applied sciences

MDPI

Article

Development of a Control Algorithm for a Semi-Active Mid-Story Isolation System Using Reinforcement Learning

Hyun-Su Kim [1,*] and Uksun Kim [2]

[1] Division of Architecture, Sunmoon University, Asan-si 31460, Republic of Korea
[2] Civil and Environmental Engineering Department, California State University, Fullerton, CA 92831, USA
* Correspondence: hskim72@sunmoon.ac.kr; Tel.: +82-41-530-2315

Abstract: The semi-active control system is widely used to reduce the seismic response of building structures. Its control performance mainly depends on the applied control algorithms. Various semi-active control algorithms have been developed to date. Recently, machine learning has been applied to various engineering fields and provided successful results. Because reinforcement learning (RL) has shown good performance for real-time decision-making problems, structural control engineers have become interested in RL. In this study, RL was applied to the development of a semi-active control algorithm. Among various RL methods, a Deep Q-network (DQN) was selected because of its successful application to many control problems. A sample building structure was constructed by using a semi-active mid-story isolation system (SMIS) with a magnetorheological damper. Artificial ground motions were generated for numerical simulation. In this study, the sample building structure and seismic excitation were used to make the RL environment. The reward of RL was designed to reduce the peak story drift and the isolation story drift. Skyhook and groundhook control algorithms were applied for comparative study. Based on numerical results, this paper shows that the proposed control algorithm can effectively reduce the seismic responses of building structures with a SMIS.

Keywords: semi-active mid-story isolation system; control algorithm; reinforcement learning; seismic response reduction; magnetorheological damper; deep q-network

Citation: Kim, H.-S.; Kim, U. Development of a Control Algorithm for a Semi-Active Mid-Story Isolation System Using Reinforcement Learning. *Appl. Sci.* **2023**, *13*, 2053. https://doi.org/10.3390/app13042053

Academic Editors: Agostino Forestiero and Antonio Mannella

Received: 11 January 2023
Revised: 31 January 2023
Accepted: 2 February 2023
Published: 4 February 2023

1. Introduction

Various seismic protection systems for building structures have been developed and used in practical engineering fields. Studies on semi-active control systems (SACS) are being ardently performed [1–3], because SACS have the reliability of passive systems and the controllable features of active systems. SACS are applied in various types of practical seismic protection systems for building structures. For example, a semi-active damper replaces a passive damper in a tuned mass damper or seismic isolation system, resulting in a semi-active tuned mass damper [4,5] or a semi-active seismic isolation system [6,7], respectively. When SACS are applied to building structures subjected to earthquake loads, the seismic response reduction performance is mainly affected by the applied control algorithms. Therefore, much research on the development of semi-active control algorithms has been conducted to date [8–10]. Each semi-active control strategy has its own advantages and shortcomings, depending on the applied problems and target responses. Among various types of semi-active control algorithms, soft computing-based controllers, such as artificial neural network, genetic algorithms, and fuzzy logic controllers, have been successfully applied to SACS. Because soft computing-based controllers can effectively handle nonlinear systems and uncertainties, they can consider structural parameter change due to possible damage to the structure and provide good performance under various types of excitation [11].

Recently, machine learning (ML), which is one of the theories of soft computing, has been successfully incorporated into a variety of applications in various engineering

fields. Reinforcement learning (RL) is one of the machine learning methods using an agent that interacts with the environment for learning. Reinforcement learning is a model-free framework for solving the control problem presented by a Markov decision process (MDP) [12]. Such RL has been applied to various types of active control problems, such as vehicles [13], controllable tensegrity structures [14], and robotics [15], presenting successful control performances.

Recent advancements in reinforcement learning have attracted the attention of structural engineers. Eshkevari et al. [16] used a RL framework for an active control system (ACS) of a five-story building. The proposed method was successful in reducing dynamic responses in comparison with a conventional ACS. Khalatbarisoltani et al. applied reinforcement learning to the real-time control of an active mass driver system as a seismic control system [17]. In their work, a control method was developed for the ACS, considering structural uncertainties. Although several studies on the application of reinforcement learning to an ACS for seismic response reduction have been conducted, to the best of the authors' knowledge, application of reinforcement learning to a SACS for seismic protection of building structures has rarely been reported. Most of the studies on the application of reinforcement learning to SACS have focused on vehicle suspension systems [18,19].

Based on this background, the authors developed a semi-active control algorithm for the seismic damage protection of building structures using reinforcement learning. From among various semi-active earthquake-resisting systems for building structures, a seismic isolation system was used in this study. While a base isolation system is successfully used for seismic response reduction of low- and mid-rise structures, a mid-story isolation system is applied to high-rise building structures [20,21]. An existing building structure with a mid-story isolation system was selected as a sample structure (Shiodome Sumitomo building in Japan [20]). The nonlinear hysteretic dampers in the isolation system of the Shiodome Sumitomo building were replaced by magnetorheological (MR) dampers to compose a semi-active mid-story isolation system (SMIS). Artificial earthquakes were generated for numerical simulation. To optimize the control commands sent to the MR dampers in the SMIS, a deep Q-network (DQN) algorithm was selected in this study. The DQN showed a successful application to many Atari games and various problems by combining RL and deep neural networks [22,23]. The reinforcement learning environment was constructed by using the sample building structure with the SMIS and the generated artificial earthquakes. Because maximum story drift is closely correlated with structural damage, it was used to design the reward of reinforcement learning. Conventional semi-active skyhook and groundhook control algorithms [24] were employed for comparative study.

2. Reinforcement Learning Framework with DQN

Machine learning generally means a computer learning from data using algorithms to perform a task [25]. When a machine learns, an algorithm acquires skills or knowledge from experience. There are many different types of machine learning techniques. Figure 1 presents the three main learning types in machine learning. Supervised learning makes a model learn a mapping between input data and the output targets [26]. Unsupervised learning develops a model that can extract relationships from input data without outputs or target data. Unlike supervised learning, unsupervised learning does not have a teacher and is generally used for clustering complex data. RL develops an agent that learns using feedback with an environment. An agent learns how to map states to actions in order to maximize a reward in RL. The agent is not guided on which action is good or bad but has to find out which action provides the most reward [22].

State, action, and reward are the main basic concepts in RL. The state shows the current situation in the environment. In this study, the states are the seismic responses of the sample structure and ground motions that can explain the characteristics of the excitation. Action means what an agent can do in each state. The semi-active control algorithm, which is the agent in this study, decides the control command for the SMIS based on the given state. The control command produced by the agent is the command voltage sent to the MR

dampers in the SMIS. The MR damper used in this study has a voltage range of 0 to 5 V. The finite actions the agent can take are determined to be eleven voltage values in the range between 0 and 5 V with 0.5 V intervals. When an agent takes an action in a state, it receives a reward. Generally, an agent in RL is trained to maximize the reward. Because the purpose of the SMIS is to reduce the seismic responses of the sample structure, the maximization of the reward should be achieved by decreasing seismic responses of the sample structure. Therefore, it is required to use the seismic responses of the sample structure as negative values or denominators in fractions for reward design. Figure 2 illustrates the configuration of the environment and the agent, including the feedback data types used in this study.

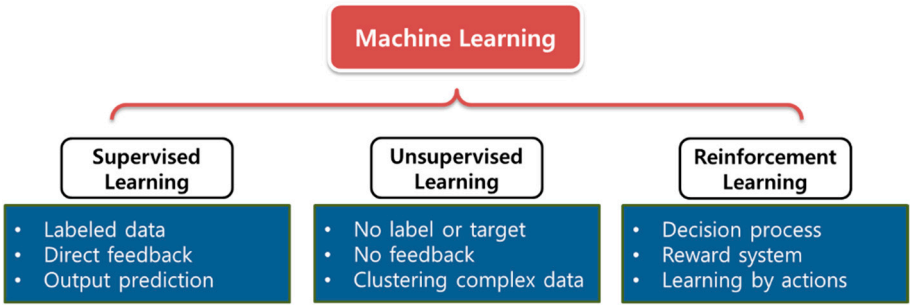


Figure 1. Machine learning categories.

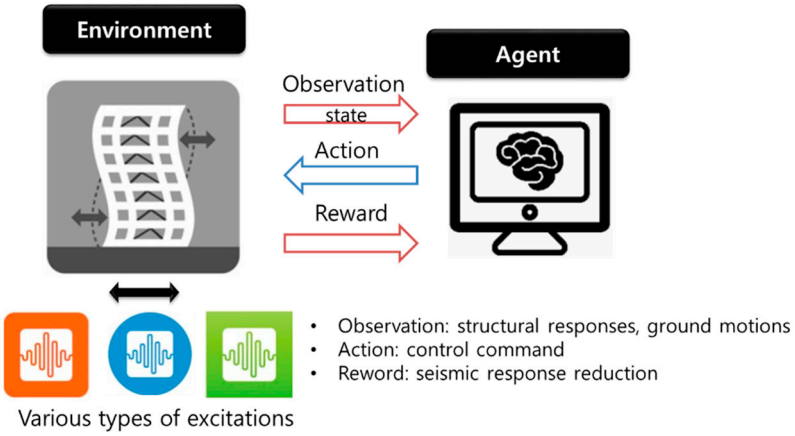


Figure 2. Relationship of the environment and agent in reinforcement learning.

