Convolutional Neural Network Based Damage Detection of IASC-ASCE Benchmark by Encoding Time-series into Images

Chen Shi[1], Younes Aoues[1], Renata Troian[1], Didier Lemosse[1]

[1] *Normandie Univ, INSA Rouen Normandie, Laboratoire de Mécanique de Normandie (LMN), 685 avenue de l'Université, 76800 Saint-Étienne-du-Rouvray, France,* chen.shi@insa-rouen.fr

ABSTRACT Civil engineering structures are essential to the infrastructure and directly impact people's lives and safety. Notably, many structures are in worrisome working conditions, with damage deterioration and occasionally collapse. Combining deep learning with structural health monitoring can provide unprecedented tools for structural damage detection and identification. This paper explores the use of acceleration features to predict the damage state of a structure using time-series acceleration data collected from the Phase I IASC-ASCE Structural Health Monitoring (SHM) benchmark model, where the primary identified damage is the removal of diagonal braces and beam connections. In this study, to overcome the limitations of using neural networks hampered by the small amount of data collected in shaker tests, three methods were used to encode time-series acceleration data as images, i.e., two-dimensional numerical matrix (2D Matrix), Short-Time Fourier Transform (STFT), and Wavelet Transform (WT). The encoded images were then used as input to the Convolutional Neural Networks (CNN) model. Two training/test scenarios are used to show the efficiency of the proposed method of structural damage recognition. The results show that the WT combined with the CNN is the best approach regarding accuracy and efficiency.

Keywords Convolutional Neural Networks (CNN), Structural Health Monitoring (SHM), Short-Time Fourier Transform (STFT)，Wavelet Transform (WT), structural damage detection

I. INTRODUCTION

Structural damage detection and Structural Health Monitoring (SHM) of infrastructure have always been an issue that cannot be ignored because they are closely related to our lives. Many monitoring and detection methods have been developed to provide practical early warning of structural damage or abnormalities. Considerable effort has been invested in vibration-based methods, which use the vibration response of the monitored structure to assess its condition and identify structural damage. Meanwhile, Machine Learning (ML) and especially Deep Learning (DL) algorithms have become more feasible with the development of emerging computing power and sensing technologies in the past decade. They are widely used for vibration-based structural damage detection with excellent performance and often with stringent accuracy.

Convolutional Neural Networks (CNN) have contributed much to damage detection research in computer vision. One of the research directions is the analysis of surface images of structures directly, as in Sony et al. (2021), Cha et al. (2018), Li et al. (2018), and Wang et al. (2018). However, some non-visible structural damage may also seriously affect the daily operation of the structure as well as its remaining service life. Moreover, the images required in the literature on vision-based methods for damage identification are often available at specific angles and distances, which makes the collection of structural surface images difficult. To further monitor structural damage that is not visible to the naked eye and break through the limitations of visual-based damage, researchers have turned their attention to data collected by sensors placed on structures, as in Chun et al. (2015), Abdeljaber et al. (2017), and Khodabandehlou et al. (2019). In such methods, the data are generally stored as time histories, such as acceleration data. In addition to directly processing acceleration data, many researchers preprocess sensor data by dynamic recognition techniques to extract damage features for subsequent analysis, as in Mantawy et al. (2022), Azimi et al (2020), and Chong et al. (2014).

However, these data conversion methods cannot achieve high accuracy and few computations simultaneously when facing large volumes of data. To overcome the drawbacks of methods utilizing time-series data in training machine learning models, especially the small data size, the paper utilizes various techniques to encode time-series data into images to feed the input layer of CNN models, which are very effective tools in computer vision.

METHODOLOGY

In this study, our approach is validated on the IASC-ASCE benchmark problem. Next, we describe our methodology, which consists of introduction of the benchmark structure, the framework for handling the structural damage detection problem, the technique for encoding time series data into images, and the CNN-based SHM.

