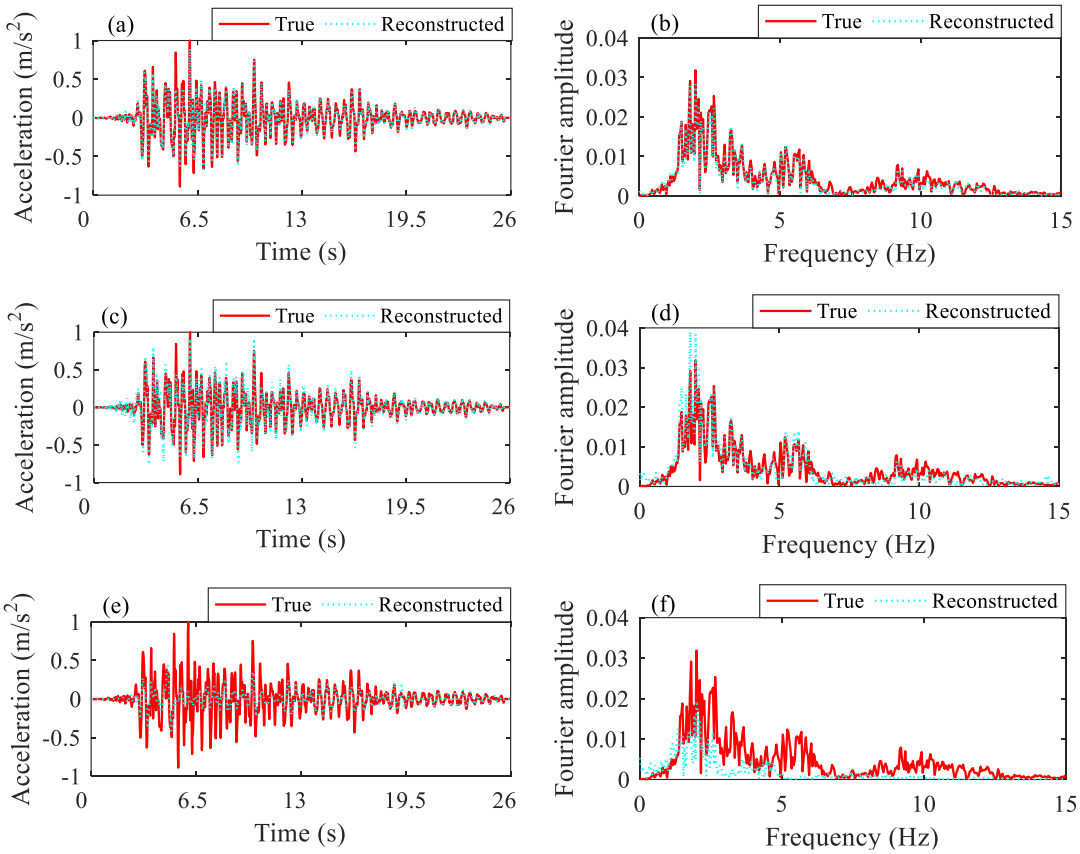


Figure 6-15. Comparison of true and reconstructed earthquake responses in time and frequency domains using: SegGAN (a) and (b); DenseNet (c) and (d); and CNN (e) and (f).

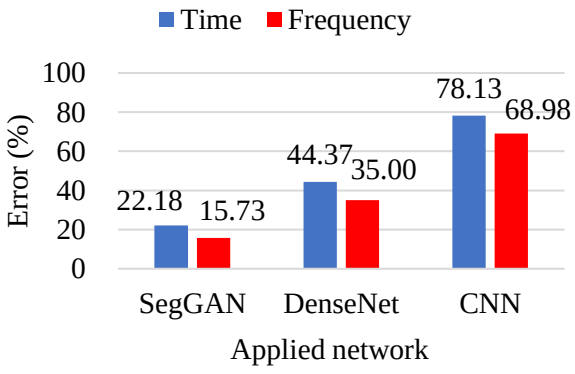


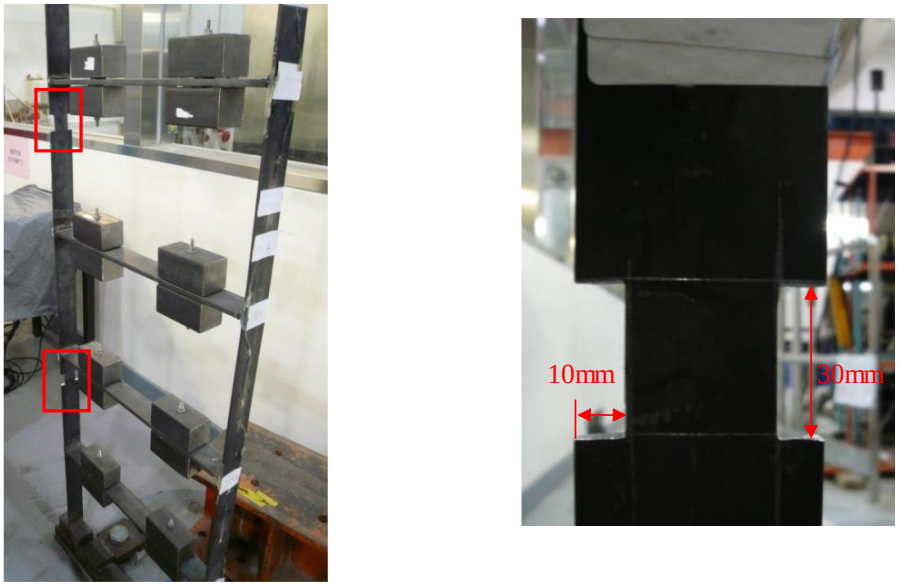
Figure 6-16. Reconstruction errors of nonlinear data in time and frequency domains.