A lot of methods of RL have been proposed to date. Among them, the deep Q-network (DQN) is widely used because it has shown good performance in various control problems. DQN is a reinforcement learning algorithm that combines Q-learning with neural networks. The basic idea behind DQN is to use a neural network to approximate the Q-value function for a given state-action pair, rather than using a table as in traditional Q-learning. The DQN algorithm methodologically involves the following steps:

1. Initialize the Q-network with random weights and a replay memory with a fixed capacity;
2. At each time step, the agent selects an action based on the current state and the Q-network's estimates of the Q-values;
3. The agent then receives a reward and a new state, and the transition (state, action, reward, next state) is stored in the replay memory;
4. Periodically, a batch of transitions is randomly sampled from the replay memory and used to update the Q-network's weights. This is performed by computing the target Q-value for each transition using the Bellman equation and then minimizing the difference between the target Q-value and the Q-network's estimate;
5. The Q-network's weights are also periodically updated with the weights of a fixed target network, which helps to stabilize the learning process.

In the DQN, deep neural networks are used for mapping between input and output data, and it may cause to RL to be more unstable. A lot of data is required to train a deep neural network, but even then, the network weights sometimes cannot converge [27]. To solve this, Mnih et al. (2015) [22] introduced two main techniques, i.e., experience replay and target network separation, into the DQN. When the training data is independent and equally distributed, the reinforcement learning results are good. However, when the reinforcement learning is performed, the series of experience data can be highly correlated to each other. For example, each dynamic response of every time step of the sample structure subjected to an earthquake does not differ much, because the numerical integration time step for time-history analysis is usually less than 0.01 or 0.02 s. The conventional Q-learning method that learns from each of these sequential simulation data can be influenced by the effects of this correlation. In the DQN algorithm, this problem was solved by using a large buffer for storing the past experience of the agent and selecting training data from it, instead of using the most recent experience. Figure 3 shows the concept of this replay buffer (experience buffer) technique. As shown in the figure, past experiences are stored in the replay buffer, and then a subset of these experiences are randomly selected to update the Q-network.

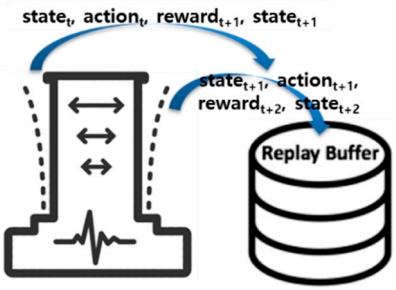


Figure 3. Experience replay of DQN.

In Q-Learning, the Bellman equation is expressed by an artificial neural network (ANN) as shown in Equation (1):

$$Q_\theta(s,a) = r + \gamma \max_{a\prime} Q_\theta(s\prime, a\prime) \tag{1}$$

where the s, a, r, and γ mean the state, action, reward, and discount factor, respectively. The meaning of primed θ variables is the variable's value in the next state. The variable θ of the action-value function $Q_\theta(s,a)$ represents the parameter vectors of ANN. In Q-Learning, the value of $r + \gamma \max_{a\prime} Q_\theta(s\prime, a\prime)$ is considered to be the target Q value. Therefore, a predicted value of ANN $Q_\theta(s,a)$ should gradually reach the target value in the learning process. For this purpose, the loss function of Q-Learning can be expressed by the difference between the target value and the predicted value, as shown in Equation (2). When the learning starts, the parameters θ of ANN for the action-value function $Q_\theta(s,a)$ are updated to minimize the loss function:

$$L(\theta) = \left(r + \gamma \max_{a\prime} Q_\theta(s\prime, a\prime) - Q_\theta(s,a)\right)^2 \tag{2}$$

In native Q-learning, a single ANN with the parameters θ is used for both the predicted value $Q_\theta(s,a)$ and the target value $Q_\theta(s\prime, a\prime)$. Therefore, when an update of ANN parameters θ to make $Q_\theta(s,a)$ closer to the target value is preformed, the target value $Q_\theta(s\prime, a\prime)$ is indirectly changed. It may cause the training to be unstable. To make the training more stable, the DQN uses a target network separation, by keeping a copy of the neural network ($\bar{\theta}$), and using it for the target value $Q_{\bar{\theta}}(s\prime, a\prime)$. In this technique, the parameters of the target network are not changed during the main network training, but they are periodically copied from the parameters of the trained main Q-network, as shown in Figure 4. This idea can improve the stability of Q-learning training [28].

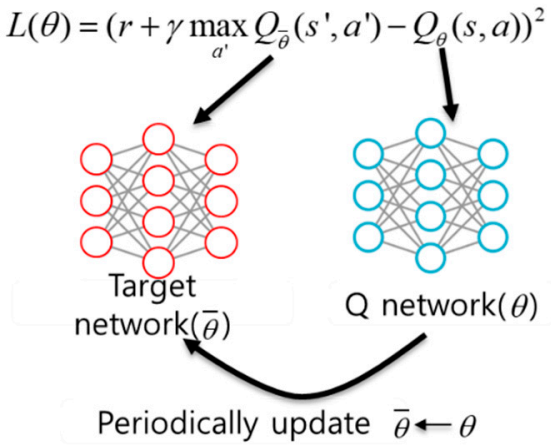


Figure 4. Target network separation.

3. Sample Building Structure and Seismic Excitation

As explained previously, the Shiodome Sumitomo building [20] was used in this study as a sample building structure to construct the environment of reinforcement learning. Figure 5a presents the framing elevation of the 125.9 m tall sample building having 26 stories. A seismic insolation system is installed between the 11th and 12th stories. The nonlinear passive hysteretic dampers in the Shiodome Sumitomo building were changed to MR dampers in the SMIS of the sample structure. Rubber bearings were directly used in the SMIS of the sample building. Figure 5b shows the analytical model of the sample building with the SMIS. One horizontal degree-of-freedom (DOF) per story was used for this analytical model. Table 1 presents the stiffnesses and masses of the sample structure that were adopted from a previous study [20]. The damping matrix was calculated by the conventional methods of Rayleigh damping. The damping ratio of the sample building was set to be 2%, and that of the seismic isolation interface was 0%. The natural periods of vibration of the first three modes were calculated as 6.04, 1.17, and 0.96 s, respectively.

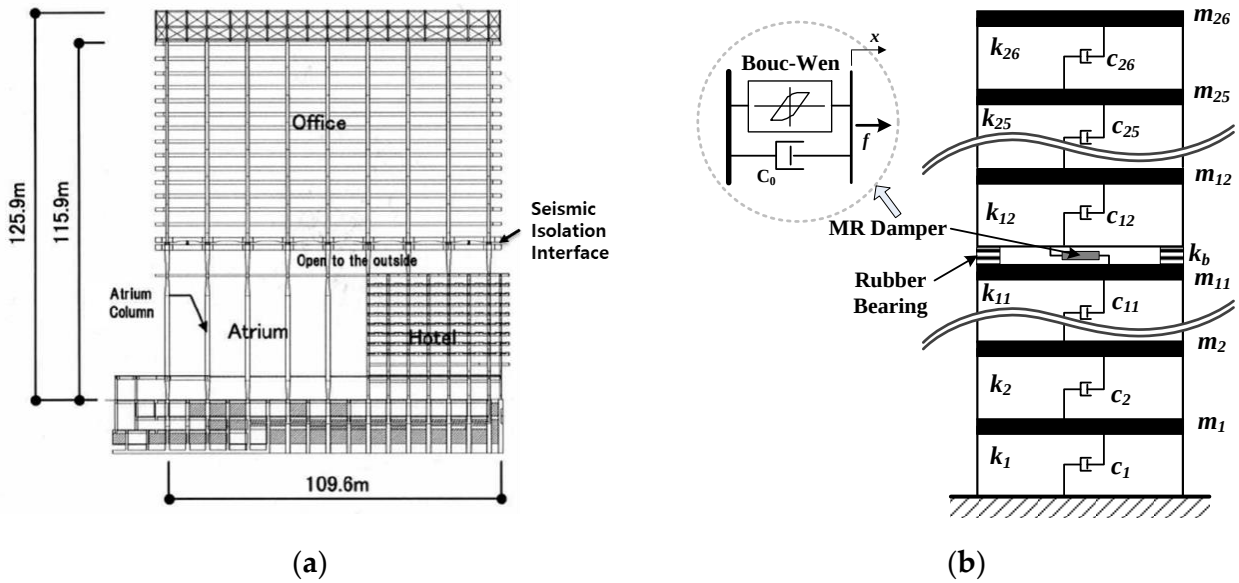


Figure 5. Sample building structure with a mid-story isolation system: (**a**) elevation of the Shiodome Sumitomo building [20], (**b**) analytical model with SMIS.

Table 1. Story mass and stiffness of the sample structure.