A. IASC-ASCE benchmark

The IASC-ASCE benchmark problem was introduced by Johnson et al. (2004). Figure 1 shows the geometry of this benchmark structure. It is a four-story, two-bay by two-bay (grade 300W) steel-frame scale model structure with a footprint of 2.5 m × 2.5 m and a frame height of 3.6 m.

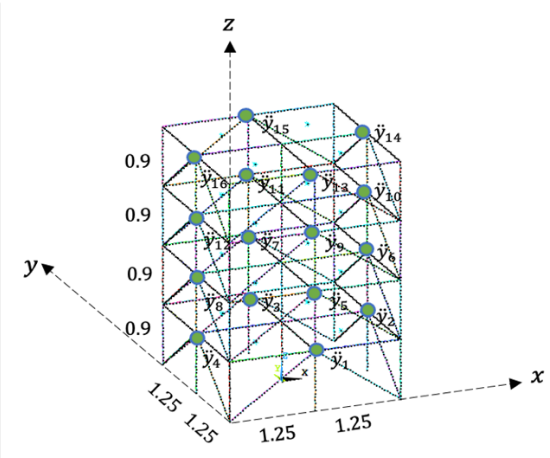


FIGURE 1. Numerical model of the ASCE SHM Benchmark structure

The sections are specifically designed for this scale model: the section type of columns is B100 × 9; the type of floor beams is S75 × 11; the support system consists of two 12.7 mm diameter threaded steel rods placed diagonally. The damage introduced in the benchmark was obtained by removing diagonal bracing at specific locations or loosening bolts at several connection locations. In this work, we have used the acceleration signal of 16 nodes of the structure, marked with green points. These measurement points are located at the middle of each side of the structure, two in the x and two in the y directions per floor.

B. *Damage detection procedure*

This section summarizes a structural damage detection procedure. The proposed pipeline can automatically learn fault characteristics and identify the structure's health from the raw vibration signals. Figure 2 shows the overall procedure of the proposed structural damage detection pipeline.

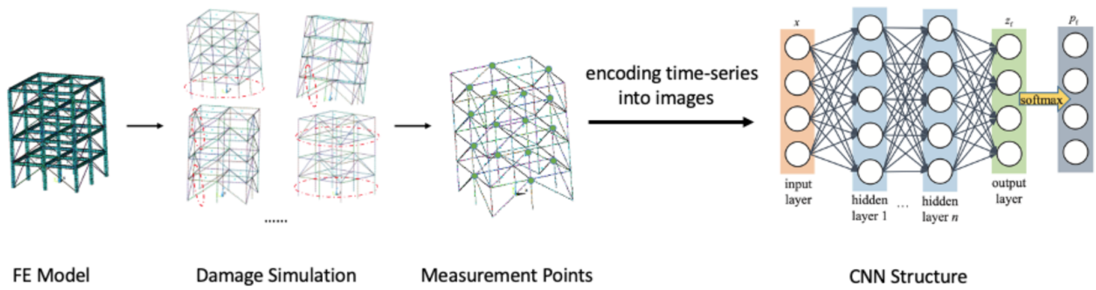


FIGURE 2. General pipeline of structural damage detection

Vibration signals with different damage modes are collected at fixed measurement points on different finite element models of the IASC-ASCE reference structure. After that, the raw vibration signals are converted into images as input to the CNN. The different techniques used in this paper for converting time series acceleration data into images and the CNN used are further described in the following sections. The main objective of this paper is to compare the different techniques for converting time series acceleration data into images when combined with CNN for structural damage detection.

C. *Encoding time-series into images*

The type and amount of recorded data play an important role in structural SHM. In most cases, the data are recorded in temporal order, i.e., before the damage occurs (no damage state), during different degrees of damage, and after the damage occurs. However, structural damage features are not easily recognized by CNN in the one-dimensional time-domain signal. One feasible consideration is to slice the one-dimensional data and convert it into a two-dimensional numerical matrix. Another approach is to convert the original data from the time domain to the frequency domain and then use it as the input image for CNN. As shown in Figure 3, the original vibration signal (Acceleration data over time) is truncated by a sliding window by defining the "Frame" length as $128 * 128$. This operation allows the generation of a set of sub-time series, then we preprocess frames as feature maps by transforming to two-dimensional numerical matrix (2D Matrix).