6.5 Experimental validations

To further validate the performance of using the proposed approach for real applications, experimental studies are conducted with limited measured response from the steel frame structure as described in Section 6.3. As shown in Figure 6-17(a), minor damages are introduced into the column of the second and fourth floors of the structure corresponding to elements No. 6 and No. 12 of the FEM, as shown in Figure 6-7. Damages are introduced into the element as two cuts with a width of 30mm and a depth of 10mm as shown in Figure 6-17(b).

Ambient and shake table tests are conducted for the above damaged structure. In the test, totally 7 sensors are mounted at one side of the column. The locations of sensors are at beam-column joint of each floor of the steel frame. The top sensor is assumed to be faulty and responses of the rest of sensors are used as available responses. Networks are trained using measured ambient responses and validated by measured responses under shake table tests with an input earthquake excitation. About 14 minutes of ambient responses of the steel frame structure with a sampling rate of 2000Hz are recorded. The first seven modes of this structure are within 30Hz. To eliminate redundant features from data and shirk the size of dataset, the measured data are pre-processed by a low-pass filter with a 50Hz cut-off frequency and then down-sampled to 100Hz before processed as training data. Shake table test is conducted to obtain responses of the structure for testing the performance of the proposed approach. The preparation of training and validation samples follows the same procedure as the numerical study as introduced in Section 6.3.1. Consequently, with the limited measured ambient vibration responses, only 128 training samples and 32 validation samples with a length of 1024 are produced. Generally, a large number of training samples are required for most of the deep learning applications in SHM to achieve a good accuracy. In this study, the limited number of samples leads to the significantly improved efficiency of feature extraction. SegGAN possesses a strong capability in feature extraction from limited training data in the numerical study of the nonlinear building model. In this experimental study, the efficiency of feature

extraction is also validated by using experimental data measured from the steel frame structures under ambient conditions.



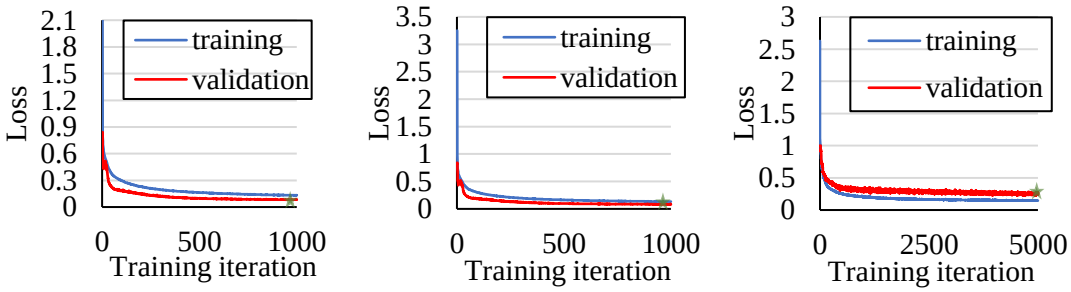
(a) (b)

Figure 6-17. (a) Frame structure with damage; (b) Detail of the damaged element.

6.5.1 Training networks

Three networks including the proposed SegGAN, DenseNet and CNN, are used in the experimental studies with the same configurations as those in numerical studies. These three networks are trained respectively with the same training data. The training time for SegGAN, DenseNet and CNN are around 10min, 8min, and 26min for 1000, 1000 and 5000 training iterations. The training and validation errors for these three networks are demonstrated in Figure 6-18. By observing Figures 6-18(a) and (b), SegGAN performs the best feature extraction capability, and then DenseNet. The losses of using SegGAN and DensetNet drop sharply for the first few iterations. The validation errors of SegGAN and DenseNet reduce gently until reaching the valley points of 0.08 and 0.09 at the 950^{th} and 930^{th} iterations, respectively. There is no overfitting issue during training these two networks, which demonstrates that SegGAN and DenseNet efficiently learn robust and representative features from training data. Similar to the results in numerical studies, the convergence efficiency of CNN is much lower than SegGAN and DenseNet, and the minimum validation error of CNN

is much larger, which retains 0.27 after around 5000 training iterations. With the measurement noise and operational conditions in the ambient vibration responses, traditional CNN becomes ineffective in extracting abstractive features from very limited measurement data. The comparison results indicate that skip and dense connections significantly improve the feature extraction and information flow among networks. In addition, considering that the number of training samples is extremely limited, SegGAN and DenseNet exhibit excellent feature extraction capability, demonstrating high application potential in reconstructing lost in-field measurement data.



(a) (b) (c)

Figure 6-18. Training processes in experimental studies using: (a) SegGAN; (b) DenseNet; (c) CNN.