Story	Mass (kg)	Stiffness (kN/m)	Story	Mass (kg)	Stiffness (kN/m)
1	3,080,514	3.18×10^6	14	3,186,563	2.32×10^6
2	2,582,900	2.68×10^6	15	3,140,676	2.63×10^6
3	1,726,352	6.34×10^6	16	3,132,518	2.65×10^6
4	1,733,490	5.92×10^6	17	3,125,381	2.59×10^6
5	1,716,155	5.71×10^6	18	3,170,247	2.59×10^6
6	1,715,135	5.36×10^6	19	3,168,208	2.48×10^6
7	1,718,195	5.20×10^6	20	3,117,223	2.49×10^6
8	1,697,801	4.95×10^6	21	3,094,790	2.34×10^6
9	1,721,254	4.79×10^6	22	3,084,593	2.24×10^6
10	3,127,420	4.45×10^6	23	3,076,435	2.17×10^6
11	3,128,440	1.08×10^6	24	3,447,606	2.11×10^6
12	4,030,874	8.07×10^5	25	3,461,882	1.73×10^6
13	3,567,930	3.11×10^6	26	5,769,463	1.51×10^6

The Bouc–Wen model [29] was used to represent the MR damper as shown in Figure 5b. The MR damper force is calculated by the following simultaneous differential equations:

$$f = c_0\dot{x} + \alpha z \tag{3}$$

$$\dot{z} = -\gamma|\dot{x}|z|z|^{n-1} - \beta\dot{x}|z|^n + A\dot{x} \tag{4}$$

where $\dot{x}$ is the velocity of the damper piston, c_0 is the viscous damping of the device, and z is the evolutionary variable that describes the history dependence of the response. The variables γ, β, n, and A define the model shape in the unloading and the smoothness of the hysteretic behavior, representing the transition from the pre-yield to the post-yield region. The parameters α and c_0 in Equation (3) are defined as functions of the applied voltage v as follows:

$$\alpha = \alpha_a + \alpha_b u \tag{5}$$

$$c_0 = c_{0a} + c_{0b} u \tag{6}$$

$$\dot{u} = -\eta(u - v) \tag{7}$$

where u is the control input defined by the first-order filter that accounts for the dynamics of the MR damper; v is the command voltage; and η is the time constant of this first-order filter. The parameters α_a, α_b, c_{0a} and c_{0b} account for the dependence of the damper force on command voltages applied magnetic current. The selected MR damper parameters are as follows: α_a = 5.5319 × 10^5 N/cm, α_b = 2.5246 × 10^6 N/(cmV), c_{0a} = 22.39 Ns/cm, c_{0b} = 223.9 Ns/(cmV), n = 1, A = 1.2, γ = 3 cm^{-1}, β= 3 cm^{-1}, and η = 50 s^{-1}. These parameters were adjusted from the parameter values of an experimental study [30] to make the maximum force of the MR damper about 2750 kN, with a saturation voltage of 5 V. The value of 2750 kN was determined based on a parameter study, and 10 MR dampers were used to make the SMIS of the sample building for optimal seismic control. The control voltage for the MR damper changes within the range of 0 and 5 V. Figure 6 shows typical displacement–force and velocity–force hysteresis loops for this MR damper model. The resisting force of the MR damper is proportional to the velocity and applied command voltage as shown in the figure.

The ground acceleration was modeled as an excitation by a filtered Gaussian process. The Kanai–Tajimi shaping filter is widely used to make an artificial ground motion because it can successfully describe features of seismic ground motions [31]. The Kanai–Tajimi shaping filter [32] shown in Equation (8) was employed to make the artificial earthquake:

$$F(s) = \frac{2\zeta_g\omega_g s + \omega_g^2}{s^2 + 2\zeta_g\omega_g s + \omega_g^2} \tag{8}$$

where the parameters ω_g and ζ_g mean the frequency and damping ratio of the soil, and they are set to be 17 rad/s and 0.3, respectively.

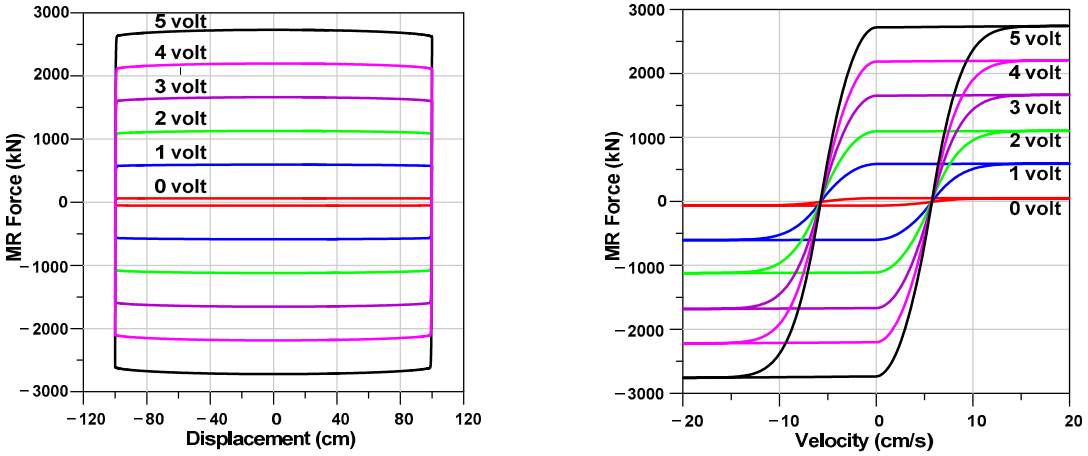


Figure 6. Hysteresis loops of the MR damper: (**a**) displacement—force hysteresis loop, (**b**) velocity–force hysteresis loop.

Gaussian white noise having a time-step of 0.005 s was filtered by using the shaping filter in Equation (8). The PGA of the artificial earthquake was set to be 0.7 g, and the envelope introduced in a previous study [33] was applied to the generated signals for a more practical ground motion. Through this procedure, three artificial earthquakes were developed for training and evaluation of the DQN-based semi-active controller, respectively. Figure 7 shows three developed artificial earthquakes.

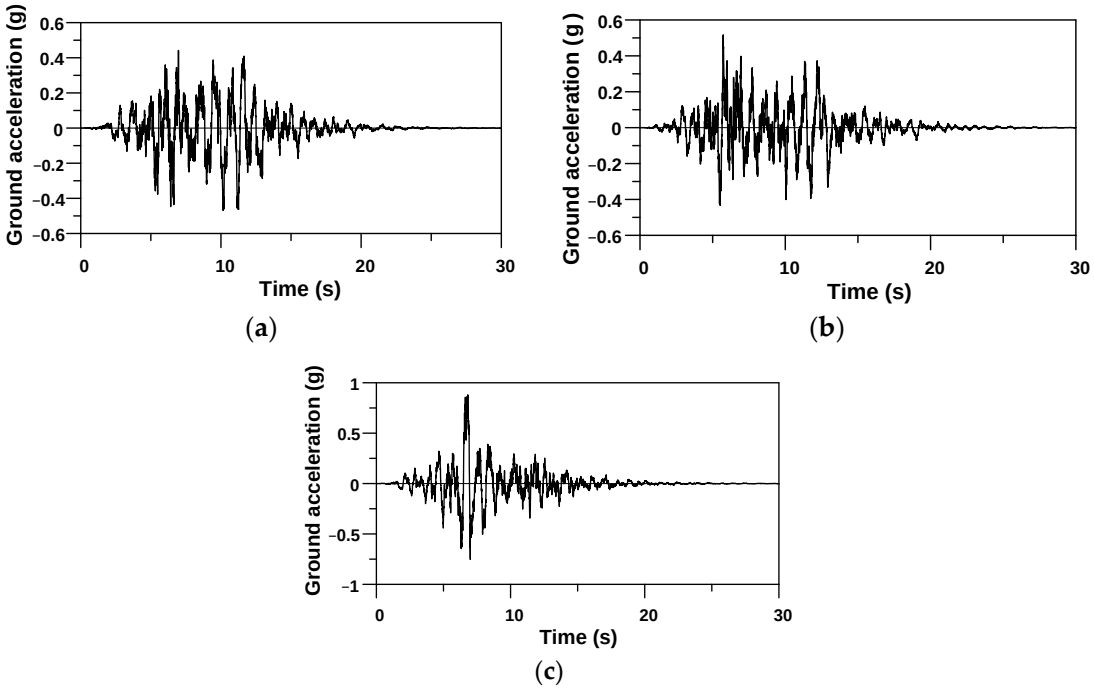


Figure 7. Time histories of artificial ground acceleration. (**a**) Earthquake for training. (**b**) Earthquake for verification1 (EV1). (**c**) Earthquake for verification2 (EV2).

4. Configuration of DQN Agent and Environment

The sample building structure with the SMIS and the artificial ground motion developed in the previous section were used to construct the environment of reinforcement learning. When the time history analyses of the sample structure subjected to the artificial earthquake are performed for numerical simulation in the environment, a lot of dynamic responses of the sample structure can be obtained. Among them, the essential dynamic responses should be selected as the states for the agent to take the appropriate action for optimal semi-active control. In this study, four structural responses and a series of ground accelerations were selected for the states, as shown in Figure 8.

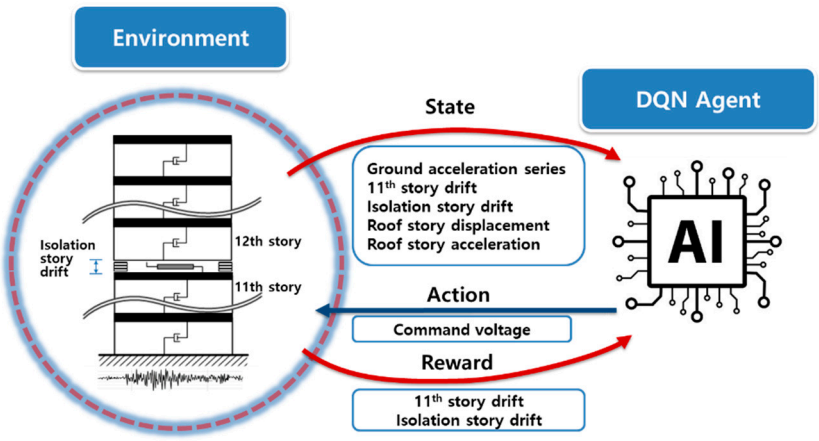


Figure 8. Configuration of the DQN agent and environment.

