Accepted Manuscript

Response prediction of nonlinear hysteretic systems by deep neural networks

Taeyong Kim, Oh-Sung Kwon, Junho Song

PII: S0893-6080(18)30336-8
DOI: https://doi.org/10.1016/j.neunet.2018.12.005
Reference: NN 4076

To appear in: *Neural Networks*

Received date : 18 June 2018
Revised date : 4 December 2018
Accepted date : 7 December 2018

Please cite this article as: T. Kim, O.-S. Kwon and J. Song, Response prediction of nonlinear hysteretic systems by deep neural networks. *Neural Networks* (2018), https://doi.org/10.1016/j.neunet.2018.12.005

*Title Page (With all author details listed)

Submitted for Publication in *Neural Networks*, June 18, 2018
Revised Manuscript Submitted, August 27, 2018, Second Revision Submitted. October 6, 2018,
Third Revision Submitted, December 4, 2018

Response Prediction of Nonlinear Hysteretic Systems by Deep Neural Networks

Taeyong Kim[1], Oh-Sung Kwon[2], Junho Song[3]

Corresponding author: Oh-Sung Kwon, Department of Civil and Mineral Engineering, University of Toronto, Toronto, Canada. (os.kwon@utoronto.ca)

[1]Graduate Student, Department of Civil and Environmental Engineering,

Seoul National University, Seoul, Korea

[2]Associate Professor, Ph.D., Department of Civil and Mineral Engineering

University of Toronto, Toronto, Canada

[3]Professor, Ph.D., Department of Civil and Environmental Engineering,

Seoul National University, Seoul, Korea

Submitted for Publication in *Neural Networks*, June 18, 2018

Revised Manuscript Submitted, August 27, 2018, Second Revision Submitted, October 6, 2018,

Third Revision Submitted, December 4, 2018

Response Prediction of Nonlinear Hysteretic Systems by Deep Neural Networks

Taeyong Kim[a], Oh-Sung Kwon[b*], Junho Song

[a]*Department of Civil and Environmental Engineering, Seoul National University, Seoul, Korea.*
[b]*Department of Civil and Mineral Engineering, University of Toronto, Toronto, Canada*

Abstract: Nonlinear hysteretic systems are common in many engineering problems. The maximum response estimation of a nonlinear hysteretic system under stochastic excitations is an important task for designing and maintaining such systems. Although a nonlinear time history analysis is the most rigorous method to accurately estimate the responses in many situations, high computational costs and modelling time hamper adoption of the approach in a routine engineering practice. Thus, in an engineering practice, various simplified regression equations are introduced to replace a nonlinear time history analysis, but the accuracy of the estimated responses is limited. This paper proposes a deep neural network trained by the results of nonlinear time history analyses as an alternative of such simplified regression equations. To this end, the convolutional neural network (CNN) which is usually applied to abstract features from visual imagery is introduced to analyze the information of the hysteretic behavior of the system, then merged with neural networks representing a stochastic random excitation to predict the responses of a nonlinear hysteretic system. For verification, the proposed deep neural network is applied to the earthquake engineering field to predict the structural responses under earthquake excitations. The results confirm that the proposed deep neural network provides a superior performance compared to the simplified regression equations which are developed based on a limited dataset. Moreover, to give an insight of the proposed deep neural network, the extracted features from the deep neural network are investigated with various numerical examples. The method is expected to enable engineers to effectively predict the response of a hysteretic system without performing nonlinear time history analyses, and provide a new prospect in the engineering fields. The supporting source code and data are available for download at (the URL that will become available once the paper is accepted).

Keywords: deep learning, convolutional neural network, hysteretic behavior, stochastic excitation, nonlinear time history analysis, single degree of freedom system (SDOF)

1. Introduction

* (Corresponding Author) Associate Professor, Email: os.kwon@utoronto.ca Phone: +1 416-978-5908 Fax: +1 416-978-7046

Assessment of nonlinear hysteretic systems under stochastic excitations is one of the essential topics in many engineering fields such as mechanical, robotics and civil. In such systems, the output values depend not only on the instantaneous input at the given time but also on its past history between an input and an output (i.e. path-dependent system). A shape memory alloy (SMA) is a typical example of a smart material having hysteretic behaviors, which is widely used in industrial fields vibration attenuation (Aizawa et al., 1998) and SMA-based microactuators (Ma et al., 2010; Shin and Garman, 2001). A wire cable vibration isolator which has found many applications in industrial machinery due to its dry friction damping performance is another example of a nonlinear hysteretic system. As the inherent interfacial damping is exerted by sliding or rubbing, the deformation and the force relationship of the wire cable isolator is a nonlinear hysteretic (Chungui et al., 2009). In civil engineering, the structural member is a typical element that shows a nonlinear hysteretic behavior under stochastic excitations. It is, therefore, important for predicting the responses of the system in order to improve the design of the system and make the system reliable during the operation.

When a nonlinear hysteretic system is subjected to a relatively low-intensity stochastic excitation and thus the maximum normalized force is smaller than the yield force, the system exhibits vibration in the linear elastic range. The response estimation of the system under such small external forces is simple and obvious so that a time history analysis is not required. On the other hand, a relatively large-intensity stochastic excitation tends to make a hysteretic system behave nonlinearly during an excitation, which makes it difficult to accurately predict the responses. A time history analysis using both refined numerical models and recorded stochastic excitations is considered as the most accurate way to estimate the responses of the system. This approach solves the dynamic equilibrium equation at every time step using a numerical integration scheme, then the corresponding responses of the nonlinear hysteretic system can be numerically estimated. It is, however, noted that the nonlinear time history analysis involves exceedingly high computational efforts, which makes the approach impractical in most routine engineering processes. In addition, due to the randomness of external vibrations and the uncertainties in the hysteretic behavior of a system, adopting a highly precise yet computationally inefficient method cannot be fully justified.

Some researchers have investigated artificial neural network (ANN) as an alternative to time-consuming analysis procedures (Adeli, 2001; Adeli and Panakkat, 2009; Berényi et al., 2003; Chungui et al., 2009; Dibi and Hafiane, 2007; Tan et al., 2012; Xie et al., 2013; Xiu and Liu, 2009). Since the most of such ANNs are trained for a specific type of a system or a corresponding hysteretic behavior, the applicability of a trained neural network to other hysteretic systems is limited. Furthermore, a relatively small number of hidden layers cannot fully identify the intricate relationships between the output (responses of a nonlinear hysteretic system) and the input information (characteristics of the system and stochastic excitations).

Recently, *deep* neural networks, a variant of ANN which stacks multiple hidden layers, have shown unprecedented performance in terms of predicting the intricate relationship between input and output variables in diverse fields including business, medical, and engineering (LeCun et al., 2015; Schmidhuber, 2015). In particular, the development of a deep *convolutional* neural network (CNN) achieved practical successes, especially in face recognition (Lawrence et al., 1997), image classification (Krizhevsky et al., 2012) and speech recognition (Sainath et al., 2015). Since the CNN shows a clear advantage in dealing with the data having strong spatial correlation by using multiple overlapping filters, we employ a CNN to extract features of a nonlinear hysteretic system from its hysteresis loops. The features extracted from hysteretic information are, then, merged with those representing a stochastic excitation information to form a deep ANN that can predict the responses of a system. Thereby, trained deep neural networks can cover comprehensive information including both the external time-series vibration and the characteristics of a nonlinear hysteretic system.

This paper first presents the overview of the process to predict responses of a nonlinear hysteretic system under stochastic random excitations. A new architecture of the neural network is proposed for which a hysteresis loop of the single degree of freedom (SDOF) system and the intensity features of stochastic excitations are employed as inputs. Next, we demonstrate the applicability and efficiency of the proposed method by applying the earthquake engineering field to predict the seismic demand. To this end, the nonlinear seismic responses database is developed, then using the dataset the proposed network is trained. The extracted features of hysteretic behaviors and stochastic excitations are investigated thoroughly to give a further insight of the trained neural network. Finally, a summary and concluding remarks are provided to promote further researches on the proposed framework.