6.5.2 Testing data evaluation

By using the trained networks, measured earthquake responses are reconstructed as shown in Figure 6-19. With less nonlinearity in this test, SegGAN and DenseNet reconstruct responses accurately in both time and frequency domains. Comparing the results in Figures 6-19(a) and (b) with those in Figures 6-19(c) and (d), SegGAN performs better in reconstructing the response time history amplitudes than DenseNet. The superiority of using SegGAN can be attributed to the involved segment based discriminator which facilitates the generator to learn detail or local features of responses. On the other hand, as shown in Figures 6-19(e) and (f), CNN presents less accuracy to extract the detail information of earthquake responses and reconstructs inaccurate responses time history amplitudes. The calculated relative errors as shown in Figure 6-20 are consistent with the observed results in Figure 6-19, where SegGAN holds the smallest errors in both time and frequency domains in reconstructing measured

earthquake responses. Recalling the critical condition for training and testing, where limited data from ambient excitations only are available and no earthquake response is included in the training process, the good results demonstrate the effectiveness and robustness of using SegGAN for structural response reconstruction. SegGAN extracts robust response features efficiently and accurately, which is therefore very applicable for response reconstruction of structures under extreme loading conditions, even though only responses under ambient excitations are included in the training datasets.

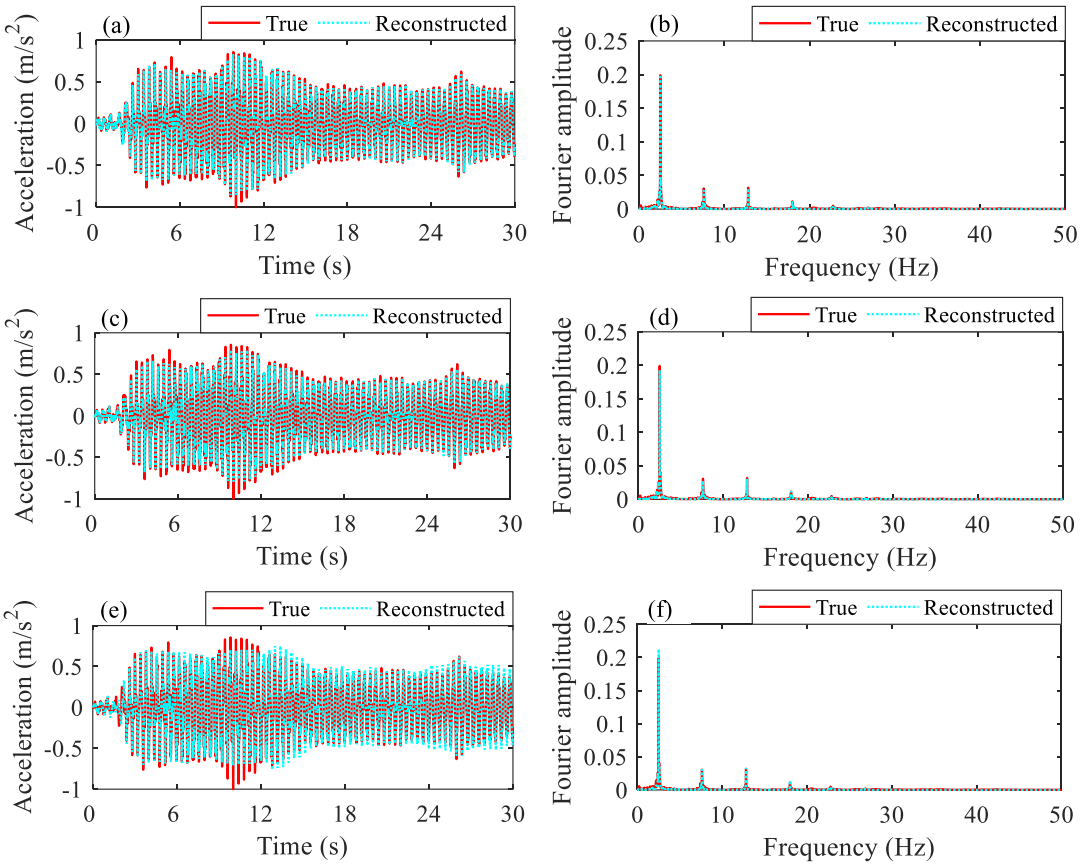


Figure 6-19. Comparison of true and reconstructed earthquake responses in time and frequency domains using: SegGAN (a) and (b); DenseNet (c) and (d); and CNN (e) and (f).

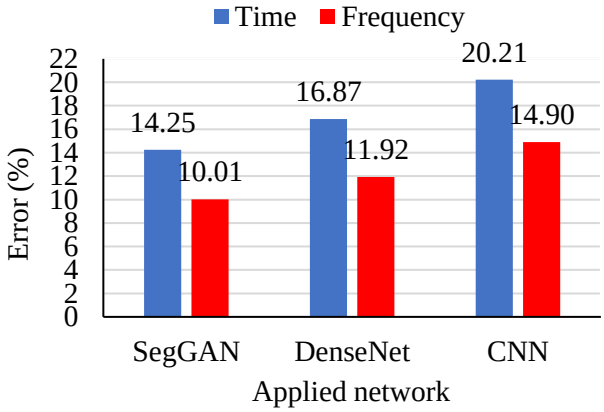


Figure 6-20. Reconstruction errors of experimental response under earthquake excitations in time and frequency domains.

6.6 Conclusions

This chapter proposes a specially designed SegGAN for reconstructing structural dynamic response under extreme loading conditions. Numerical studies on a steel frame structure and a nonlinear building structural model, as well as experimental studies on a laboratory tested steel frame structure are conducted to validate the effectiveness and accuracy of the proposed approach. The reconstruction results of using the proposed SegGAN are compared with those obtained by DenseNet and a traditional CNN. SegGAN featured with skip connection, dense connection and bottleneck architecture and trained by an adversarial strategy exhibits excellent feature extraction capability, even with limited sample datasets. With efficient and accurate feature extraction, SegGAN outperforms the other two networks and produces distinct dynamic response reconstruction results in both time and frequency domains. In numerical and experimental studies, SegGAN effectively and automatically extracts linear and nonlinear response features from limited training data with measurement noise under operational conditions, and can be used to reconstruct the responses under extreme loading condition, e.g., an earthquake. The advantages of using less measurement data under operational conditions and not necessarily including responses under extreme loadings for training make this developed approach more applicable for practical applications.