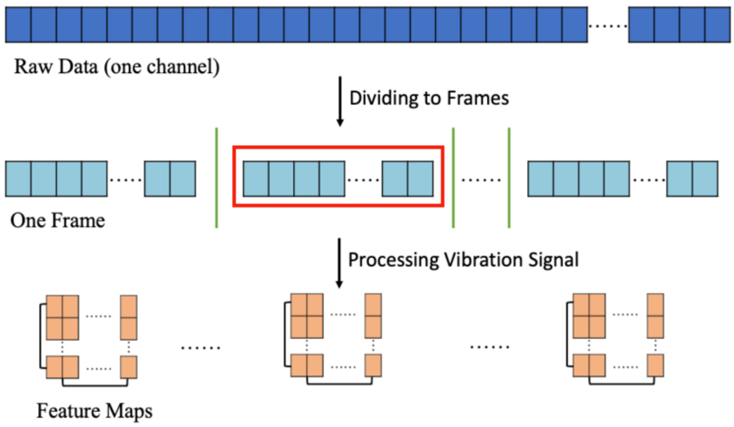


FIGURE 3. Schematic diagram of processing data into two-dimensional numerical matrices

We also apply Short-Time Fourier Transform (STFT) and Wavelet Transform (WT) to these generated sub-samples. Analyzing signals in the time domain is a common and effective way of modern signal processing. Fourier transform is a powerful tool, and it can reflect the overall spectral characteristics of the signal well. However, the Fourier transform can only obtain information about the frequency components of the signal to be processed, but not time information of each frequency component, which results in two signals with very different time domains may have the same spectrum. For non-smooth signals such as structural fault vibration signals, the signal frequency changes with time, and the instantaneous frequency and amplitude information at each moment are also important, which will increase the success rate of damage identification. By selecting the wavelet basis function $\psi_{a,b}(t)$ and adjusting the scaling factor a and translation factor b, wavelet transform with different time-frequency widths can be obtained, which can be matched with the original signal at any time to complete the time-frequency localized two-dimensional analysis of the signal, as shown in Figure 4.

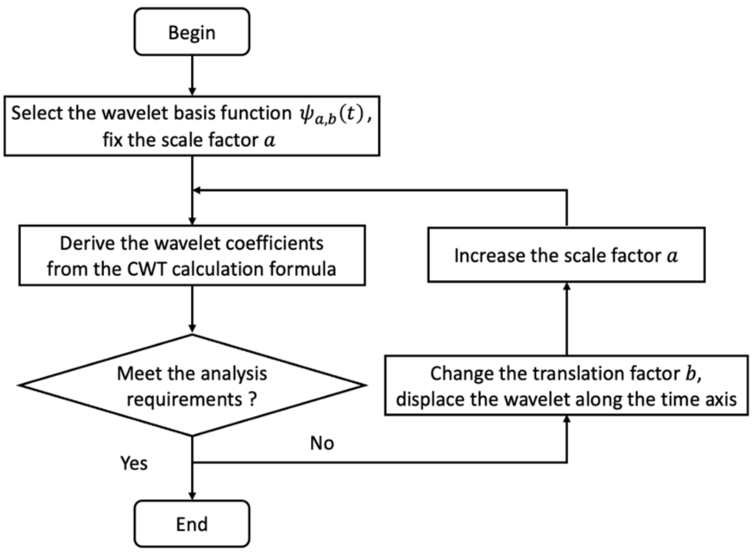


FIGURE 4. Numerical Basic flow of Wavelet Transform

D. Proposed CNN-Based SHM

CNNs are feedforward artificial neural networks inspired by the visual cortex of animals. CNNs have been widely used in object detection in images, videos, and the classification of images (Li 2021). As shown in Figure 5, each CNN consists of five kinds of layers: the input layer, the convolution layer, the pooling layer, the fully connected layer, and the output layer. The input layer is a passive layer that accepts the input image and passes the input to convolutional layers. The convolutional layers apply different filters to the input image to extract characteristic features. These filters are not fixed and are trained by a backpropagation algorithm. The subsampling (pooling) layers reduce the dimension of the convolutional layer output following each convolutional layer. The fully connected layers are similar to the hidden layers of a multilayer perceptron and together with the output layer, produce the output vector of the CNN.