2. Prediction of responses of a nonlinear hysteretic system under stochastic excitations

A nonlinear hysteretic system excited by random vibration forces produces stochastic responses. In order to simulate such seismic responses of a system, a sound analytical model and recorded excitations are needed. However, most of the routine engineering problems utilize the idealized SDOF system rather than complicated numerical models to reduce the computational cost and modeling efforts. This section, therefore, describes the overall procedure of a nonlinear time history analysis using an SDOF system, then provides the general framework to replace such time-consuming method by using a deep learning method.

2.1 *Time history analysis of a nonlinear hysteretic system*

The governing differential equation of an SDOF system subjected to an excitation $F(t)$ is written as:

$$m\ddot{u} + c\dot{u} + f_s = F(t) \tag{1}$$

where m is the mass, c is the damping coefficient, f_s represents the restoring force function which may depend on the history of the responses, and u, $\dot{u}$ and $\ddot{u}$ respectively denote the displacement, velocity and acceleration of the mass relative to the ground. Equation (1) can be rearranged after dividing both sides by the mass m:

$$\ddot{u} + 2\xi\omega\dot{u} + F_s = F(t)/m \tag{2}$$

where ξ represents the damping ratio (typically 5% is used in practice), ω is the circular natural frequency of the nonlinear hysteretic system, and F_s denotes the normalized restoring force, i.e. f_s/m. Given that ω depends on the relationship between u and F_s, F_s is the only term in the equation that affects the responses of a nonlinear hysteretic system subjected to the specific external force $F(t)$. Figure 1 provides three different hysteretic models which are widely used in the engineering fields: linear elastic, bilinear kinematic hardening, and bilinear stiffness degrading systems. In fact, the hysteretic behavior of a real system can be much more complicated. The three hysteretic behaviors in Figures 1(a), (b) and (c) are respectively denoted as HM1, HM2, and HM3 in this paper. Using the implicit or explicit numerical integration method, a full-time history of responses (i.e. u, $\dot{u}$ and $\ddot{u}$) can be obtained.

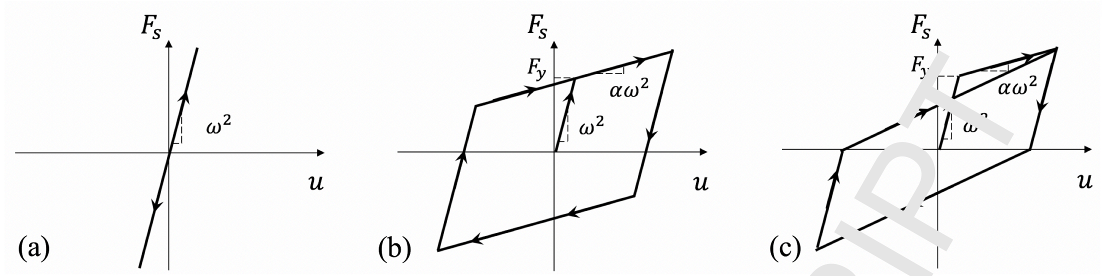

Figure 1. Hysteretic loops of a system: (a) linear (HM1), (b) bilinear kinematic hardening (HM2), and (c) bilinear stiffness degrading system (HM3). The linear system is parametrized by the slope of the hysteretic model which is called stiffness. The stiffness can readily be calculated from the circular frequency ω^2 and the mass m. The bilinear hysteretic systems consist of piecewise linear and continuous relationships, characterized by the circular frequency ω^2, the normalized yield force F_y, and the post-yield stiffness ratio α. The difference between HM2 and HM3 is that the stiffness of the HM3 tends to degrade as damages are accumulated.

2.2 *Deep learning-based response prediction of a nonlinear hysteretic system*

From the engineering perspective, the maximum responses of a nonlinear hysteretic system are the most important outputs during stochastic excitations because the maximum responses are typically used for design or assessment of the hysteretic system. Thus, the goal of the proposed method is to predict the maximum responses of nonlinear hysteretic systems. This section describes the input parameters that are employed in the deep neural network and its architecture to predict the responses of a system.

2.2.1 Input data description

Two different information is needed to estimate the responses of a nonlinear hysteretic system subjected to a stochastic excitation: a nonlinear hysteretic system and a stochastic excitation. The former is represented as a hysteretic loop, i.e. the relationship between the imposed displacement and the restoring force. The hysteretic behavior can be obtained by performing a quasi-static cyclic analysis (i.e. push and pull the nonlinear hysteretic system) of an SDOF system using a predefined displacement step, which captures characteristics of a nonlinear hysteretic system. The example of a hysteretic behavior used as an input is illustrated in Figure 2. Most of the existing methods directly employ the parameters from simplified mathematical models of a nonlinear hysteretic system such as stiffness and yield strength. However, in the proposed method, the hysteretic behavior (i.e. the displacement and the force vectors in Figure 2) is employed as an input parameter to CNN which is often used to extract complicated natural data to lower-level features. Thereby, CNN would automatically extract the useful information for estimating the responses of the nonlinear hysteretic system during training the neural network. On the other hand, the latter is represented by a set of useful features that can illustrate its intensities or characteristics. Since the goal is to estimate the maximum displacement rather than entire responses, it is more efficient to use the intensity measures that are highly correlated with the output value. Moreover, the features of a stochastic excitation are grouped along with their characteristics, then applied as inputs (e.g. group 1: frequency contents, group 2: peak values of a stochastic excitation, and group 3: source information of a stochastic excitation).

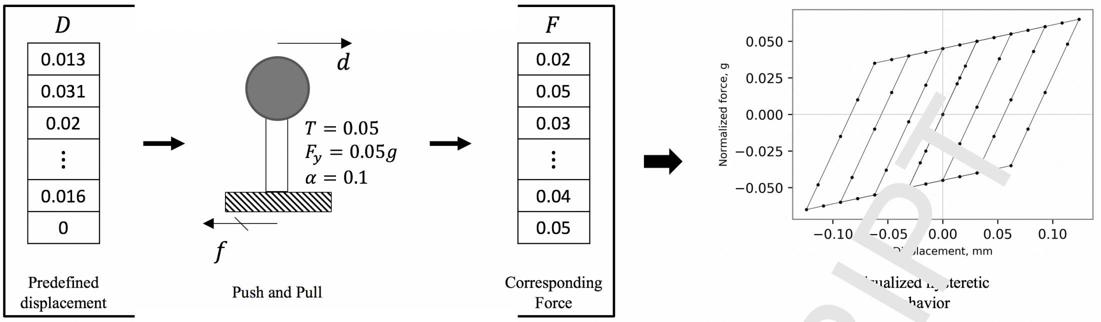

Figure 2. The procedure of generating a hysteretic behavior of a nonlinear hysteretic system as an input of the deep neural network when the system having period 0.05, yield strength 0.05g, and post-yield stiffness 0.1 is selected. Using the predefined displacement steps, the corresponding force can be estimated. The displacement and the estimated force are visualized in the right-hand side of the figure as a scatter plot.

2.2.2 *Architecture of the deep neural network*

Motivated by the natural phenomena of a dynamic excitation to a nonlinear hysteretic system, a new deep neural network architecture in Figure 3 is proposed by incorporating two different information. As mentioned in Section 2.2.1, the features of a nonlinear hysteretic system are extracted from CNN and depicted as an orange box in Figure 3. On the other hand, for the features related to a random excitation, each of the feature group is first processed by ANN (green boxes in Figure 3), then merged with the extracted hysteretic information (orange box in Figure 3) and again processed by ANN, i.e. extracted hysteretic information is added to each of the feature group when the data is processed by ANN. This is to reflect the fact that each stochastic excitation information produces distinct features for a given hysteretic behavior. Finally, each processed feature of a random excitation along with a hysteretic system (dark blue boxes in Figure 3) is merged with extracted hysteretic units (orange box in Figure 3) again to form a single ANN that can predict responses.