6.7 References

Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M., & Inman, D. J. (2017). Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration, 388*, 154-170.

Bao, Y., Li, H., Sun, X., Yu, Y., & Ou, J. (2013). Compressive sampling–based data loss recovery for wireless sensor networks used in civil structural health monitoring. *Structural Health Monitoring, 12*(1), 78-95.

Bao, Y., Shi, Z., Wang, X., & Li, H. (2018). Compressive sensing of wireless sensors based on group sparse optimization for structural health monitoring. *Structural Health Monitoring, 17*(4), 823-836.

Bao, Y., Tang, Z., & Li, H. (2020). Compressive-sensing data reconstruction for structural health monitoring: a machine-learning approach. *Structural Health Monitoring, 19*(1), 293-304.

Bao, Y., Tang, Z., Li, H., & Zhang, Y. (2019). Computer vision and deep learning–based data anomaly detection method for structural health monitoring. *Structural Health Monitoring, 18*(2), 401-421.

Cha, Y. J., Choi, W., Suh, G., Mahmoudkhani, S., & Büyüköztürk, O. (2018). Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. *Computer-Aided Civil and Infrastructure Engineering, 33*(9), 731-747.

Chen, W., Lu, Z., Lin, W., Chen, S., Ni, Y., Xia, Y., & Liao, W. (2011). Theoretical and experimental modal analysis of the Guangzhou New TV Tower. *Engineering Structures, 33*(12), 3628-3646.

Chen, Z., Bao, Y., Li, H., & Spencer Jr, B. F. (2018). A novel distribution regression approach for data loss compensation in structural health monitoring. *Structural Health Monitoring, 17*(6), 1473-1490.

Cross, E., Koo, K., Brownjohn, J., & Worden, K. (2013). Long-term monitoring and data analysis of the Tamar Bridge. *Mechanical Systems and Signal Processing, 35*(1-2), 16-34.

Fan, G., Li, J., & Hao, H. (2019). Lost data recovery for structural health monitoring based on convolutional neural networks. *Structural Control and Health Monitoring, 26*(10), e2433.

Gao, Y., & Mosalam, K. M. (2018). Deep transfer learning for image-based structural damage recognition. *Computer-Aided Civil and Infrastructure Engineering, 33*(9), 748-768.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). *Generative adversarial nets.* Paper presented at the Advances in neural information processing systems, Montréal, Canada.

He, J., Guan, X., & Liu, Y. (2012). Structural response reconstruction based on empirical mode decomposition in time domain. *Mechanical Systems and Signal Processing, 28*, 348-366.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). *Densely connected convolutional networks.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, USA.

Huang, Y.-W., Wu, D.-G., & Li, J. (2011). Structural healthy monitoring data recovery based on extreme learning machine. *Computer Engineering, 16*, 241-243.

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). *Image-to-image translation with conditional adversarial networks.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, USA.

Kammer, D. C. (1997). Estimation of structural response using remote sensor locations. *Journal of Guidance, Control, and Dynamics, 20*(3), 501-508.

Kuleshov, V., Enam, S. Z., & Ermon, S. (2017). Audio super resolution using neural networks. *https://arxiv.org/abs/1708.00853*.

Kullaa, J. (2010). Sensor validation using minimum mean square error estimation. *Mechanical Systems and Signal Processing, 24*(5), 1444-1457.

Law, S., Li, J., & Ding, Y. (2011). Structural response reconstruction with transmissibility concept in frequency domain. *Mechanical Systems and Signal Processing, 25*(3), 952-968.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., & Wang, Z. (2017). *Photo-realistic single image super-resolution using a generative adversarial network.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, USA.

Li, J., Hao, H., Fan, G., Ni, P., Wang, X., Wu, C., Lee, J.-M., & Jung, K.-H. (2017). Numerical and experimental verifications on damping identification with model updating and vibration monitoring data. *Smart Structures and Systems, 20*(2), 127-137.

Li, J., & Law, S. (2011). Substructural response reconstruction in wavelet domain. *Journal of Applied Mechanics, 78*(4), 041010.

Li, J., Law, S., & Ding, Y. (2012). Substructure damage identification based on response reconstruction in frequency domain and model updating. *Engineering Structures, 41*, 270-284.

Lin, Y. Z., Nie, Z. H., & Ma, H. W. (2017). Structural damage detection with automatic feature-extraction through deep learning. *Computer-Aided Civil and Infrastructure Engineering, 32*(12), 1025-1046.

Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). *Rectifier nonlinearities improve neural network acoustic models.* Paper presented at the International conference on machine learning, Atlanta, USA.

Mao, X., Shen, C., & Yang, Y. B. (2016). *Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections.* Paper presented at the Advances in neural information processing systems, Barcelona, Spain.

Nguyen, T., Chan, T. H. T., Thambiratnam, D. P., & King, L. (2015). Development of a cost-effective and flexible vibration DAQ system for long-term continuous structural health monitoring. *Mechanical Systems and Signal Processing, 64-65*, 313-324.

Odena, A., Dumoulin, V., & Olah, C. (2016). Deconvolution and checkerboard artifacts. *Distill, 1*(10), e3.