First, the roof story dynamic responses, i.e., displacement and acceleration that can represent the overall structural behavior of the sample building, were selected. Because the inter-story drift is directly related to the structural damage of the building, the 11th-story drift where the peak story drift occurs and the isolation story drift were also selected as the states to check if the maximum story and isolator drifts exceed their limit or not. Because the ground acceleration is one of the major causes of structural responses, it will give the agent good information to find the appropriate action to decrease the seismic responses of the sample structure. However, one ground acceleration value at a specific step is not enough to predict the seismic responses of the sample structure at that moment. The responses of the specific ground acceleration value are not identical all the time. For example, even if the ground acceleration is 0.2 g, the seismic responses at that moment vary, depending on whether the acceleration is increasing or decreasing. Therefore, not only the ground acceleration value at the current step, but also a series of ground acceleration values of the previous four steps, were selected as the states to give the agent sufficient information to select the optimal control command. As explained previously, the command voltage for the MR damper in the SMIS is determined by one of eleven voltage values, i.e., the voltage range between 0 and 5 V with a step of 0.5 V. Therefore, the agent selects one of the finite actions (command voltage) based on the input state.

When the agent selects the control command for the SMIS, its control forces are transmitted to the sample structure, and the seismic responses are changed. Because the goal of this reinforcement learning is to minimize the seismic responses of the sample structure, the reward should be designed to reflect this goal. The inter-story drift is one of the most critical seismic responses related to building failure. Therefore, the 11th-story drift and the isolation story drift were used to calculate the reward, as shown in Equation (9). Because the agent in reinforcement learning is trained to maximize the reward, the drift responses of the 11th story and the isolation story are expressed as denominators in the

fraction in Equation (9). By squaring these drift responses, the reward decreases quickly as the drift values increase:

$$\text{reward} = 1/\left\{(\text{11th story drift})^2 + (\text{Isolation story drift})^2\right\} \quad (9)$$

In the DQN, the agent is modeled by an artificial neural network. Figure 9 shows that the structure of the agent neural network used for the DQN implementation in this study has two hidden layers. The neural network of the agent maps the input state to the output action. As explained previously, four seismic responses (roof story displacement and acceleration, and the 11th story and the isolation story drifts) and five steps of the ground acceleration are used for the input state, and thus, the input size of the agent is nine. The output size of the agent neural network is eleven, because one hot encoding is used for the agent's action determined by one of eleven voltage values.

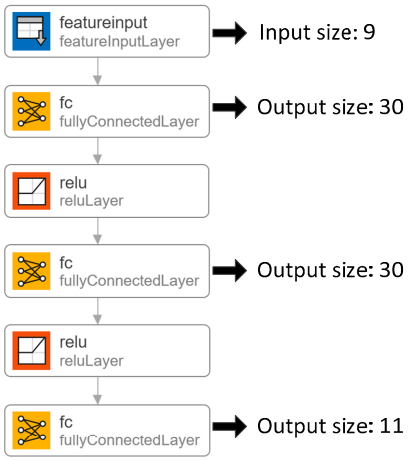


Figure 9. Network configuration of DQN agent.

Control performance of the DQN agent is significantly affected by the number of hidden layers and nodes (neurons) in hidden layers in the agent neural network. Generally, the more hidden layers and neurons are used, the better the control performance of the agent that can be obtained. However, the use of networks that are too complex can induce overfitting, resulting in poor control performance of the agent for unlearned data. The number of hidden layers and neurons in the agent neural network were determined to be two and thirty, respectively, based on a parameter study. A hidden layer usually has an activation function for introduction of non-linearity into the outputs of a network. Therefore, the rectified linear unit (ReLU) activation function was applied to each hidden layer as shown in Figure 9, because it is simple to apply and effective compared to other activation functions [34].

Determination of hyperparameter values and proper function for learning are not easy to develop for the DQN model for the semi-active control algorithm. Table 2 lists the hyperparameter values and functions selected for the DQN training. An Adam (Adaptive Moment Estimation) optimizer with the learning rate of 0.001 was employed as an optimization algorithm. The Adam optimizer is popularly used in the field of deep learning and has shown stable and good performance. Overfitting is a critical issue to be considered with DQN learning. Because the dropout method is known as an effective regularization technique to decrease overfitting and improve deep learning performance, it was used in this study. The dropout rate means the rate of nodes used for network output calculation. A dropout rate of 1.0 implies that there is no dropout node, and it is generally used for the verification process. An appropriate dropout value is known to be in the range between 0.5 and 0.8. The dropout rate of 0.8 was selected in the DQN training.

Table 2. Default hyperparameter values and function.

Item	Value
Learning rate	0.001
Target update frequency	4
Discount factor	0.99
Mini batch size	256
Activation function	ReLU
Optimizer	Adam
Gradient threshold	1
Max. episode	10,000

In reinforcement learning, the agent usually discovers which action provides the most valuable reward by trial and error. When an agent always takes the known action that provides a big reward and does not explore any other actions, i.e., greedy behavior or exploration, the agent cannot improve its knowledge and may fall into sub-optimal results. Therefore, the agent requires enough exploration to find a global optimal solution. However, if the exploration rate is too high, the known high-value actions are not effectively used, and thus, the improvement speed of the agent's performance is decreased in the RL training process. When the agent explores, it gets closer to global optimal solutions, and when it exploits, it could obtain more reward. Because there is a trade-off between exploration and exploitation, they cannot be selected simultaneously [35]. In this study, the epsilon-greedy method was used to balance exploration and exploitation. In the epsilon-greedy method, epsilon means the probability of exploration with random command voltage. Because the epsilon is calculated by Equation (10) in this study, the probability of exploration decreases as the episode increases, resulting in the decreased range of fluctuation of the DQN model control performance.

$$\varepsilon = 1/(episode/10) \tag{10}$$

The Adam optimization algorithm that is a training algorithm for deep neural networks in DQN uses the idea of moving averages of the gradients and squared gradients to adapt the learning rate for each parameter. It also uses bias-correction to ensure that the algorithm starts with a fair estimate of the moving averages. The update rule for the weights is a combination of the gradient with respect to the loss and the bias-corrected moving averages of the gradients and squared gradients. The Adam algorithm uses two different moving averages of the gradients and the squared gradients, denoted as $m(t)$ and $v(t)$, respectively. The moving averages are initialized with 0 and are updated at each iteration by the following formulas:

$$m(t) = \beta_1 \times m(t-1) + (1-\beta_1)\Delta L(t) \tag{11}$$

$$v(t) = \beta_2 \times v(t-1) + (1-\beta_2)\Delta L(t)^2 \tag{12}$$

where L is the loss function; $\Delta L(t)$ is the gradient of the loss function at iteration t; and β_1 and β_2 are hyperparameters that control the decay rate of the moving averages. In order to prevent bias in the initialization, the Adam uses bias-corrected versions of $m(t)$ and $v(t)$ denoted as $\hat{m}(t)$ and $\hat{v}(t)$, respectively:

$$\hat{m}(t) = m(t)/(1-{\beta_1}^t) \tag{13}$$

$$\hat{v}(t) = v(t)/(1-{\beta_2}^t) \tag{14}$$

The update rule for the weights is then defined as follows:

$$\omega(t+1) = \omega(t) - \alpha \times (\hat{m}(t)/\sqrt{\hat{v}(t)+\varepsilon}) \tag{15}$$

where $\omega(t)$ is the weight at iteration t, α is the learning rate, and ε is a small value added to the denominator to prevent division by zero.

The training Q-network and the target Q-network provide eleven Q-values corresponding to eleven command voltages, respectively. The DQN agent selects the command voltage corresponding to the maximum Q-values to control the MR damper. The purpose of DQN learning is to make the Q-values of the target network and those of the training network the same. To this end, the training Q-network's weights are updated by minimizing the mean-squared error (MSE) loss between the predicted Q-values and the target Q-values as presented in Figure 10. This minimization process is conducted based on the Adam optimization algorithm described above. The variation of the MSE loss value according to the episode is shown in Figure 11. The flat part of the graph showing the early episodes means Not a Number (NaN), representing an undefined number in floating-point operations. Figure 11 shows that the MSE loss value decreased rapidly up to 1000 episodes and then gradually decreased to 8000 episodes. After that, the MSE loss value remained almost unchanged. This variation tendency of the MSE loss value indicates that the network training was well done.

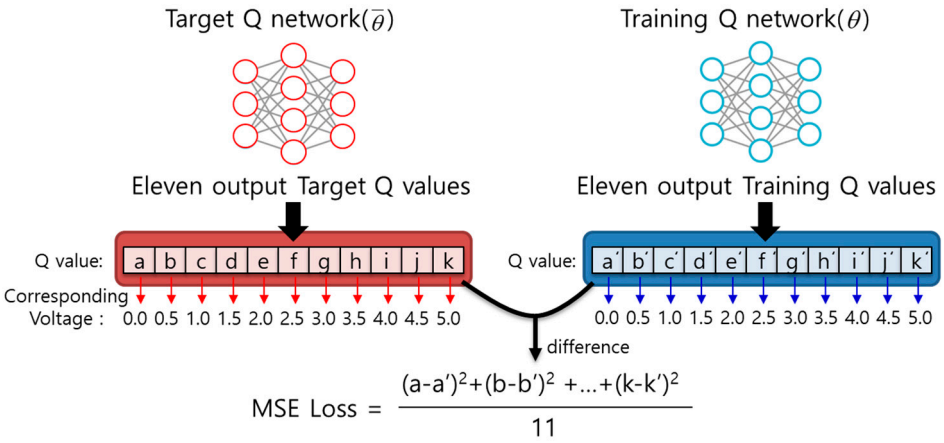


Figure 10. Calculation of MSE loss.