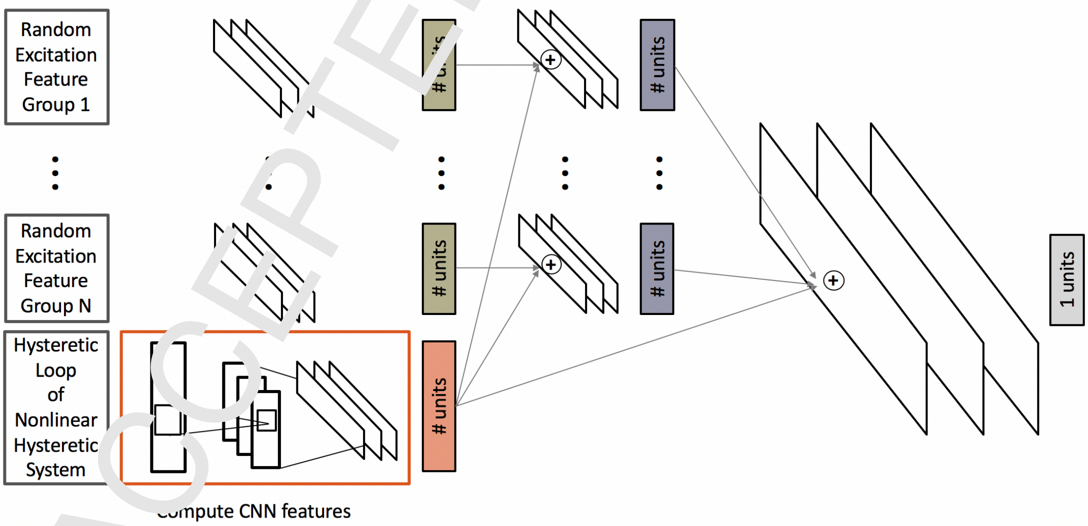

Figure 3. Schematic diagram of the proposed deep neural network.

3. Proposed deep neural network and its application to the earthquake engineering field

To validate the proposed method, we develop a deep neural network to accurately estimate the seismic responses of a nonlinear structural system. In this example, stochastic excitations are earthquake ground motions, and a nonlinear hysteretic system is a structural system such as a building. Although a three-dimensional structural model with detailed structural properties is desirable, an idealized structural model, such as an equivalent SDOF system, is also often used, especially when the structure is being analyzed in the design phase (Ibarra et al., 2005; Ruiz-García and Miranda, 2003; Tothong and Cornell, 2006; Vamvatsikos and Cornell, 2006). For earthquake ground motions, three different types of information are employed, i.e. source information of excitations as earthquake characteristics, stochastic force information as peak values of ground motion time histories, and frequency contents of excitations as a response spectrum, which have been demonstrated to correlate well with the responses of a nonlinear structural system (ASCE 41-13, 2013; ATC 40, 1997; FEMA 440, 2005; Power et al., 2006; Riddell, 2007). The features selected to represent *earthquake characteristics* are magnitude (M), epicentral distance from the site of the structure (R), and the soil class of the site which has five different class (BSSC, 2004). For the *peak values of ground motion time history,* the maximum value of each recorded acceleration (peak ground acceleration; PGA), velocity (peak ground velocity; PGV) and displacement (peak ground displacement; PGD). Lastly, 5%-damped spectral accelerations in the period range of 0.005s to 10s (total 110 steps; $S_a(T)$), i.e. a *response spectrum* is used for frequency contents.

To train the neural network, we generated a dataset of nonlinear time history analysis results. It was found that the extracted features from the CNNs were overfitted when the structural model was randomly selected along with the earthquake ground motion. To address this issue, the same set of structural models is applied to each ground motion when training the network. This section first shows the detailed architecture of deep neural network based on Figure 3. After training the network, the performance of the trained neural network is investigated by comparing with existing methods in terms of the accuracy. Note that the work reported in this paper was performed using Tensorflow (Abadi et al., 2016) and the Compute Canada's GPU cluster.

3.1 *Detailed architecture of the deep neural network*

A detailed diagram of the proposed deep neural network design is provided in Figure 4. All convolutional and neural network layers are followed by a rectified linear operator (ReLU; Nair and Hinton, 2010) except for the first part of the convolutional layers for which a tangent hyperbolic operator (Tanh) is used as the activation function. In addition, the batch normalization (Ioffe and Szegedy, 2015) is adopted for all layers after the ReLU operator is applied. Since the initial stiffness of the hysteresis shows large variability (the ratio of the largest to the smallest stiffness that we cover in this study is about 40,000), the Tanh activation function is used for normalizing parameters describing the hysteretic behaviors. It was observed that, if the ReLU is adopted rather than Tanh at the first convolutional part, the hysteretic features extracted from CNN was overfitted, especially for those having a large initial stiffness.

Four types of filter sizes (2, 4, 8, and 16) are employed to capture the features having different scales. This is helpful, particularly when extracting features from the hysteretic behaviors because the input hysteretic behavior can have different scales (e.g. number of cycles). Moreover, 2×1 max pooling is applied to reduce the dimension of layers in the convolutional phases. The first fully connected layer (FC) that concatenates the output of the four convolutional layers containing 64 neurons is followed by the second FC with the 48-dimensional output. This architecture is motivated by autoencoder (Li et al., 2013; Socher et al., 2013) which forces the layer to engage features having different scales by reducing the neurons. Finally, the output of the network is a single unit (gray box in Figure 4) which is fed to a linear activation function to estimate the continuous variable, i.e. responses of a nonlinear hysteretic system.