Oh, B. K., Glisic, B., Kim, Y., & Park, H. S. (2020). Convolutional neural network–based data recovery method for structural health monitoring. *Structural Health Monitoring*, 1475921719897571.

Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., & Efros, A. A. (2016). *Context encoders: feature learning by inpainting.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, USA.

Pathirage, C. S. N., Li, J., Li, L., Hao, H., Liu, W., & Ni, P. (2018). Structural damage identification based on autoencoder neural networks and deep learning. *Engineering Structures, 172*, 13-28.

Pathirage, C. S. N., Li, J., Li, L., Hao, H., Liu, W., & Wang, R. (2019). Development and application of a deep learning–based sparse autoencoder framework for structural damage identification. *Structural Health Monitoring, 18*(1), 103-122.

Peeters, B., & De Roeck, G. (2001). One-year monitoring of the Z24-Bridge: environmental effects versus damage events. *Earthquake engineering & structural dynamics, 30*(2), 149-171.

Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. *https://arxiv.org/abs/1511.06434*.

Ribeiro, A., Silva, J., & Maia, N. (2000). On the generalisation of the transmissibility concept. *Mechanical Systems and Signal Processing, 14*(1), 29-35.

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-net: convolutional networks for biomedical image segmentation.* Paper presented at the International conference on medical image computing and computer-assisted intervention, Munich, Germany.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., & Wang, Z. (2016). *Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network.* Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition, Las Vegas, USA.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research, 15*(1), 1929-1958.

Sun, S.-B., He, Y.-Y., Zhou, S.-D., & Yue, Z.-J. (2017). A data-driven response virtual sensor technique with partial vibration measurements using convolutional neural network. *Sensors, 17*(12), 2888.

Wan, H.-P., & Ni, Y.-Q. (2019). Bayesian multi-task learning methodology for reconstruction of structural health monitoring data. *Structural Health Monitoring, 18*(4), 1282-1309.

Wan, Z., Li, S., Huang, Q., & Wang, T. (2014). Structural response reconstruction based on the modal superposition method in the presence of closely spaced modes. *Mechanical Systems and Signal Processing, 42*(1-2), 14-30.

Wang, Z.-C., Xin, Y., & Ren, W.-X. (2015). Nonlinear structural model updating based on instantaneous frequencies and amplitudes of the decomposed dynamic responses. *Engineering Structures, 100*, 189-200.

Xia, Y., Chen, B., Zhou, X., & Xu, Y. (2013). Field monitoring and numerical analysis of Tsing Ma Suspension Bridge temperature behavior. *Structural Control and Health Monitoring, 20*(4), 560-575.

Xu, Y., Wei, S., Bao, Y., & Li, H. (2019). Automatic seismic damage identification of reinforced concrete columns from images by a region-based deep convolutional neural network. *Structural Control and Health Monitoring, 26*(3), e2313.

Xu, Y., Zhang, X., Zhu, S., & Zhan, S. (2016). Multi-type sensor placement and response reconstruction for structural health monitoring of long-span suspension bridges. *Science bulletin, 61*(4), 313-329.

Yeh, R. A., Chen, C., Lim, T.-Y., Schwing, A. G., Hasegawa-Johnson, M., & Do, M. N. (2017). *Semantic image inpainting with deep generative models.* Paper presented at the The IEEE conference on computer vision and pattern recognition, Honolulu, USA.

Zhang, C. D., & Xu, Y. (2016). Optimal multi-type sensor placement for response and excitation reconstruction. *Journal of Sound and Vibration, 360*, 112-128.

Zhang, D., Bao, Y., Li, H., & Ou, J. (2012). Investigation of temperature effects on modal parameters of the China National Aquatics Center. *Advances in Structural Engineering, 15*(7), 1139-1153.

Zhang, R., Chen, Z., Chen, S., Zheng, J., Büyüköztürk, O., & Sun, H. (2019). Deep long short-term memory networks for nonlinear structural seismic response prediction. *Computers & Structures, 220*, 55-68.

Zhang, T., Biswal, S., & Wang, Y. (2019). SHMnet: condition assessment of bolted connection with beyond human-level performance. *Structural Health Monitoring*, 1475921719881237.

Zhang, Y., Miyamori, Y., Mikami, S., & Saito, T. (2019). Vibration-based structural state identification by a 1-dimensional convolutional neural network. *Computer-Aided Civil and Infrastructure Engineering, 34*(9), 822-839.

Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018). *Residual dense network for image super-resolution.* Paper presented at the Proceedings of the IEEEconference on computer vision and pattern recognition, Salt Lake City, USA.

Zhang, Z., & Luo, Y. (2017). Restoring method for missing data of spatial structural stress monitoring based on correlation. *Mechanical Systems and Signal Processing, 91*, 266-277.

Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). *Unpaired image-to-image translation using cycle-consistent adversarial networks.* Paper presented at the Proceedings of the IEEE international conference on computer vision, Venice, Italy.

CHAPTER 7 CONCLUSIONS AND RECOMMENDATIONS

7.1 Main findings

In this thesis, five methods for addressing different technical issues of SHM is proposed for improving the applicability of SHM techniques for civil engineering structures. The major contributions and findings in this thesis are briefly summarized below:

Chapter 2 proposes an improved automated OMA approach based on data-driven SSI and clustering techniques. The results in numerical and experimental studies reveal that the proposed approach is capable of identifying the vibration modes accurately even the data are polluted with high level of noise. The proposed approach outperforms the reference method because the statistics-based clustering and clustering merging procedures provide more comprehensive considerations of modal parameters that can distinguish the physical and spurious modes more effectively. The proposed method is fully automated and efficient, which can be used for processing long-term vibration data.