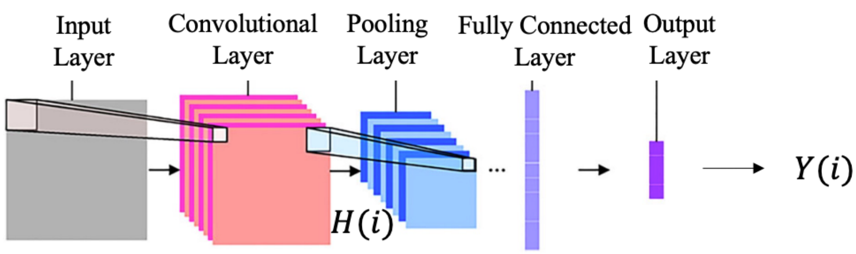


FIGURE 5. General CNN structure

The process of a convolutional neural network is to transform the original matrix H_0 through multiple levels of data become H_i , and map it to a new mathematical model of feature representation Y_i. In this study, the acceleration data obtained from the different damage patterns of the IASC-ASCE benchmark structure are encoded as input "images". The training data fed into the CNN corresponding to 9 damage patterns. For this purpose, a CNN with four convolutional layers (kernel size = 3), four max pooling layers (kernel size = 2), one fully connected layer (number of neurons = 9), and a SoftMax layer as the output layer is adopted as show in Figure 6. We adopt Negative Log-Likelihood Loss (NLLLoss) as our loss function, Stochastic Gradient Descent (SGD) as our optimization way. We use a learning rate scheduler to adapt the learning rate during the training process with an initial value of 0.1, it will be reduced by a factor of 0.1 every 10 epochs. In addition, to prevent the model from overfitting the training data, we apply an early stopping strategy: stop training when the model's performance on the training set changes within 0.001 for ten consecutive times instead of continuing training until the training loss is minimized.