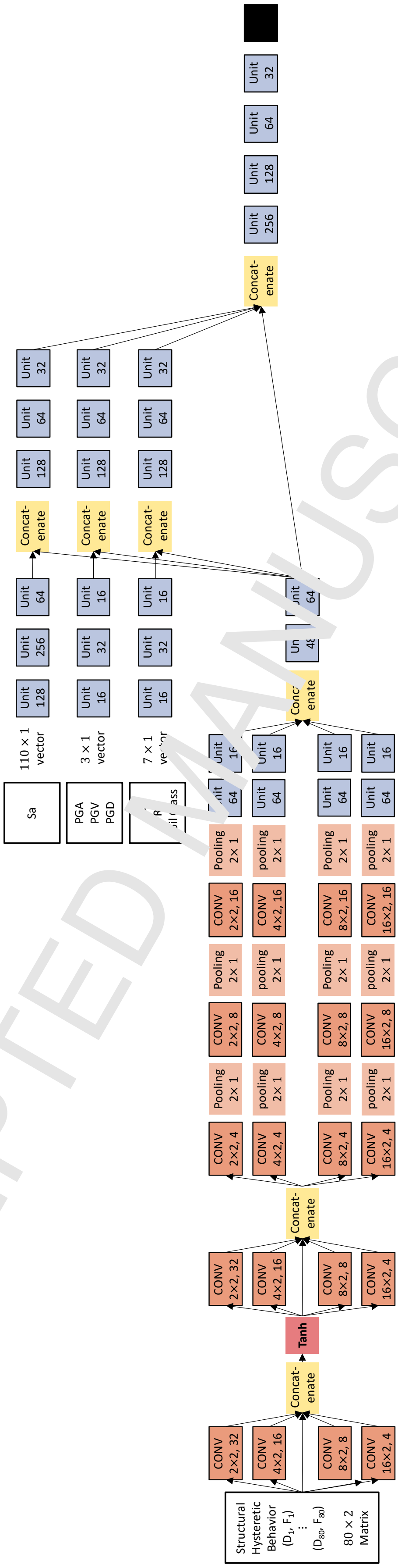

Figure 4. Detailed diagram of the proposed deep neural network architecture

3.2 *Datasets*

To train the neural network, we constructed a database of nonlinear structural responses by performing a large number of nonlinear time history analyses using OpenSees (Mazzoni et al., 2006). The database contains responses of three different hysteretic models (HM1, HM2, and HM3 in Figure 1) subjected to 1,499 ground motions (GMs) obtained from the NGA database (Power et al. 2006). To cover every practical range of structural characteristics, 90 steps of structural period from 0.05 sec to 10 sec, 30 steps of yield strength from 0.05 g to 1.5 g, and 10 steps of post-yield stiffness ratio from 0 to 0.5 were selected. The upper and lower limits of these values were determined based on the capacity curves in HAZUS-MH 2.1 (FEMA 2012) which can describe force-deformation relationships of a wide range of building structures and the intermediate values are uniformly distributed between the upper and lower limits. Therefore, the total number of hysteretic behaviors considered in this study is 54,090 (90 for HM1 and $90 \times 30 \times 10 = 27{,}000$ for each bilinear system). For earthquake ground motion data, based on the authors' experience, even the entire set of GMs (1,499) is insufficient to generalize the ground motion information due to the nature of a *random* excitation We randomly split the 1,499 GMs into a train set of size 1,199 (80%) and a test set of size 300 (20%) in order to check the over-fitting of the proposed deep learning model by monitoring the loss on the test set.

To achieve the improved accuracy with the relatively small number of training time, the dataset is processed before being employed as the input and the output of the proposed network. Since it is widely reported that some of the inputs and the output dataset follow the lognormal distribution (Ellingwood, 2001), the natural logarithm is applied to several input parameters, i.e., response spectrum, PGA, PGV, and PGD, as well as the output, i.e., structural responses, to resolve the skewness. Moreover, in order to allow the network to better generalize the structural hysteretic behaviors, the data augmentation technique (Krizhevsky et al., 2012) is applied by shifting and flipping the hysteresis. After training the proposed network using the transient displacement as the output which is the most important features when designing and assessing structures, a comprehensive investigation is carried out to validate our model. Since any type of structural responses can be used as an output for our model, the trained results using other structural responses, such as the maximum acceleration and the maximum velocity, are also presented in Section 4.3.

3.3 *Initialization*

It is widely known that the performance of a neural network is highly affected by the initialization of weights (Glorot and Bengio, 2010; He et al., 2015; Lecun et al., 1998). In this study, it was found that using randomly initialized weights, the convolutional part of the network cannot extract the features properly, which, in turn, made the performance of the network much worse than those of existing simple regression methods (ATC 40 1997, FEMA 440 2005). To address such an issue, a pre-training is carried out with a small number of samples of the hysteretic behaviors based on the study by Wagner et al. (2013). Only HM1 and HM2 with zero post-yield stiffness (i.e. elasto-perfectly plastic) are selected with randomly selected 894 GMs for pre-training (the number of data set is 2,790×894=2,494,260). The Adam optimizer (Kingma and Ba, 2014; Reddi et al., 2018) is used as the optimization algorithm to reduce the mean squared error (MSE) of the output with 512 batch sizes and 54 epochs of training.

3.4 *Training*

Although a total of 54,090 hysteretic behaviors should be trained for the neural network, we sparsely select the hysteretic behaviors to increase the efficiency of the training. This is based on the premise that the neural network has an ability to estimate the intermediate variables of distinct input values through interpolation. Therefore, 15 steps of yield strength and 7 steps of post-yield stiffness ratio are considered rather than 30 and 10 steps of the original dataset, respectively (total 18,990 structural models). When sparsely selecting the structural models, relatively smaller yield strengths are selected more often than bigger ones to incorporate enough number of cases in which the structural system behaves nonlinearly during an earthquake excitation. That is, hysteretic behaviors of the dataset used for training are prone to behave nonlinearly during earthquake excitation when compared with the dataset that is not used for training. This is because it was found that the nonlinear cases are more difficult to estimate (i.e. bigger MSE) than the linear cases. Moreover, we randomly select 280 GMs among 1,199 GMs which leads to total 5,317,200 datasets for each epoch of training to overcome the limitation of the computational resources. Like the *Initialization* part, the Adam optimizer (Kingma and Ba, 2014; Reddi et al., 2018) is used as the optimization algorithm to reduce the MSE of the output with 512 batch sizes and 300 epochs of training.

3.5 *Prediction accuracy*

3.5.1 Training results

To check whether the trained neural network is overfitted or not, the MSE and the mean absolute error (MAE) are calculated for both train and test dataset. Note that, in this paper, the MSE and MAE are calculated for the natural logarithms of the structural responses and the results are shown in Table 1.

The MSE and MAE of training results show that using the trained deep neural network, one can predict the structural responses that are fairly close to those by a nonlinear time history analysis. Although the MSE and MAE of the test dataset may be significantly higher than those of the train datasets, the discrepancy between the train and test sets is not critical, in that the ratio of the predicted value to the correct value for training and test data set in the original scale (i.e. not applying the natural logarithm) are around $e^{0.05} \approx 1.05$ and $e^{0.12} \approx 1.13$, respectively. This also indicates that 1,199 ground motions may not be enough for properly training the proposed network in terms of a *random* earthquake ground motion information. The authors believe that if more ground motions are used for training, the better results can be obtained for test datasets.

Table 1. The comparison of MSE and MAE between train set and test set

GM Set	Mean squared error	Mean absolute error
Train (1,199 GM)	0.0044	0.0485
Test (200 GM)	0.0253	0.1202

Since the hysteretic behavior is sparsely selected during the training process, the MSE is computed for both trained and non-trained hysteretic models in order to demonstrate the trained network's ability to interpolate over the hysteretic behaviors. 200 GMs are randomly selected from the entire set of the ground motions to obtain the results shown in Table 2. The hysteretic curves of the structural system are generated uniformly, but we intentionally select the systems that are prone to exhibit responses in nonlinear range when training the proposed neural network. Given that the trained deep neural network predicts the linear cases with less MSE than nonlinear cases, the MSE of the nontrained case in Table 2 is less than that of trained cases even though the hysteretic behaviors are not used for training. Moreover, the results confirm that the trained neural network has the ability to interpolate regarding hysteretic behaviors. Therefore, it is expected that the trained neural network will be able to predict structural responses even if the specific hysteretic behaviors were not used in training.

Table 2. The comparison of the MSE between trained and non-trained hysteretic behaviors

Hysteretic behaviors	Mean squared error
Trained	0.0090
Non-trained	0.0077

3.5.2 Comparison with existing methods of seismic response prediction

To demonstrate the superior accuracy of the proposed method, the prediction results are compared with those by the three methods: (1) R-mu-T relationship developed by Nassar and Krawinkler (1991), (2) capacity spectrum method (FEMA 440 2005), and (3) the coefficient method (ASCE 41-13 2013). Such methods have been developed using regression equations with a rather small number of earthquake events to estimate the transient displacement without performing dynamic analysis. The elasto-perfectly plastic systems with randomly selected 200 GMs are provided as inputs for the methods. Among 558,000 cases (= 2,790 × 200) of the entire dataset, the cases showing nonlinearity during excitation (8.03% of the dataset, i.e. 44,842 cases) are investigated. To demonstrate the improvement visually, the differences between the natural logarithms of the transient displacements by nonlinear dynamic analyses and those by the three methods above and the deep neural network are shown in Figure 5. The comprehensive comparison results are presented in Table 3 in terms of the MSE. The results demonstrate that even though the structures handled in this example show a simple hysteresis, the existing methods cannot predict the transient displacement properly compared to our deep learning model. In other words, the proposed method can predict the nonlinear responses with the smaller size of the uncertain error compared to the existing methods.