Chapter 3 proposes an innovative approach based on specially designed CNN for recovering the lost SHM data during the transmission process. The results of two case studies show that the proposed CNN can effectively recover vibration data in both time and frequency domains. A comparison study of the effect of sampling rates demonstrates that the proposed approach can accurately recover signals with a low sampling rate, as long as the excited vibration modes are covered in the target frequency range of interest as half of the sampling rate. The data lost recovery results and modal identification accuracy using the recovered signals reveal that the proposed CNN has a decent performance to effectively recover the lost data in SHM, even with a 90% data loss ratio.

Chapter 4 proposes a ResNet based vibration signal denoising approach to remove noise in vibration measurement data. Experimental studies on Guangzhou New TV Tower

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 in 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. 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.

Chapter 5 proposes a novel approach based on DenseNets for reconstructing the acceleration responses at locations without recorded data, by using recorded responses at other locations. The developed DenseNets can alleviate the vanishing-gradient issue, strengthen feature extraction and propagation, and substantially reduce the number of parameters, which significantly increases the training efficiency and accuracy. The results of the experimental studies demonstrate that the proposed method can extract the high-level features from the input data and develop the nonlinear relationship between these features and the output responses to be reconstructed. The proposed method is also proved to have a strong noise immunity, which is practical for civil engineering applications.

Chapter 6 proposes a specially designed SegGAN for reconstructing linear and nonlinear response under extreme loadings. The reconstruction results of SegGAN are compared with those generated by DenseNet and a traditional CNN. SegGAN featured with skip connection, dense connection and bottleneck architecture and trained by an adversarial strategy exhibits overwhelming feature extraction capability. With efficient and accurate feature extraction, SegGAN outperforms the other two networks and produces distinguished reconstruction results in both time and frequency domains. SegGAN effectively and automatically extracts linear and nonlinear response features from limited training data with measurement noise and varying environmental and operational conditions. The advantages of less data requirement, complex environmental adaptability and automation lead this method to more applicable for practical applications.

7.2 Recommendations for future works

In the current research, clustering and deep learning techniques have shown strong capability on addressing SHM issues. Further study may leverage these advantages and further extend the researches and applications in the following areas:

- The identified long-term modal parameters by using the automated modal identification method involve the long-term vibration characteristics of structures. They can be utilized as the input to networks for a more robust and automated structural condition assessment;
- For signal denoising, 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;
- In response reconstruction by DenseNet, selection of input channels are manually conducted, and the computation would be time-consuming when the number of available sensors is large. Therefore selecting optimal input channels can be a future research topic;
- Response reconstruction by networks using inter-channel correlations is similar to response transmission. Further study may be conducted to estimate the structural transmissibility function by using deep learning networks;
- SegGAN with strong capability on nonlinear feature extraction has potential to be extended to nonlinear system identification.

APPENDIX I

THE PSEUDOCODE OF THE STATISTICS-BASED CLUSTERING

The pseudocode of the statistics-based clustering

Generate system modes matrix *S* from SSI

Remove spurious modes by hard validation criteria

Randomize the row order of the cleaned system modes matrix *S*

Initialize the cluster number *k* as 1

For *ii = 1* to n_{total} (scan the system mode matrix from the first to the last model order)

- **For** *jj* = 1 to the last mode (scan modes from the first to the last mode in each model order)
  - **If** mode at *S[ii,jj]* is ungrouped
    - Create a new cluster C_k
    - Save this mode as the first element of this cluster
    - Initialize the validation criteria $T_f, T_d, T_{MAC}, T_{MPC}$ as predefined values
    - **For** *i* = (*ii* +1) to n_{total} (scan from the successive model order)
      - **For** *j* = 1 to the last element
        - **If** mode at *S[i,j]* is ungrouped
          - Compute the distance of the mode with cluster C_k by Equation (9)
        - **End If**
      - **End For**
      - Find the mode *m* with the minimum distance to C_k
      - **If** The modal parameters of mode *m* satisfy all the thresholds in Equation 10
        - Collect *m* to cluster C_k
        - **If** the number of elements in $C_k \geq 10\% n_{total}$
          - Compute the difference vectors by Equation (11)
          - Update the validation criteria by Equation (12)
        - **End If**
      - **End If**
    - **End For**
    - Finish the clustering of C_k and update *k* as *k = k+1*
  - **End If**
- **End For**

End For

Eliminate the clusters containing elements < $10\% n_{total}$

APPENDIX II

ATTRIBUTION OF AUTHORSHIP

To whom it may concern

I, Gao Fan, conducted numerical, experimental investigations, data processing, analysis and wrote manuscripts of the papers titled as follows, which were revised and edited by the second and the third co-authors. They also provided insights on experimental preparation, data processing and data analysis.

1. **Improved automated operational modal identification of structures based on clustering**
2. **Lost data recovery for structural health monitoring based on convolutional neural networks**
3. **Vibration signal denoising for structural health monitoring by residual convolutional neural networks**
4. **Dynamic response reconstruction for structural health monitoring using densely connected convolutional networks**

(……...……...)

I, as a co-author, endorse that this level of contribution by the candidate indicated above is appropriate.

(Prof. Hong Hao) (……...……...)

(Associate Prof. Jun Li) (……...……...)