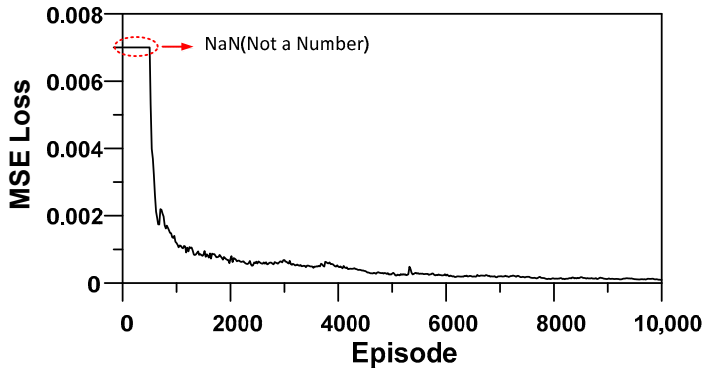


Figure 11. History of MSE loss value.

5. Control Performance Evaluation of the DQN Model

Comparative control algorithms were used to verify the effectiveness of the DQN model (controller). Skyhook and groundhook control algorithms were selected as comparative control algorithms, because they have typically been used for semi-active control problems and have shown good performance. A skyhook controller provides effective performance for dynamic-response reduction of the mass isolated from the vibrating base, such as in a semi-active base isolation system. A groundhook controller effectively decreases dynamic responses of the structure with the auxiliary mass, such as with a semi-active tuned mass damper [36]. Figure 12 shows the ideal configurations of the skyhook and

groundhook controllers. These conceptual arrangement of an ideal skyhook and groundhook cannot be constructed in practice, because the damper cannot be connected to the sky or an absolute fixed base. The purpose of skyhook or groundhook controllers is to imitate the ideal structural configuration in Figure 12 using the practical SMIS shown in Figure 5b. Koo et al. showed that the displacement-based on–off controller can provide the best performance for the benchmark problem among various versions of the skyhook and groundhook control algorithms [37].

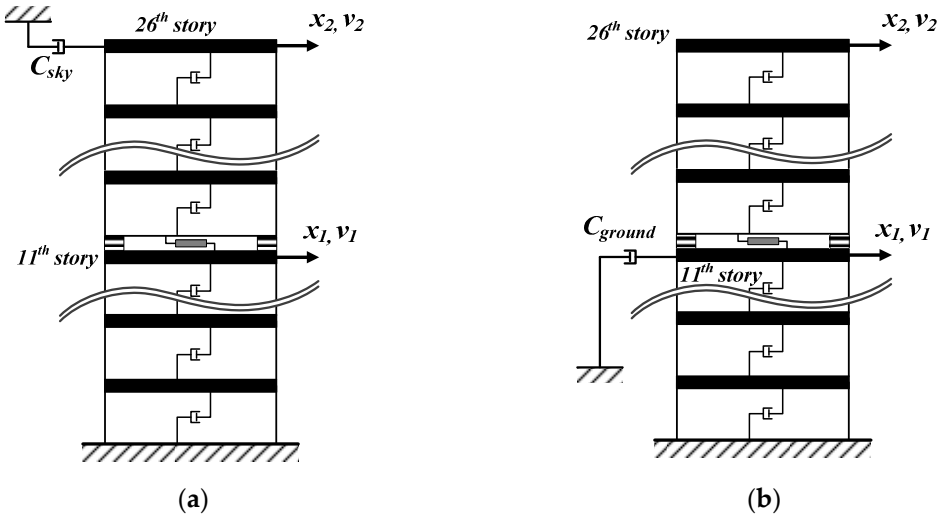


Figure 12. Idealized semi-active control configurations: (**a**) skyhook, (**b**) groundhook.

The relative velocity between the roof story and the 11th story, i.e., $v_1 - v_2$ in Figure 12, is an important value for the development of the skyhook and groundhook controllers. The displacement-based skyhook algorithm is formulated by the relative velocity and the absolute displacement of the upper isolated structure (x_2). The on–off skyhook control algorithm selects the minimum or the maximum command voltage. This calculation is performed by using Equation (16):

$$V = \begin{cases} V_{\max} \text{ if } x_2(v_1 - v_2) \leq 0 \\ V_{\min} \text{ if } x_2(v_1 - v_2) > 0 \end{cases} \quad (16)$$

where V is the control command voltage; $V_{\max}$ is the maximum voltage, 5 V; and $V_{\min}$ is the minimum voltage, 0 V. The displacement-based on–off groundhook control algorithm can be obtained by changing x_2 to x_1 in the skyhook controller. The groundhook controller can be expressed by:

$$V = \begin{cases} V_{\max} & \text{if } x_1(v_1 - v_2) \geq 0 \\ V_{\min} & \text{if } x_1(v_1 - v_2) < 0 \end{cases} \quad (17)$$

After 10,000 episodes of reinforcement learning training, the DQN model was obtained to control the SMIS. Python 3.5.0 was employed to program the DQN model and environment generation codes, with Tensorflow 1.6.0 as a machine learning library. Time history analyses of the sample building structure with the SMIS were performed to investigate the seismic response control performance of the DQN model. An artificial ground acceleration not used for training was used for dynamic time history analysis.

Figures 13–16 compare the sample structure's seismic response time histories using the SMIS controlled by the DQN model versus the skyhook and groundhook control algorithms. Figure 13 shows the inter-story drift time histories of three control cases at the 11th story, where the peak story drift occurs. In this figure, the 11th-story drift controlled by the groundhook controller is generally greater than those of the other controllers. The DQN model developed in this study provided much better control performance, in comparison with the two typical semi-active controllers. In particular, the 11th-story drift between

about 7 and 10 s, when the response amplitude is significant for both the EV1 and EV2 earthquakes, was effectively reduced by the DQN model. Figure 14 presents isolation story drift time histories of the three controllers. Two comparative semi-active control algorithms present contradictory performance in the decrease of the isolation story drift versus the 11th-story drift. The isolation story drift controlled by the skyhook controller is the greatest among the three controllers, and the peak response occurs at around 8 s for both the EV1 and EV2 earthquakes. The peak isolation story drift of the DQN model is smaller than those of the two comparative controllers, as in the case of the previous 11th-story drift.

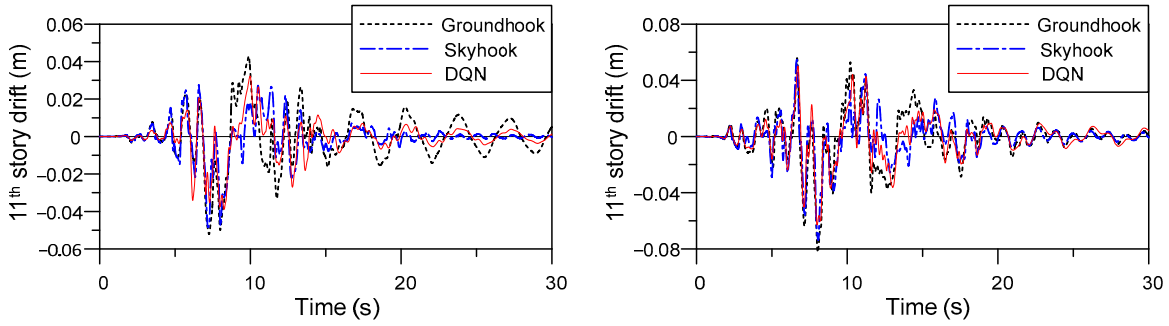


Figure 13. Comparison of the 11th-story drift time histories: (**a**) responses to EV1, (**b**) responses to EV2.

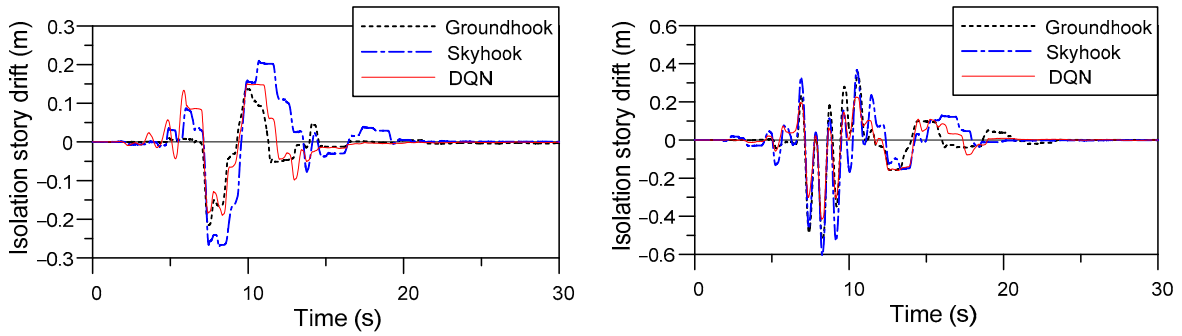


Figure 14. Comparison of the isolation story drift time histories: (**a**) responses to EV1, (**b**) responses to EV2.