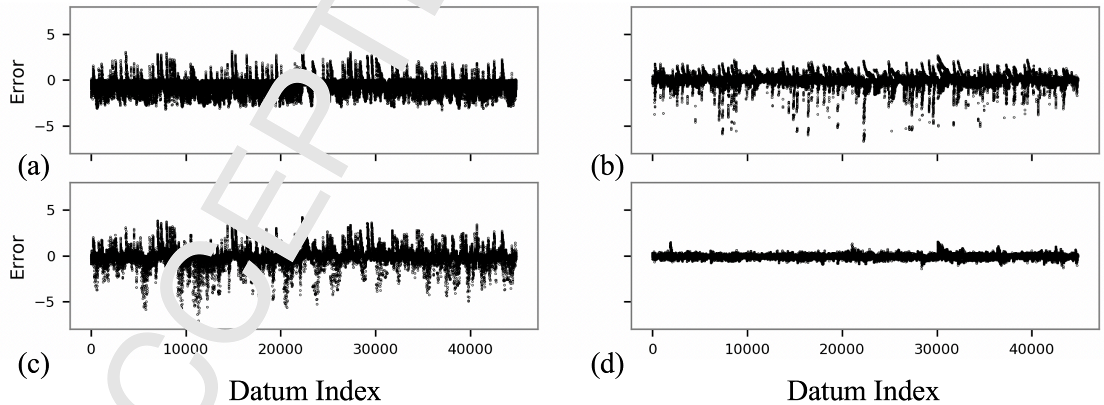

Figure 5. The differences between the natural logarithms of the transient responses by nonlinear time history analyses and those by (a) R-μ-T relationship by Nassar and Krawinkler (1991), (b) capacity spectrum method (FEMA 440 2005), (c) the coefficient method (ASCE 41-13 2013), and (d) proposed deep neural network

Table 3. The comparison of the MSE between proposed and existing methods

Methods	Mean squared error

R-μ-T	0.9857
Capacity spectrum method	0.3526
Coefficient method	0.6110
Deep neural network	0.0407

4. Validation of trained deep neural network

After training the neural network to estimate the structural responses, an investigation has been made to give a further insight of the proposed network. To validate the feature extraction of a nonlinear hysteretic system part of the trained neural network, the extracted features from the CNN are analyzed thoroughly by introducing the other hysteretic models which are not used in training. Moreover, the analysis of the intermediate features in CNN is also carried out to demonstrate the role of different filter sizes used during training. For the feature extraction of a stochastic excitation part, we train the neural network using the sub-samples of the feature groups of the stochastic excitation, i.e. total 6 different combinations of the input parameters and the corresponding neural networks are produced. It is numerically verified that the proposed method has superior performance in terms of its applicability and effectiveness.

4.1 *Validation of feature extraction in a nonlinear hysteretic system part*

4.1.1 Investigation of extracted features from CNN

Despite the great success of deep neural network, there exist concerns and criticisms that the method provides *black boxes* (Alain and Bengio, 2016; Shwartz-Ziv and Tishby, 2017). In order to understand the input parameters passed through the CNN in the developed deep neural network, extracted features from CNN are plotted and investigated based on the knowledge of the dynamics. First of all, the output of the final FC which was obtained right after application of the ReLU activation function, i.e. 64 vectors (orange box in Figure 3), is plotted in Figure 6 (i.e. 64 lines). Herein, hysteretic models of HM1 (90 hysteretic behaviors) are imposed as the input of the trained neural network. We plot the extracted values of the hysteretic behaviors whose natural frequency (period) is big (short) to small (long) as left to right. As shown in Figure 6, each element of the 64 units is activated along with the different natural frequency of a nonlinear hysteretic system (bell shape plot at specific structural system), which resembles the modal analysis finding the various periods at which the system naturally resonates. It is expected that since the extracted features represent the frequency contents of a nonlinear hysteretic system, merging with the frequency contents information of a stochastic excitation may yield the proper estimation of system's responses, which will be investigated in Section 4.2.

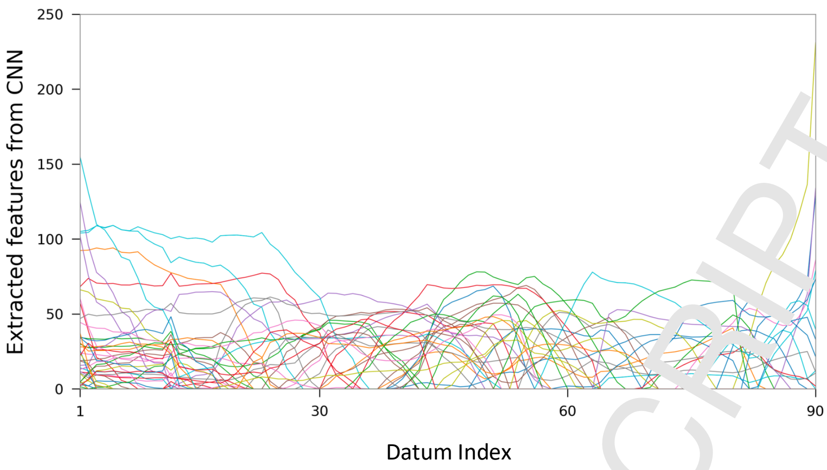

Figure 6. Extracted features from CNN when HM1 is used as an input. x-axis represents each of hysteretic behavior and y-axis represents the values obtained from the CNN. To find the pattern clearly, the values of the same element in the 64 units are connected each other.

To give further insight, the extracted features from CNN are compared along with the hysteretic types. The experiment is based on the assumption that if different hysteretic models share some features such as the same initial stiffness (ω^2 in Figure 1 for HM1, HM2, and HM3), it is expected that some of the extracted features from CNN of different hysteretic models may also be equivalent to each other and show meaningful trends.

Figure 7 shows the extracted features from CNN when all hysteretic models of HM1 (90 hysteretic behaviors) and zero post-yield stiffness (i.e. $\alpha = 0$) cases of HM2 (2,700 hysteretic behaviors) and HM3 (2,700 hysteretic behaviors) are used as inputs of CNN. The figure is plotted as follows: When one of the behaviors of HM1 is applied as an input of CNN, the 64×1 vectors can be obtained as an extracted feature vector. Then, we plot each element of the vector in separate boxes. After the extracted features of every hysteretic behavior of HM1 are mapped out in Figure 7(a), we obtain the 64 separate boxes whose number of indices along the *x*-axis is 90. Using the same procedure, Figure 7(b) and (c) represents the output vector when HM2 and HM3 are imposed as the input of CNN, respectively. In addition, the extracted features of the hysteretic behaviors are plotted as the following order from left to right for each box: From hysteretic behaviors whose period is short to long, and the yield strength of each period is small to large, especially for HM2 and HM3, sequentially. In order to reduce the repeated plots in Figure 7, we only illustrate 5 indices among 64 whose values can represent the distinct information.

As shown in Figure 7, It is observed that the overall shapes of the corresponding box of first, second, and third rows of extracted features from CNN are similar except the relatively short period range (the 34th boxes) and smaller yield strength (the plots fluctuates along with yield strength) in which the structure is prone to behave nonlinearly during vibration. Since bilinear models need additional information (i.e. nonlinearity of the hysteretic system) compared to a linear model, the 22nd index captures such information as shown in only extracted features of HM2 and HM3 fluctuate. Based on this investigation, it is concluded that the extracted features from CNN may give some physical meanings of hysteresis even though it is hard to interpret as widely used mathematical parameters such as initial stiffness and yield strength.