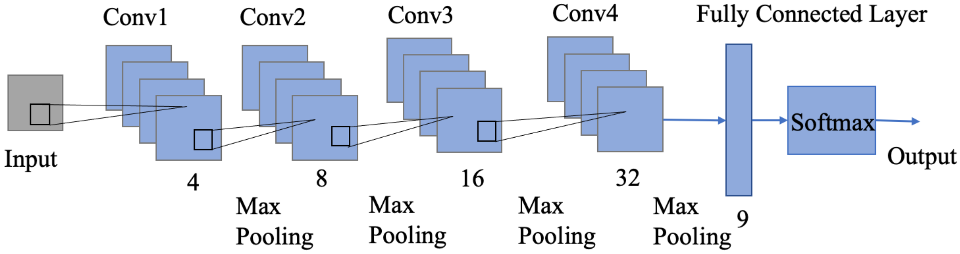


FIGURE 6. The proposed CNN structure

III. DISCUSSION & RESULTS

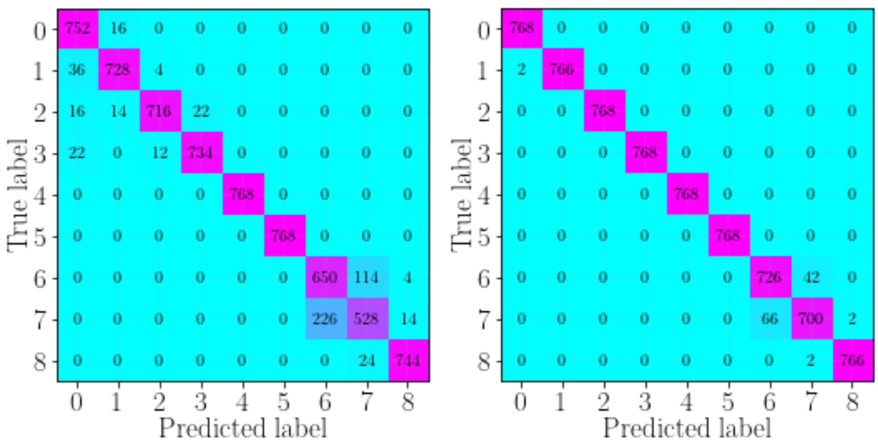
A MATLAB-based finite element analysis code is developed by the IASC-ASCE SHM Task Group to generate the simulated structural response data (Johnson 2004). In this study, we use nine different damage patterns (including the undamaged pattern) of the IASC-ASCE benchmark structure proposed in Abdeljaber et al. (2018) as the training and testing dataset for the proposed CNN. In order to overcome the limitations of using neural networks hampered by the small amount of data collected in shaker tests, three methods were used to encode time-series acceleration data as images, i.e., 2D Matrix, STFT, and WT. The sampling frequency of raw data is 1000Hz. The time-length of SFTF is $128 * 128/1000 = 16.384$s, the frequency-length of STFT is 50Hz (discard useless information in high frequency). Therefore, for one channel of data (400s), we can get $40,000/(128 * 128)/2 \approx 48$ spectrograms; for each damage pattern i, we have 16 channels, so we can get $16 * 48 = 768$ spectrograms; for all damage patterns, we can get $9 * 16 * 48 = 6912$ spectrograms. We use Daubechies 5 (bd5) wavelet with its own scaling and wavelet function. The encoded images were then used as input to the CNN model. The damage prediction results of the CNN using three different methods to encode the acceleration data into images are shown in Table 1. "Length of Samples" is the size of the vibration signal in seconds. "Model" represents the technique used to convert the time series into images. "Accuracy" measures the proportion of correct predictions out of the total number of predictions made. We will further explain the accuracy in the form of the confusion matrix, which shows the number of correct and incorrect predictions made by the classifier, broken down by each class. "Training time" means the time taken to run the code. We conduct two sets of experiments with data lengths of 40s and 400s, and each experiment calculates the prediction accuracy and training time for different models (2D Matrix, STFT, WT). When the duration of the acceleration data to be processed increases from the 40s to 400s, the prediction accuracy increases from 68% to 75% using the 2D Matrix model, and the training time increases from 20.34s to 3min53s; the prediction accuracy increases from 73% to 92% using the STFT model, and the training time increases from 3.82s to 39.11s; the prediction accuracy using the WT model increases from 89% to 98%, the training time increases from 3.30s to 35.03s. The results show that the WT combined with the CNN is optimal in accuracy and efficiency.

TABLE 1. Comparison of damage prediction results for benchmark structure

Length of Samples	Model	Accuracy	Training time
40s	2D Matrix	68%	20.34s
	STFT	73%	3.82s
	WT	**89%**	**3.30s**
400s	2D Matrix	75%	3min53s
	STFT	92%	39.11s
	WT	**98%**	**35.03s**

To better describe the performance of STFT and WT combined with the CNN on structural damage prediction, we use their corresponding confusion matrices for further analysis. It is a visualization of the relationship between actual and predicted values in a classification problem. The confusion matrix of all training dataset of CNN predicted damage pattern result (data length: 400s, data preprocessing: STFT) is shown in Figure 7(a). For ease of expression, we abbreviate the damage pattern as DP. We can find that the prediction accuracy of DP 0 (i.e., no damage to the structure) is