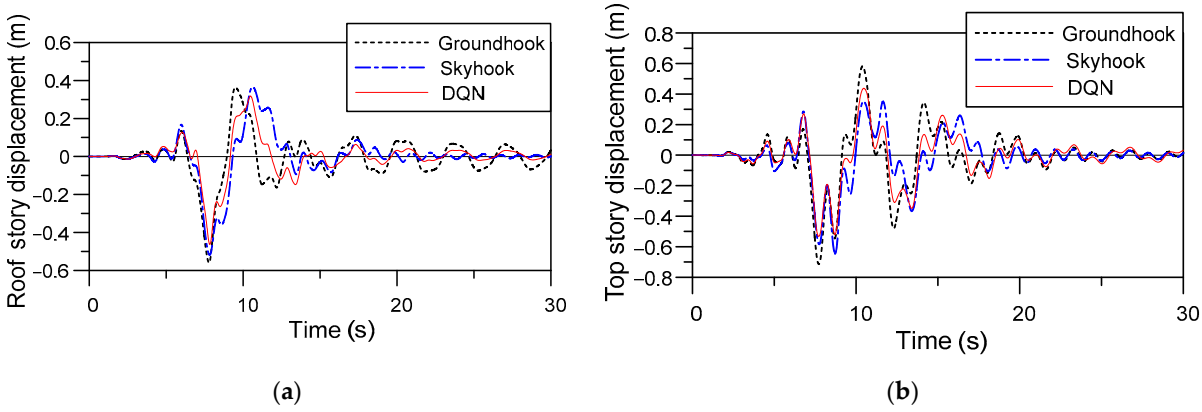
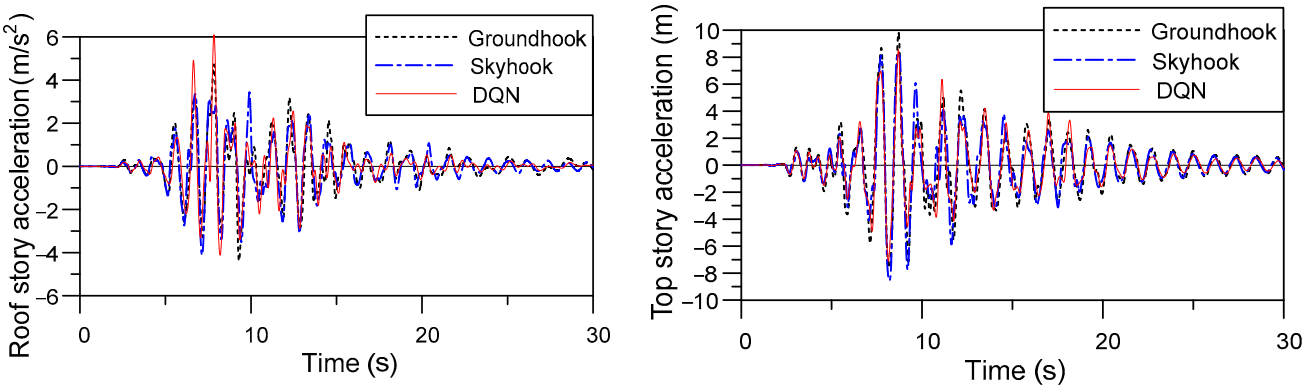


Figure 15. Comparison of the roof story displacement time histories: (**a**) responses to EV1, (**b**) responses to EV2.



(**a**) (**b**)

Figure 16. Comparison of the roof story acceleration time histories: (**a**) responses to EV1, (**b**) responses to EV2.

Figures 15 and 16 compare the roof story displacement and acceleration time histories of the three controllers. It seems that the peak roof story displacement of the groundhook controller is similar to that of the skyhook controller for EV1, but that of the groundhook controller is larger than that of the skyhook controller for EV2. The DQN model can effectively reduce roof story displacement, compared to the two comparative controllers for both EV1 and EV2 cases. In contrast to the displacement control performance, the peak roof story acceleration controlled by the DQN model is greater than that of the two comparative controllers for EV1. However, the DQN model controlled the peak roof story acceleration better than that of the two comparative controllers in case of EV2. The skyhook controller can effectively reduce the peak roof story acceleration compared to the groundhook controller. The control performance of the DQN model for seismic responses of the roof story is not consistent, compared to that for the story drift response reduction. This may be because the reward equation of the DQN model includes not the roof story responses but the story drift responses. In multi-purpose optimization, the optimization efficiency of each purpose generally decreases as the number of purposes increases. Therefore, if too many structural responses are included in the reward equation, the control performance of the DQN model may be disrupted.

To quantitatively compare the control performance of each algorithm, Tables 3 and 4 list their peak responses. "Passive", in the control algorithm item, means the seismic responses obtained from the original Shiodome Sumitomo building with a passive mid-story isolation system. "Passive-On" or "Passive-Off" means the command voltage applied to the MR damper is kept constant at a maximum (5 V) or a minimum voltage (0 V), respectively. The ratio of each controlled response to the response of the original passive mid-story isolation system is presented in the corresponding parentheses.

Table 3. Comparison of peak responses according to control algorithms (EV1).

Control Algorithm	Peak 11th-Story Drift (m)	Peak Isolation Story Drift (m)	Peak Roof Displacement (m)	Peak Roof Acceleration (m/s^2)
Passive [1]	0.0554 (1.00) [2]	0.2754 (1.00)	0.6289 (1.00)	5.4558 (1.00)
Groundhook	0.0521 (0.94)	0.2152 (0.78)	0.5598 (0.89)	4.7248 (0.87)
Skyhook	0.0485 (0.88)	0.2719 (0.99)	0.5189 (0.83)	4.1149 (0.75)
Passive-On	0.0534 (0.96)	0.2183 (0.79)	0.5525 (0.88)	4.7144 (0.86)
Passive-Off	0.0467 (0.84)	0.4542 (1.65)	0.6604 (1.05)	5.7801 (1.06)
DQN	0.0387 (0.70)	0.1898 (0.69)	0.4597 (0.73)	6.1052 (1.12)

[1] Original passive mid-story isolation system in Shiodome Sumitomo building. [2] Ratio of semi-active controlled responses to responses of passive mid-story isolation system.

Table 4. Comparison of peak responses according to control algorithms (EV2).

Control Algorithm	Peak S (m)	Peak Isolation Story Drift (m)	Peak Roof Displacement (m)	Peak Roof Acceleration (m/s^2)
Passive	0.0861 (1.00)	0.6279 (1.00)	0.7945 (1.00)	11.2183 (1.00)
Groundhook	0.0826 (0.96)	0.5184 (0.83)	0.7205 (0.91)	9.7622 (0.87)
Skyhook	0.0741 (0.86)	0.5995 (0.95)	0.6332 (0.80)	8.5115 (0.76)
Passive-On	0.0837 (0.97)	0.4872 (0.78)	0.7166 (0.90)	9.404 (0.84)
Passive-Off	0.0645 (0.75)	0.8256 (1.31)	0.7430 (0.94)	12.3944 (1.10)
DQN	0.0630 (0.73)	0.4162 (0.66)	0.5331 (0.67)	8.5569 (0.76)

It is evident that most of the semi-active control cases provide better control performance than the original passive mid-story isolation system. A groundhook controller can reduce the peak isolation drift by 22% in comparison with the reference original passive model for EV1, while decreasing the peak 11th-story drift by 6%. On the other hand, a skyhook controller reduces the peak 11th-story drift by 12% in comparison with the original passive, maintaining the peak isolation story drift close to the original passive. A skyhook shows effective performance for the decrease of roof story responses in comparison with the other control cases, because it emulates the ideal configuration of a passive damper connected between the roof story of the structure and the sky, as shown in Figure 12a. Since the passive-on case provides the maximum MR damper force to the isolation story, the peak isolation-story drift is well controlled, compared to the other responses. In the passive-off case, the peak isolation-story drift significantly increases, because the minimum MR damper force is applied to the SMIS. However, the peak 11th-story drift of the passive-off case is controlled better than in the other control cases, except for the DQN case. This is because the minimum MR damper force in the SMIS allows the upper isolated structure to move like a tuned mass damper (TMD). The isolated upper story structure's moving like an auxiliary mass results in the largest roof story responses, compared to all other control cases. Tables 3 and 4 show that the DQN model can reduce both the peak 11th-story and isolation story drifts by about 30 %, compared to the original passive case. This shows that the DQN model is the most effective control algorithm among the five control cases. Tables 3 and 4 reveal that the reduction of the isolation story drift is in conflict with the decrease of the 11th-story drift. That is, a reduction of the isolation story drift cannot be accomplished without an increase of the 11th-story drift, whereas the 11th-story drift can be decreased with an increment of the isolation story drift. Tables 3 and 4 show that the DQN control algorithm can effectively reduce both the peak 11th story and isolation story drifts that are in conflict, by including them in the reward equation.

Figure 17 shows the peak inter-story drifts of each control case along the story. The isolation story drifts of each model are excluded in this figure, because they are much larger than the other inter-story drifts, but they are quantitatively compared in Tables 3 and 4. The largest inter-story drift can be seen at the 11th story from every control algorithm in Figure 17, because the 11th story stiffness of the Shiodome Sumitomo building is the smallest. The DQN model provides good control performance not only for the 11th-story drift, but also for all the story drifts, compared to the other control algorithms. The control effectiveness of the DQN model is much better than that of the comparative controllers, especially in the lower structure. Compared to the other control cases, both the DQN and skyhook controllers can effectively decrease the inter-story drifts in the upper structure.