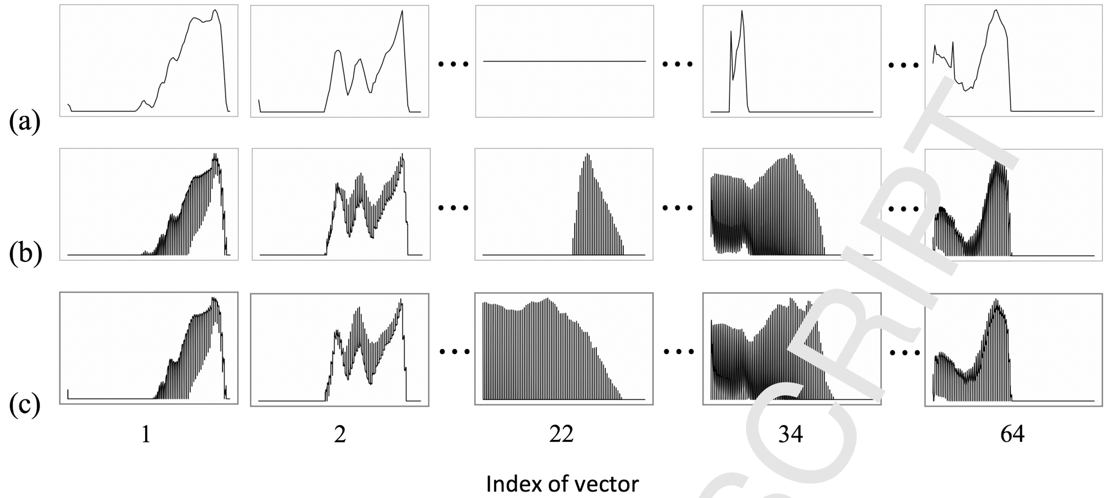

Figure 7. The plot of extracted features from CNN, i.e. the output of the final FC layer when (a) linear elastic (HM1), (b) elasto-perfectly plastic (HM2 with $\alpha = 0$), and (c) elasto-perfectly plastic with stiffness degradation (HM3 with $\alpha = 0$) are used as inputs. In words, Figure 6 is the combination of Figure 7(c).

Another way to investigate the extracted features from CNN is to compare the distance between the extracted features of trained hysteresis and that of a new hysteresis such as Bouc-Wen model (Baber and Noori, 1985; Song and Der Kiureghian, 2006, Wen, 1976). The main premise is that, if the CNN is properly trained, the force-displacement relationship of a new hysteretic behavior, which was not used for training of the CNN, should be well described by the CNN whose extracted features has the smallest distance (i.e. Euclidian $l2$-norm) from that of the new hysteresis of concern. Figure 8 shows two cases from Bouc-Wen hysteresis model (black solid line) and the corresponding equivalent hysteretic behavior (red dashed line). This result also clearly confirms that the CNN can properly extract important features of hysteresis behaviors.

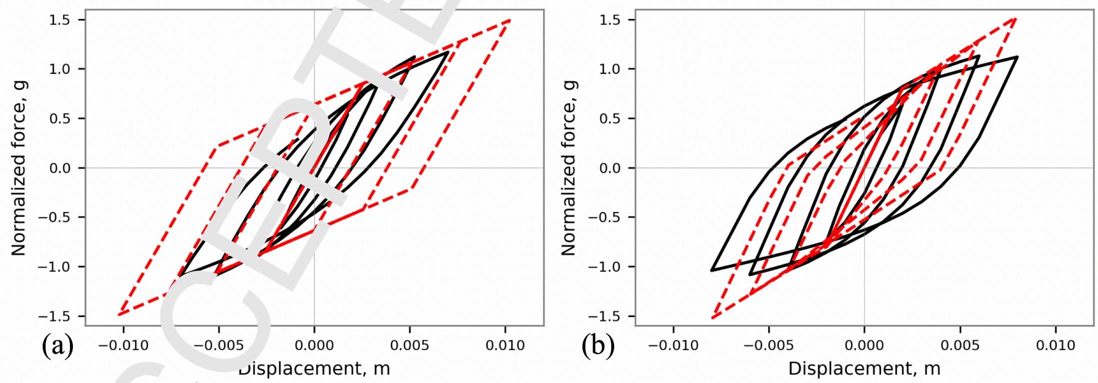

Figure 8. The hysteretic behavior generated from Bouc-Wen model (black solid line) and the equivalent hysteretic behaviors (red dashed line). We compute the Euclidian distance between extracted features of the equivalent hysteresis of all hysteresis in the database (54090 cases) and that of Bouc-wen model. Then, finds the hysteretic behavior having the smallest distance and denotes such hysteresis as *the equivalent hysteretic behaviors*. The equivalent hysteretic behaviors of this example are as follows: (a) bilinear model with period 0.11 sec, yield strength 0.8g and post-yield stiffness ratio 0.25, and (b) bilinear with stiffness degradation model with period 0.10 sec, yield strength 0.75g, and post-yield stiffness ratio 0.3.

4.1.2 Influence of different filter sizes in CNN

Since four types of filter sizes, i.e. 2×2, 4×2, 8×2 and 16×2, are used to train the proposed neural network as shown in Figure 4, a parametric study is carried out to investigate the sensitivity with respect to filter size of CNN. The intermediate features of hysteretic behaviors which are extracted from the different filter sizes (i.e. the layers having 16 units before concatenating each other) are plotted in Figure 9 when hysteretic models of HM1 are employed as an input. Similarly to Figure 6, x-axis represents each hysteretic behavior and y-axis represents the values obtained from the different filter sizes. While the 2×2 filter captures the overall characteristics of the hysteretic behaviors, other filters capture specific frequency contents of the system. For example, most of the extracted features from 8×2 filters depicted in Figure 9(c) are activated when the hysteretic behaviors having large period are imposed as an input. Although some values are activated (e.g. purple, blue, and green lines in Figure 9(c)) when the hysteretic behaviors having a small period is used, they give almost the same values even the input of hysteretic behaviors are changed. In words, the CNN part of the proposed network first extracts the different kinds of frequency contents along with the filter size (four different vectors) as shown in Figure 9, then combines the information to produce the overall frequency contents of a nonlinear hysteretic system in the final layer (one vector) as shown in Figure 6.

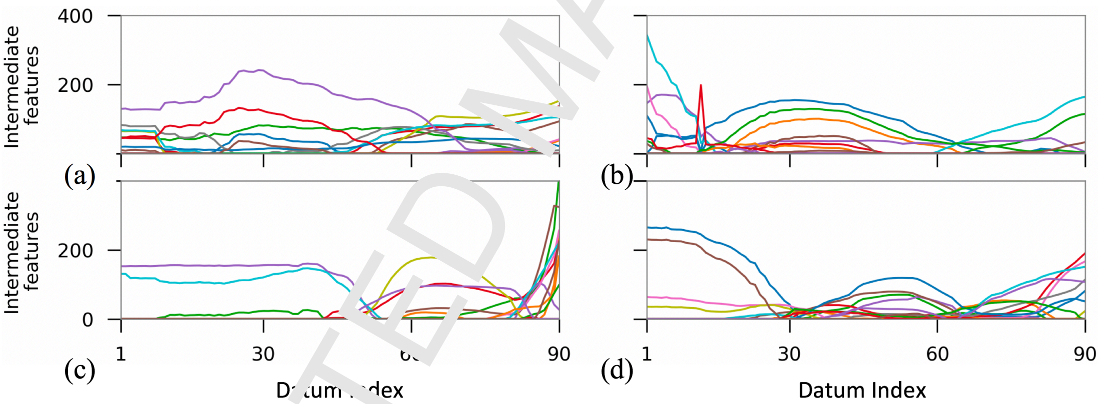

Figure 9. Intermediate features in CNN when HM1 is used as an input along with the filter size in CNN: (a) 2×2, (b) 4×2, (c) 8×2, and (d) 16×2 is used in CNN

4.2 *Deep neural network trained with different features of a stochastic excitation*

Three different categorized features of a stochastic excitation are used for training in the previous section: source information of excitations (SE), peak values of excitation force (PE) and frequency contents (FC). To investigate the relationship between input features and the output in terms of accuracy, the neural network is trained using 6 different types of input parameters (i.e. subsamples of three categorized features). According to the fundamentals in Structural Dynamics and discussed earlier in the Section 4.1.1, the frequency contents are closely related with the responses of a nonlinear hysteretic system than any other features. Thus, it is expected that the neural network trained with a response spectrum (i.e. FC) shows a superior performance. The same training environments that are used in Section 3 are applied to 6 different neural networks, and the MSE and MAE for the test dataset are shown in Table 4.