98%, DP 1 is 95%, DP 2 is 93%, and DP 3 is 96%, with a small amount of mutual confusion occurring between their four DPs. The prediction accuracy of DPs 4 and 5 is 100%. DP 6 has a prediction accuracy of 85%, and DP 7 has a prediction accuracy of 69%, much lower than the other impairment modes. DPs 6 and 7 are relatively easy to confuse for the CNN, with the number of each damage pattern being 768, with 226 cases of DP 6 being mistaken for 7, and 114 cases of DP 7 being mistaken for 6. The prediction accuracy of DP 8 is 97%. In summary, the final training time is 39.11s, and the overall accuracy is 92%. The confusion matrix of all training dataset of CNN predicted fault mode result (data length: 400s, data preprocessing: WT) is shown in Figure 7(b). We can find that the prediction accuracy of DPs 0, 1, 3, 4, 5, and 8 is 100%. The prediction accuracy of DP 6 is 95%, and the prediction accuracy of DP 7 improves from 69% to 91% when comparing STFT and WT to data processing. In summary, the final training time is 35.03s, and the overall accuracy is 98%.



(a) (b)

FIGURE 7. Confusion matrix of the trained CNN prediction failure mode results for 400s length of samples with different preprocess models, (a): STFT, (b): WT.

IV. CONCLUSIONS

This paper compares the effectiveness of three methods to encode time-series acceleration data into images when combined with CNN for structural damage identification. When the duration of the acceleration data to be processed increases from 40s to 400s, the prediction accuracy using the WT method increases from 89% to 98%, the training time increases from 3.30s to 35.03s. The results show that the WT combined with the CNN method is optimal in accuracy and efficiency. Therefore, when collecting data is expensive, pre-processing data with WT should be preferred. In addition, the CNN architecture used in this paper is simple and less likely to lead to overfitting. Moreover, the method used in this paper encodes the time series data as images, i.e., it takes into account that the loading period may also be one of the causes of the damage and does not ignore the time effect of the time series, which provides the possibility to apply the proposed method to fatigue damage of structures.

REFERENCES

Mantawy, I.M., & Mantawy, M.O. (2022). Convolutional neural network based structural health monitoring for rocking bridge system by encoding time-series into images. *Structural Control and Health Monitoring*, 29(3), e2897. https://doi.org/10.1002/stc.2897

Sony, S., Dunphy, K., Sadhu, A., & Capretz, M. (2021). A systematic review of convolutional neural network-based structural condition assessment techniques. *Engineering Structures*, 226, 111347. https://doi.org/10.1016/j.engstruct.2020.111347

Johnson, E.A., Lam, H.F., Katafygiotis, L.S., & Beck, J.L. (2004). Phase I IASC-ASCE structural health monitoring benchmark problem using simulated data. *Journal of engineering mechanics*, 130(1), 3-15. https://doi.org/10.1061/(ASCE)0733-9399(2004)130:1(3)

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems.* https://ieeexplore.ieee.org/abstract/document/9451544

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. https://onlinelibrary.wiley.com/doi/full/10.1111/mice.12334

Li, S., Zhao, X., & Zhou, G. (2019). Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network. *Computer-Aided Civil and Infrastructure Engineering*, 34(7), 616-634. https://onlinelibrary.wiley.com/doi/full/10.1111/mice.12433

Wang, N., Zhao, Q., Li, S., Zhao, X., & Zhao, P. (2018). Damage classification for masonry historic structures using convolutional neural networks based on still images. *Computer-Aided Civil and Infrastructure Engineering*, 33(12), 1073-1089. https://onlinelibrary.wiley.com/doi/full/10.1111/mice.12411

Chun, P.-J., Yamashita, H., & Furukawa, S. (2015). Bridge damage severity quantification using multipoint acceleration measurement and artificial neural networks. *Shock and Vibration*, 2015, https://doi.org/10.1155/2015/789384

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. https://www.sciencedirect.com/science/article/pii/S0022460X16306204

Azimi, M., & Pekcan, G. (2020). Structural health monitoring using extremely compressed data through deep learning. *Computer-Aided Civil and Infrastructure Engineering*, 35(6), 597–614. https://onlinelibrary.wiley.com/doi/full/10.1111/mice.12517

Chong, J.W., Kim, Y., & Chon, K.H. (2014). Nonlinear multiclass support vector machine–based health monitoring system for buildings employing magnetorheological dampers. *Journal of Intelligent Material Systems and Structures*, 25(12), 1456–1468.