The action of the DQN agent is promptly determined based on the states, including the ground motion series and the dynamic responses of the sample structure. In this study, the action of the DQN agent is interpreted as the command voltage to control the MR damper. Figure 18 presents the command voltage time histories of the DQN and two comparative control algorithms for the artificial earthquake of EV1. It can be seen that the command voltage generated from the groundhook or skyhook algorithm is similar to the output of a simple on–off control algorithm, which provides 0 or 5 V. However, the command voltage generated from the DQN algorithm varies irregularly between 0 and 5 V with a step of 0.5 V.

When the isolation story drift is too big, the DQN agent increases the command voltage resulting in an increase of the MR damper force to reduce the isolation story drift and the movement of the isolated upper structure. On the other hand, when the 11th-story drift increases, the DQN agent decreases the command voltage, resulting in a decrease of the MR damper force to make the isolated upper structure move like a mass damper. Because the DQN algorithm generates the control command minutely, its control performance is better than that of the comparative algorithms. The MR damper stroke–force relationships of the three control algorithms are presented as hysteretic loops in Figure 19. The area enclosed by the hysteretic loop is considered to be the amount of seismic energy dissipated in the SMIS. Figure 19 shows that the MR damper in the SMIS behaves properly to dissipate the seismic energy. It can be seen that the hysteretic loops of the groundhook and skyhook algorithms are somewhat symmetrical about the vertical axis because of their calculation logic. Meanwhile, the hysteretic loop of the DQN algorithm shows an atypical and very irregular shape, because the MR damper force instantly varies based on the states of the RL environment, including the ground motion and seismic responses of the sample structure to make an optimal control force.

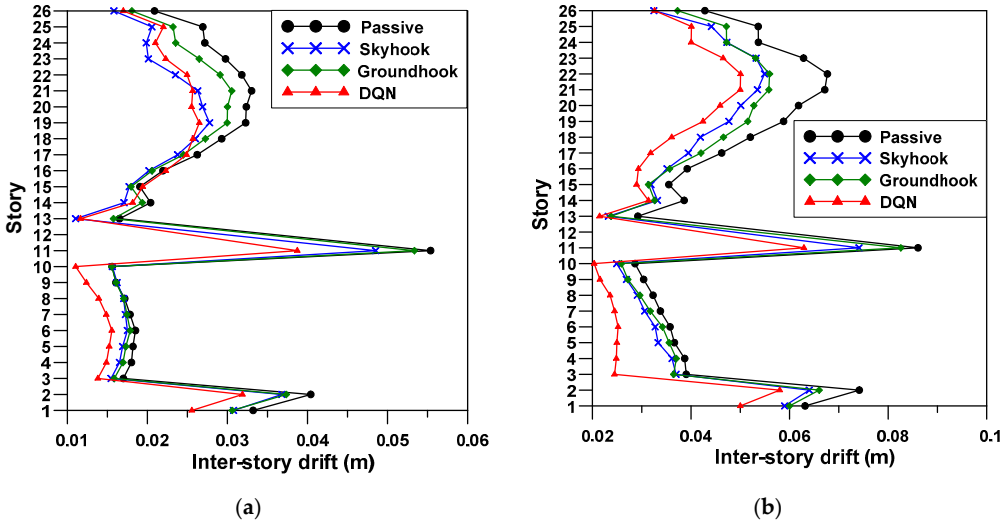


Figure 17. Peak inter-story drift of each control algorithm: (**a**) responses to EV1, (**b**) responses to EV2.

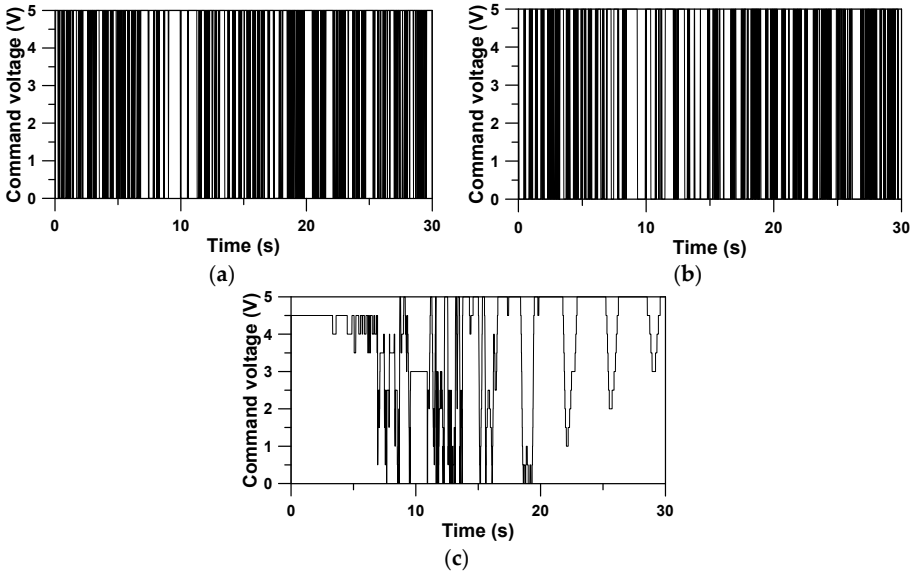


Figure 18. Time histories of command voltage (EV1): (**a**) groundhook algorithm, (**b**) skyhook algorithm, (**c**) DQN algorithm.

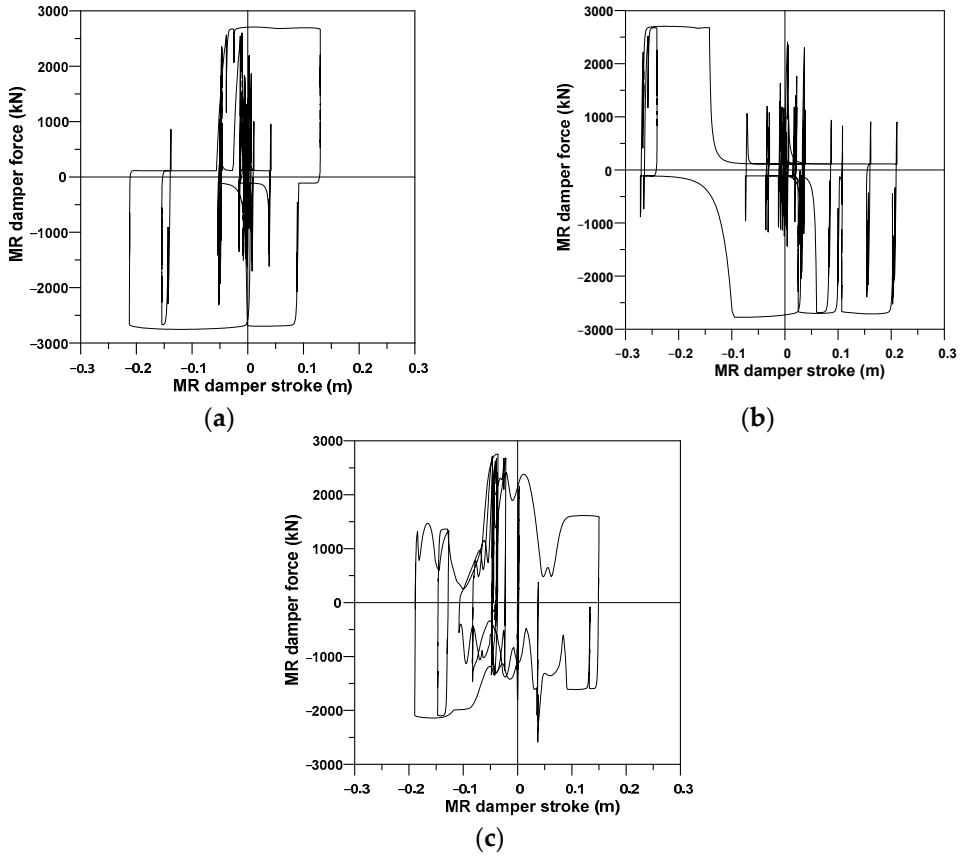


Figure 19. MR damper stroke-and-force relationship (EV1): (**a**) groundhook algorithm, (**b**) skyhook algorithm, (**c**) DQN algorithm.

6. Conclusions

A reinforcement-learning-based semi-active control algorithm was developed in this study. The DQN was selected from among reinforcement learning methods for semi-active control, because it has shown many successful applications to various control engineering fields. The seismic response control performance of the DQN for the SMIS was investigated with comparative algorithms. For more practical study, the structural properties of the existing Shiodome Sumitomo building and the mid-story isolation system were directly used in this study. The sample SMIS is composed of MR dampers replacing nonlinear passive dampers in the existing building. The DQN reward was designed to reduce the isolation story and the peak inter-story (11th story) drifts that are in conflict. The control performance of the DQN controller was compared with typical semi-active control algorithms, i.e., skyhook, groundhook, passive-on, and passive-off. Seismic responses of the original passive mid-story isolation system were used as the reference values to evaluate the control effectiveness of the SMIS controlled by five semi-active control cases. Two artificial ground motions were generated as seismic loads for training and evaluation of the DQN-based semi-active controller.