Table 4. The comparison of the MSE and the MAE of the deep neural networks which are trained with different stochastic excitation information

Input	MSE	MAE
SE	0.9597	0.7459
PE	0.1574	0.3070
FC	0.0496	0.1644
SE & PE	0.1341	0.2803
PE & FC	0.0400	0.1604
FC & SE	0.0522	0.1715

As expected, the neural networks that are trained with frequency contents show better performance than other trained networks. In particular, even the predicted responses of the neural network trained only with FC are more accurate than the one trained with the other two excitation information. In addition, all MSE and MAE values of Table 5 are bigger than those in Table 1. Since the SE, PE and, FC stand for different characteristics of a stochastic excitation, the deep neural network can properly find the relationship between the input and the output with enough information. However, it is not always good to use all features as inputs of the deep neural network. The MSE and MAE of the neural network trained with FC & SE are bigger than those by the network trained only with FC, which means the deep neural network recognizes that SE does not significantly influence the maximum responses. Therefore, a comprehensive understanding of a problem is needed even using non-overlapping features as inputs of the network. Using the deep neural network, one can identify input parameters making a significant influence on the output value, but a thorough investigation is needed to capture the loss of information quantitatively, which will be carried out in the future studies.

4.3 *Prediction of other structural responses by neural network*

In earthquake engineering, the transient acceleration and the transient velocity are also important responses to predict because they are closely related with the base shear capacity and the performance of non-structural components. To reduce the computational time for training, a transfer learning method (Pan and Yang, 2010; Yosinski et al., 2014) is employed to train each response using the neural network trained to predict the transient displacements. The parameters of the trained neural network are fixed except those after the final merged layers. This is because the layers before the final merged layers extract the features from both earthquake information and structural information, which should not be different even if the different structural responses are predicted. The training is carried out in the same environments that are used for the transient displacement. The MSE and the MAE are computed for the natural logarithms of the predicted transient acceleration (Table 5) and transient velocity (Table 6), after 30 epochs of training. Moreover, Figure 10 and Table 7 show the errors in the transient acceleration and the transient velocity employing the input dataset used in Section 3.5.2. The errors are almost the same as those of the prediction of the transient displacement, which confirms that the proposed neural network can provide highly accurate predictions regardless of the responses type.

Table 5. MSE and MAE of transient acceleration for train sets and test set

GM Set	Mean squared error	Mean absolute error
Train (1,199 GM)	0.0131	0.0903
Test (300 GM)	0.0267	0.1243

Table 6. MSE and MAE of transient velocity for train sets and test set

GM Set	Mean squared error	Mean absolute error
Train (1,199 GM)	0.0278	0.1319
Test (300 GM)	0.0433	0.1624

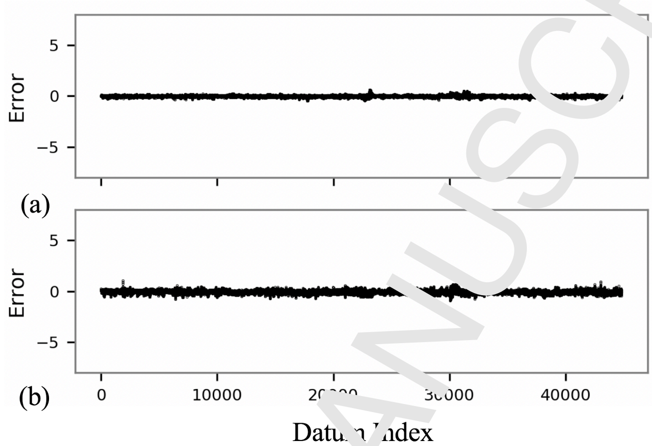

Figure 10. The errors for the natural logarithms of the predicted responses: (a) acceleration and (b) velocity when applying the dataset used in Section 3.5.2.

Table 7. The MSE of dataset for the elasto-perfectly plastic system of HM2 and HM3 subjected to 200 GMs used in Section 3.5.2.

GM Set	Mean squared error
Transient acceleration	0.0101
Transient velocity	0. 0273

5. Conclusions and future research

To facilitate efficient prediction of the responses of nonlinear hysteretic systems without compromising accuracy, this study proposes a new deep-learning-based model for the responses of single degree of freedom system by introducing a convolutional neural network. The displacement-force relationship (i.e. hysteresis) of a nonlinear system and the intensity features of a stochastic excitation are imposed as an input to estimate the maximum responses of the system which are the most important values in the engineering fields. The proposed architecture of the deep neural network is applied to the earthquake engineering problem to demonstrate its applicability and effectiveness. To train the proposed network, the database of nonlinear structural responses is constructed through a large number of nonlinear dynamic analyses. A small set of data is applied to initialize the weight of the network and a proper training scheme is used to compensate the limitation of computational resources. The accuracy of the proposed method is superior to that of the three existing simple methods that are widely used in practice. It is expected that a variety of applications in the earthquake engineering field can be developed using the proposed method such as efficient and accurate regional seismic loss estimation. The numerical investigations demonstrate that the proposed method successfully extracts hysteretic behaviors to the lower-level features and is not overfitted to a certain data set. Furthermore, one can find the most important feature set of stochastic excitations by comparing the accuracy of neural networks training with different input feature set. Currently, a further study is underway to extend the proposed framework to multi degree of freedom systems using the complete quadratic combination with the results of modal analysis of a nonlinear hysteretic system (Chopra and Goel, 2002). Moreover, the proposed framework is being further developed for uncertainty quantification by introducing Bayesian neural network (Kendall and Gal, 2017) to deal with the propagation of the input randomness to the responses. Although this paper focuses on the maximum responses of a nonlinear hysteretic system, the authors believe that a full-time history of responses can be obtained by incorporating a recurrent neural network to the proposed framework.

Acknowledgement

The first author is supported by the International Research & Development Program of the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (Grant number: 2017K1A3A1A12034682), the second author is supported by the Korean Federation of Science and Technology Societies (KOFST) grant funded by the Korean government (MSIP: Ministry of Science, ICT and Future Planning), and the third author is supported by the Institute of Engineering Research at Seoul National University and the project "Development of Life-cycle Engineering Technique and Construction Method for Global Competitiveness Upgrade of Cable Bridges" of the Ministry of Land, Infrastructure and Transport (MOLIT) of the Korean Government (Grant No. 16SCIP-B119960-01). Finally, this research was enabled in part by support provided by WestGrid (www.westgrid.ca) and Compute Canada (www.computecanada.ca). Any opinions, findings, and conclusions expressed in this paper are those of the authors, and do not necessarily reflect the views of the sponsors.

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2016. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. ArXiv160304467 Cs.
2. Adeli, H., 2001. Neural Networks in Civil Engineering: 1989–2000. Comput.-Aided Civ. Infrastruct. Eng. 16, 126–142.
3. Adeli, H., Panakkat, A., 2009. A Probabilistic Neural Network for Earthquake Magnitude Prediction. Neural Netw. 22, 1018–1024.
4. Alain, G., Bengio, Y., 2016. Understanding Intermediate Layers Using Linear Classifier Probes. ArXiv161001644 Cs Stat.
5. ASCE, 2013. Seismic Evaluation and Rehabilitation of Existing Buildings, ASCE/SEI 41-13 (Public Comment Draft), American Society of Civil Engineers, Reston, VA.