To whom it may concern

I, Gao Fan, conducted numerical, experimental investigations, data processing, analysis and wrote manuscripts of the paper titled as follows, which were revised and edited by the second and the third co-authors. They also provided insights on experimental preparation, data processing and data analysis. The last author helps design the nonlinear model in the numerical studies.

Dynamic response reconstruction under extreme loading conditions using segment based generative adversarial network

(……...……...)

I, as a co-author, endorse that this level of contribution by the candidate indicated above is appropriate.

(Prof. Hong Hao) (……...……...)

(Associate Prof. Jun Li) (……...……...)

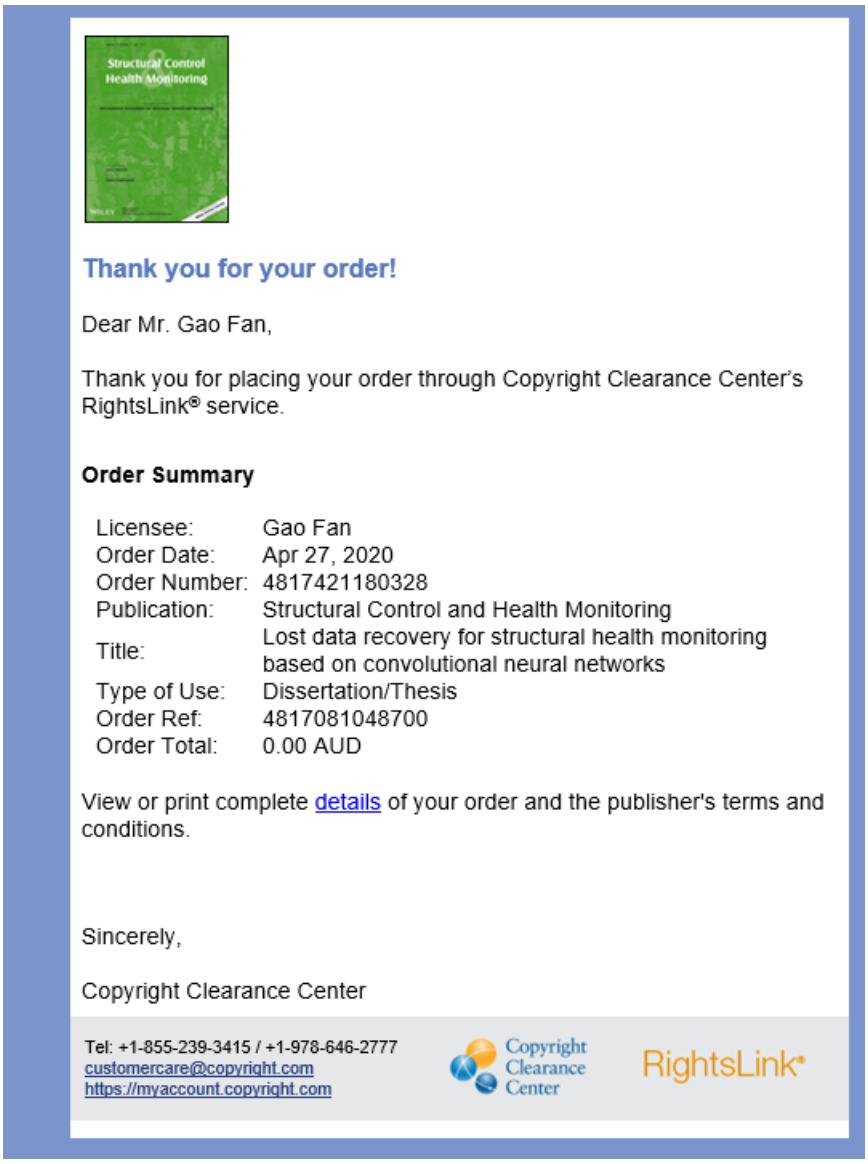
(Mr. Yu Xin) (……...……...)

APPENDIX III

COPYRIGHT CLEARANCE

The proof of the rights, granted by publishers for the publication that forms the chapters of this thesis, to reproduce the contribution in the thesis are attached below.

Fan G., Li J., Hao H. (2019). Improved automated operational modal identification of structures based on clustering. *Structural Control and Health Monitoring*. *26*(12), e2450. https://doi.org/10.1002/stc.2450.



Thank you for your order!

Dear Mr. Gao Fan,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Gao Fan
Order Date:	Apr 27, 2020
Order Number:	4817421180328
Publication:	Structural Control and Health Monitoring
Title:	Lost data recovery for structural health monitoring based on convolutional neural networks
Type of Use:	Dissertation/Thesis
Order Ref:	4817081048700
Order Total:	0.00 AUD

View or print complete details of your order and the publisher's terms and conditions.