Numerical analyses show that groundhook and skyhook controllers show good reduction performance of the isolation story and the 11th-story drifts, respectively, in accordance with their design concept. Passive-on and passive-off control cases also effectively reduce the isolation story and the 11th-story drifts, respectively, like groundhook and skyhook controllers. Because reduction of the isolation story drift is in conflict with decrease of the 11th-story drift, these two structural responses cannot be easily simultaneously controlled by typical semi-active control policies. However, the DQN controller can effectively reduce both the 11th story and isolation story drifts, compared to typical semi-active control algorithms. This means that the DQN network can successfully map the input state (seismic

responses and ground motions) to the output action (control command), resulting in effective semi-active control. In this study, limited artificial ground motions were used for training and evaluation of the DQN controller. Accordingly, to increase the reliability of the semi-active control algorithm developed by the DQN method, other artificial or historical earthquake loads will be required for performance evaluation. Because the reward in the reinforcement learning method mainly affects the control performance of the DQN controller, future research will investigate the effects of various reward designs. Although the semi-active structural control system controlled by the DQN algorithm provides better control performance than the passive control system, the cost of the semi-active system is higher than that of the passive system. Whether the high cost of the semi-active system is reasonable depends on the project circumstances; thus, it is necessary to investigate the relationship between structural performance and economic feasibility in practical applications.

Author Contributions: Conceptualization, H.-S.K. and U.K.; methodology, U.K.; software, H.-S.K.; validation, H.-S.K. and U.K.; writing—original draft preparation, H.-S.K.; writing—review and editing, H.-S.K. and U.K.; visualization, U.K.; funding acquisition, H.-S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by a National Research Foundation of Korea (NRF) grant, funded by the Korean government (MEST), grant number NRF-2019R1A2C1002385.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in our study are available on request from the corresponding author.

Acknowledgments: The authors express sincere gratitude to Department of Civil and Environmental Engineering of California State University, Fullerton for supporting sabbatical leave of H.-S.K.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bagherkhani, A.; Baghlani, A.E. Reliability assessment of MR fluid dampers in passive and semi-active seismic control of structures. *Probabilistic Eng. Mech.* **2021**, *63*, 103114. [CrossRef]
2. Karami, K.; Manie, S.; Ghafouri, K.; Nagarajaiah, S. Nonlinear structural control using integrated DDA/ISMP and semi-active tuned mass damper. *Eng. Struct.* **2019**, *181*, 589–604. [CrossRef]
3. Soto, M.G.; Adeli, H. Semi-active vibration control of smart isolated highway bridge structures using replicator dynamics. *Eng. Struct.* **2019**, *186*, 536–552. [CrossRef]
4. Shih, M.H.; Sung, W.P. Seismic resistance and parametric study of building under control of impulsive semi-active mass damper. *Appl. Sci.* **2021**, *11*, 2468. [CrossRef]
5. Wang, L.; Nagarajaiah, S.; Shi, W.; Zhou, Y. Seismic performance improvement of base-isolated structures using a semi-active tuned mass damper. *Eng. Struct.* **2022**, *271*, 114963. [CrossRef]
6. Rayegani, A.; Nouri, G. Seismic collapse probability and life cycle cost assessment of isolated structures subjected to pounding with smart hybrid isolation system using a modified fuzzy based controller. *Structures* **2022**, *44*, 30–41. [CrossRef]
7. Mohebbi, M.; Dadkhah, H. An adaptive control algorithm for smart base isolation systems based on seismic early warning system. *Structures* **2021**, *30*, 638–646. [CrossRef]
8. Bharathi, P.C.; Gopalakrishnan, N. Emotional learning based adaptive control algorithm for semi-active seismic control of structures. *Mater. Today Proc.* **2022**, *65*, 1703–1710.
9. Jansen, L.M.; Dyke, S.J. Semi-active control strategies for MR dampers: A comparative study. *J. Eng. Mech.* **2000**, *126*, 795–803.
10. Oliveira, F.; Morais, P.; Suleman, A. A comparative study of semi-active control strategies for base isolated buildings. *Earthq. Eng. Eng. Vib.* **2015**, *14*, 487–502. [CrossRef]
11. Bitaraf, M.; Ozbulut, O.E.; Hurlebaus, S.; Barroso, L. Application of semi-active control strategies for seismic protection of buildings with MR dampers. *Eng. Struct.* **2010**, *32*, 3040–3047. [CrossRef]
12. Buşoniu, L.; Bruin, T.D.; Tolić, D.; Kober, J.; Alunko, I. Reinforcement learning for control: Performance, stability, and deep approximators. *Annu. Rev. Control* **2018**, *46*, 8–28. [CrossRef]
13. Howell, M.N.; Frost, G.P.; Gordon, T.J.; Wu, H.Q. Continuous action reinforcement learning applied to vehicle suspension control. *Mechatronics* **1997**, *7*, 263–276. [CrossRef]
14. Adam, B.; Smith, F.C. Reinforcement learning for structural control. *J. Comput. Civ. Eng. ASCE* **2006**, *22*, 133–139. [CrossRef]
15. Kober, J.; Bagnell, J.A.; Peters, J. Reinforcement learning in robotics: A survey. *Int. J. Robot. Res.* **2013**, *32*, 1238–1274. [CrossRef]

16. Eshkevari, S.S.; Eshkevari, S.S.; Sen, D.; Pakzad, S.N. RL-Controller: A reinforcement learning framework for active structural control. *arXiv* **2021**, arXiv:2103.07616.
17. Khalatbarisoltani, A.; Soleymani, M.; Khodadadi, M. Online control of an active seismic system via reinforcement learning. *Struct. Control Health Monit.* **2019**, *26*, e2298. [CrossRef]
18. Lee, D.; Jin, S.; Lee, C. Deep reinforcement learning of semi-active suspension controller for vehicle ride comfort. *IEEE Trans. Veh. Technol.* **2022**, *72*, 327–339. [CrossRef]
19. Ming, L.; Yibin, L.; Xuewen, R.; Shuaishuai, Z.; Yanfang, Y. Semi-active suspension control based on deep reinforcement learning. *IEEE Access* **2020**, *8*, 9978–9986. [CrossRef]
20. Sueoka, T.; Torii, S.; Tsuneki, Y. The application of response control design using middle-story isolation system to high-rise building. In Proceedings of the 13th World Conference on Earthquake Engineering, Vancouver, BC, Canada, 1–6 August 2004.
21. Kim, H.S.; Roschke, P.N. Design of fuzzy logic controller for smart base isolation system using genetic algorithm. *Eng. Struct.* **2006**, *28*, 84–96. [CrossRef]
22. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Rusu, A.A.; Veness, J.; Bellemare, M.G.; Graves, A.; Riedmiller, M.; Fidjeland, A.K.; Ostrovski, G.; et al. Human-level control through deep reinforcement learning. *Nature* **2015**, *518*, 529–533. [CrossRef] [PubMed]
23. Introduction to RL and Deep Q Networks. Available online: https://www.tensorflow.org/agents/tutorials/0_intro_rl (accessed on 28 September 2022).
24. Liu, C.; Chen, L.; Yang, X.; Zhang, X.; Yang, Y. General theory of skyhook control and its application to semi-active suspension control strategy design. *IEEE Access* **2019**, *7*, 101552–101560. [CrossRef]
25. Deep Learning vs. Machine Learning—What's the Difference? Available online: https://levity.ai/blog/difference-machine-learning-deep-learning/ (accessed on 11 November 2022).
26. Different Types of Learning in Machine Learning. Available online: https://www.machinelearningmastery.com/types-of-learning-in-machine-learning/ (accessed on 11 November 2022).
27. Deep Q-Network (DQN)-II. Available online: https://www.towardsdatascience.com/deep-q-network-dqn-ii-b6bf911b6b2c/ (accessed on 30 November 2022).
28. How does the Bellman equation work in Deep RL? Available online: https://towardsdatascience.com/how-the-bellman-equation-works-in-deep-reinforcement-learning-5301fe41b25a/ (accessed on 30 November 2022).
29. Sues, R.H.; Mau, S.T.; Wen, Y.K. System identification of degrading hysteretic restoring forces. *J. Eng. Mech.* **1988**, *114*, 833–846. [CrossRef]
30. Yi, F.; Dyke, S.J.; Caicedo, J.M.; Carlson, J.D. Experimental verification of multi-input seismic control strategies for smart dampers. *J. Eng. Mech.* **2001**, *127*, 1152–1164.
31. Alotta, G.; Paola, M.D.; Pirrotta, A. Fractional Tajimi–Kanai model for simulating earthquake ground motion. *Bull Earthq. Eng.* **2014**, *12*, 2495–2506. [CrossRef]
32. Ramallo, J.C.; Johnson, E.A.; Spencer, B.F. "Smart" base isolation systems. *J. Eng. Mech.* **2002**, *128*, 1088–1100. [CrossRef]
33. Kim, H.S.; Roschke, P.N. GA-fuzzy control of smart base isolated benchmark building using supervisory control technique. *Adv. Eng. Softw.* **2007**, *38*, 453–465. [CrossRef]
34. How to Choose an Activation Function for Deep Learning. Available online: https://machinelearningmastery.com/choose-an-activation-function-for-deep-learning/ (accessed on 7 December 2022).
35. Epsilon-Greedy Algorithm in Reinforcement Learning. Available online: https://www.geeksforgeeks.org/epsilon-greedy-algorithm-in-reinforcement-learning/ (accessed on 7 December 2022).
36. Setareh, M. Use of semi-active tuned mass dampers for vibration control of force-excited structures. *Struct. Eng. Mech.* **2001**, *11*, 341–356. [CrossRef]
37. Koo, J.H.; Setareh, M.; Murray, T.M. In search of suitable control methods for semi-active tuned vibration absorbers. *J. Vib. Control* **2004**, *10*, 163–174. [CrossRef]