6. ATC, 1997. Seismic Evaluation and Retrofit of Concrete Buildings, Report No. ATC-40, Applied Technology Council, Redwood City, CA.
7. ATC., 2005. "Improvement of Nonlinear Static Seismic Analysis Procedures." Rep. No. FEMA-440, Washington, D.C.
8. Aizawa, S., Kakizawa, T., Higasino, M., 1998. Case Studies of Smart Materials for Civil Structures, J. Smart Mater. Struct. 7 617–626.
9. Baber, T.T., Noori, M.N., 1985. Random Vibration of Degrading, Pinching Systems. J. Eng. Mech. 111, 1010–1026.
10. Berényi, P., Horváth, G., Lampaert, V., Swevers, J., 2003. Nonlocal Hysteresis Function Identification and Compensation with Neural Networks. Period. Politech., Ser. El. Eng., 47, nos. 3–4, pp. 253.
11. BSSC., 2003. NEHRP Recommended Provisions for Seismic Regulations for New Buildings and Other Structures, Part1: Provisions, FEMA 368, Federal Emergency Management Agency, Washington, D.C.
12. Chopra, A.K., Goel, R.K. 2002. A Modal Pushover Analysis Procedure for Estimating Seismic Demands for Buildings, Earthquake Eng. Struct Dyn. 31, 561–582
13. Chungui, Z., Xinong, Z., Shilin, Z., Tong, Z., Changchun, Z., 2009. Hybrid Modeling of Wire Cable Vibration Isolation System Through Neural Network, Mathematics and Computers in Simulation, 79, 3160–3173.
14. Dibi, Z., Hafiane, M.L., 2007. Artificial Neural Network-Based Hysteresis Estimation of Capacitive Pressure Sensor. Phys. Status Solidi B 244, 468–473.
15. Ellingwood, BR., 2001. Earthquake risk assessment of building structures, Reliability Eng. and System Safet, 74(3), 251-262.
16. FEMA., 2012. Multi-Hazard Loss Estimation Methodology. Earthquake model – HAZUS MH 2.1 Technical Manual, Washington, USA
17. Glorot, X., Bengio, Y., 2010. Understanding the Difficulty of Training Deep Feedforward Neural Networks, in: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. Presented at the Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256.
18. He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. ArXiv:1502.01852 Cs.
19. Ibarra, L.F., Medina, R.A., Krawinkler, H., 2005. Hysteretic Models That Incorporate Strength and Stiffness Deterioration. Earthq. Eng. Struct. Dyn. 34, 1489–1511.
20. Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. ArXiv150203167 Cs.
21. Kendall, A., Gal, Y., 2017. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?, arXiv preprint arXiv:1703.04977, 2017
22. Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. ArXiv14126980 Cs.
23. Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in: Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q. (Eds.), Advances in Neural Information Processing Systems 25. Curran Associates, Inc., pp. 1097–1105.
24. Lawrence, S., Giles, C.L., Ah Chung Tsoi, Back, A.D., 1997. Face Recognition: A Convolutional Neural-Network Approach. IEEE Trans. Neural Netw. 8, 98–113.

25. LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep Learning. Nature 521, 436–444.
26. Lecun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based Learning Applied to Document Recognition. Proc. IEEE 86, 2278–2324.
27. Li, P., Liu, Y., Sun, M., 2013. Recursive Autoencoders for ITG-Based Translation, in: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, Seattle, Washington, USA, pp. 567–577.
28. Ma, C.C., Wang, R., 2000. Frequency Response of TiNi Shape Memory Alloy Thin Film Microactuators, in: Proceedings of IEEE, 13th Annual International Conference on Micro Electro Mechanical Systems, 370–374.
29. Mazzoni, S., Mckenna, F., Scott, M.H., Fenves, G.L., 2006. The Open System for Earthquake Engineering Simulation (OpenSEES) User Command-Language Manual.
30. Nair, V., Hinton, G.E., n.d. Rectified Linear Units Improve Restricted Boltzmann Machines 8.
31. Nassar, A., Krawinkler, H., 1991. Seismic Demands for SDOF and MDOF Systems, John A. Blume Earthquake Engineering Center Report No. 95, Stanford University, CA.
32. Pan, S.J., Yang, Q., 2010. A Survey on Transfer Learning. IEEE Trans. Knowl. Data Eng. 22, 1345–1359.
33. Power, M., Chiou, B., Abrahamson, N., Roblee C., 2006. The “Next Generation of Ground Motion Attenuation Models” (NGA) Project: An Overview. In Proceedings, Eighth National Conference on Earthquake Engineering, Paper No. 2022.
34. Reddi, S.J., Kale, S., Kumar, S., 2018. On the Convergence of Adam and Beyond. ICLR 2018 Conference Blind Submission.
35. Riddell, R., 2007. On Ground Motion Intensity Indices. Earthq. Spectra 23, 147–173.
36. Ruiz-García, J., Miranda, E., 2003. Inelastic Displacement Ratios for Evaluation of Existing Structures. Earthq. Eng. Struct. Dyn. 32, 1237–1258.
37. Sainath, T.N., Kingsbury, B., Saon, G., Soltau, H., Mohamed, A.-R., Dahl, G., Ramabhadran, B., 2015. Deep Convolutional Neural Networks for Large-scale Speech Tasks. Neural Netw. 64, 39–48.
38. Schmidhuber, J., 2015. Deep Learning in Neural Networks: An overview. Neural Netw. 61, 85–117.
39. Shin, D.D., Garman, G.P. 2001. Operating Frequency of Thin Film NiTi in Fluid Media, in: Proceedings of the ASME IMECE, MEMS Symposium.
40. Shwartz-Ziv, R., Tishby, N., 2017. Opening the Black Box of Deep Neural Networks via Information. ArXiv170300810 Cs.
41. Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C., n.d. 2013. Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank. In EMNLP 2013, pages 1631–1642
42. Song J., Der Kiureghian A., 2006. Generalized Bouc–Wen Model for Highly Asymmetric Hysteresis. J. Eng. Mech. 132, 610–618.
43. Tan, Y., Dong, R., Chen, H., He, H., 2012. Neural Network Based Identification of Hysteresis in Human Meridian Systems. Int. J. Appl. Math. Comput. Sci. 22, 685–694.
44. Tothong, P., Cornell, C.A., 2006. An Empirical Ground-Motion Attenuation Relation for Inelastic Spectral Displacement. Bull. Seismol. Soc. Am. 96, 2146–2164.
45. Vamvatsikos, D., Cornell, C.A., 2006. Direct Estimation of the Seismic Demand and Capacity of Oscillators with Multi-linear Static Pushovers Through IDA. Earthq. Eng. Struct. Dyn. 35,

1097–1117.

46. Wagner, R., Thom, M., Schweiger, R., Palm, G., Rothermel, A., 2013. Learning Convolutional Neural Networks from Few Samples. IEEE, pp. 1–7.
47. Wen, Y.-K., 1976. Method for Random Vibration of Hysteretic Systems. J. Eng. Mech. Div. 102, 249–263.
48. Xie, S.L., Zhang, Y.H., Chen, C.H., Zhang, X.N., 2013. Identification of Nonlinear Hysteretic Systems by Artificial Neural Network. Mech. Syst. Signal Process. 34, 76–87.
49. Xiu, C., Liu, Y., 2009. Hysteresis Response Neural Network and Its Applications, 2009 ISECS International Colloquium on Computing, Communication, Control, and Management. pp. 361–364.
50. Yosinski, J., Clune, J., Bengio, Y., Lipson, H., 2014. How transferable are features in deep neural networks? ArXiv14111792 Cs.