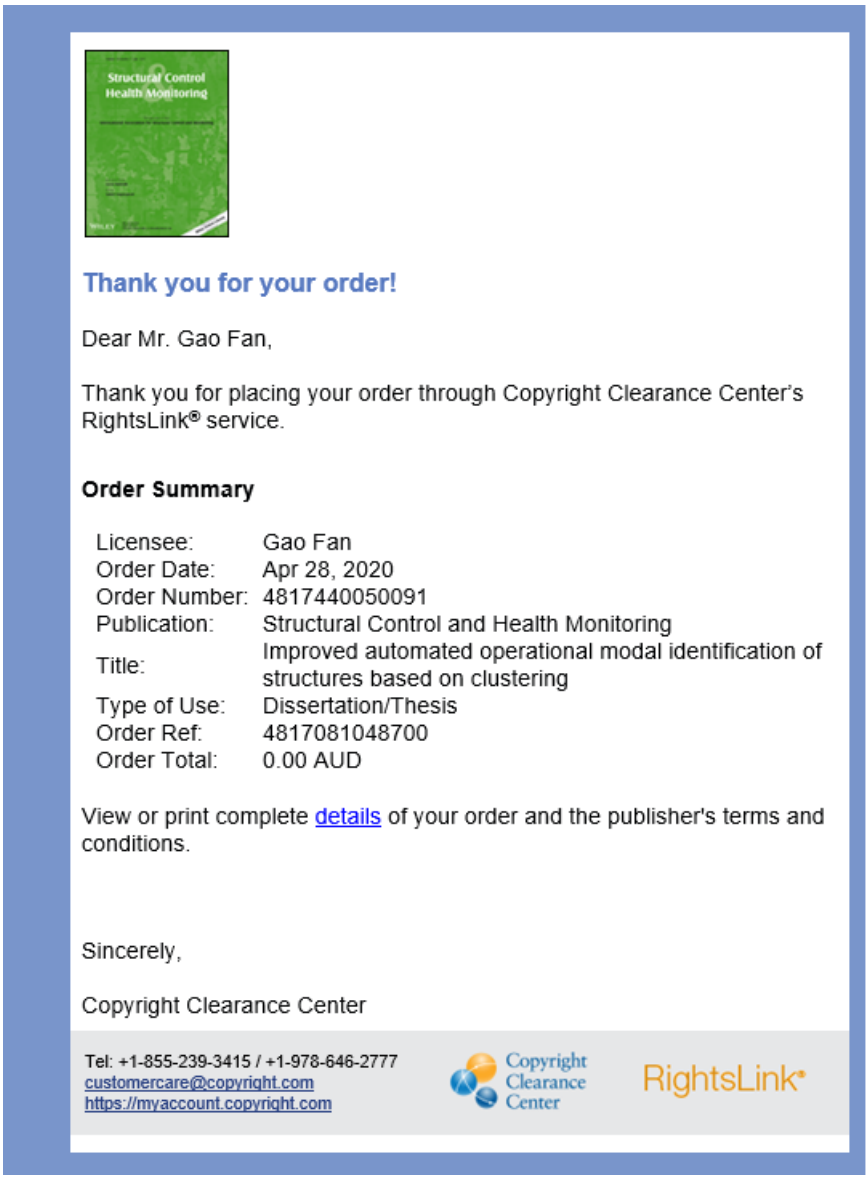
Sincerely,

Copyright Clearance Center

Tel: +1-855-239-3415 / +1-978-646-2777
customercare@copyright.com
https://myaccount.copyright.com

Copyright Clearance Center RightsLink®

Fan G., Li J., Hao H. (2019). Lost data recovery for structural health monitoring based on convolutional neural networks. *Structural Control and Health Monitoring*. *26*(10): e2433. https://doi.org/10.1002/stc.2433.



Thank you for your order!

Dear Mr. Gao Fan,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Licensee:	Gao Fan
Order Date:	Apr 28, 2020
Order Number:	4817440050091
Publication:	Structural Control and Health Monitoring
Title:	Improved automated operational modal identification of structures based on clustering
Type of Use:	Dissertation/Thesis
Order Ref:	4817081048700
Order Total:	0.00 AUD

View or print complete details of your order and the publisher's terms and conditions.

Sincerely,

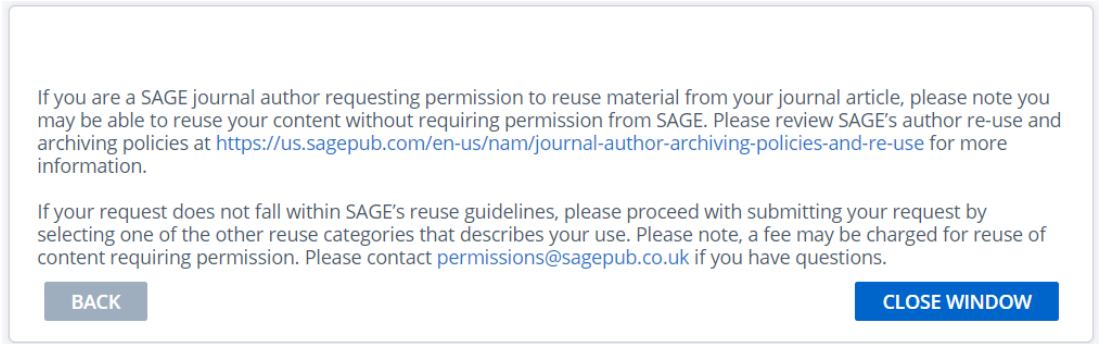
Copyright Clearance Center

Tel: +1-855-239-3415 / +1-978-646-2777
customercare@copyright.com
https://myaccount.copyright.com

Copyright Clearance Center RightsLink®

Fan G., Li J., Hao H.（2020）. Vibration signal denoising for structural health monitoring by residual convolutional neural networks. *Measurement*. 107651. https://doi.org/10.1016/j.measurement.2020.107651.

Fan G., Li J., Hao H. (online). Dynamic Response Reconstruction for Structural Health Monitoring Using Densely Connected Convolutional Networks. *Structural Health Monitoring*. https://doi.org/10.1177/1475921720916881



BIBLIOGRAPHY DISCLAIMER

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.