Hindawi
Structural Control and Health Monitoring
Volume 2023, Article ID 6770137, 15 pages
https://doi.org/10.1155/2023/6770137

Research Article

Novel Model-free Optimal Active Vibration Control Strategy Based on Deep Reinforcement Learning

Yi-Ang Zhang [1] **and Songye Zhu** [1,2]

[1]*Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China*
[2]*Research Institute for Artificial Intelligence of Things, Hong Kong Branch of National Rail Transit Electrification and Automation Engineering Technology Research Center, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China*

Correspondence should be addressed to Songye Zhu; songye.zhu@polyu.edu.hk

Received 29 September 2022; Revised 14 January 2023; Accepted 18 January 2023; Published 8 February 2023

Academic Editor: Yoshiki Ikeda

Neural networks (NNs) can provide a simple solution to complex structural vibration control problems. However, most past NN-based control strategies cannot guarantee an optimal policy in structural vibration control. In this study, a novel active vibration control strategy based on deep reinforcement learning is proposed, which utilizes the learning ability of NN controllers and simultaneously provides control performance comparable to traditional model-based optimal controllers. The proposed learning algorithm can determine the control policy through interaction with the environment without knowing dynamic system models. This study shows that the proposed model-free strategy can provide optimal control performance to various systems and excitations. The proposed control strategy is first verified on a single-degree-of-freedom model and subsequently extended to a multi-degree-of-freedom shear-building model. Its control performance with full-state feedback is nearly the same as that of a classical linear quadratic regulator. Moreover, the learned policy can outperform a traditional output feedback controller in a partially observed system. The robustness of the proposed control strategy against measurement noise is also tested.

1. Introduction

Structural vibration control aims to suppress the vibrations of civil and mechanical structures induced by dynamic loads. A variety of structural control systems have been proposed and built to alleviate structural responses under various dynamic loads (such as seismic and wind loads), particularly when original structural resistance is insufficient [1, 2]. In general, these systems can be classified into passive, semiactive, and active types [3]. Although passive and semiactive systems are cost-effective and reliable in operation, their performance is limited because they cannot adapt to different excitations [4]. Active control systems, such as active mass dampers [5], active tendon systems [6, 7], and active brace systems [8], can provide high-performance structural response reduction by calculating the desirable actuator control force according to real-time observations.

In the past five decades, various active control algorithms and strategies have been proposed to determine precise control forces with sensor measurements [9], such as linear quadratic regulator (LQR) [10], pole assignment [11], and sliding mode control [12]. However, most optimal control algorithms necessitate complete system dynamics knowledge [13], and their control performance is primarily dependent on the accuracy of model parameters. For example, the design of an LQR controller requires solving the algebraic Riccati equation (ARE) by using the state-space representation of a system [14, 15].

In contrast to these model-based algorithms, a neural network (NN) controller offers a model-free control solution that is more versatile in design and more practical in applications [16, 17]. NN controllers can eliminate the need for analytically developing control algorithms, which is often difficult for structures with unknown dynamics and strong

nonlinearity [18]. The early applications of NN to active vibration control started in the 1990s [19–21] when an emulator network was used to learn structural behavior and train a controller network by minimizing different types of error functions to achieve the desired response. Examples include the zero-one-step response [22] and the zero-average expected future response [23]. Although these methods can reduce structural response under different loadings, their control performance is mostly not comparable with model-based optimal control methods because the error functions used for training can only represent simple control schemes. Kim et al. [24] studied instantaneous optimal control by using the quadratic cost function to train an NN controller for a single-degree-of-freedom (SDOF) system under the El Centro earthquake, in which structural displacement, velocity, and ground acceleration were input into the network. Nevertheless, the performance of this controller was still not comparable with the optimal LQR algorithm that optimizes the same quadratic cost function.

Previous studies on finding model-free optimal control strategies have been mostly conducted on solving ARE through reinforcement learning (RL). This data-driven technique directly finds the optimal policy from its own experience with the environment and does not require the knowledge of full system dynamics. Vrabie et al. [25] used integral RL (IRL) to update the control policy for a full-state feedback system. The policy iteration of IRL is achieved by solving the least-squares problem while integrating the target function along the state trajectory. The same approach has also been applied to output feedback systems [26] and tracking control problems [27, 28]. However, this method is unsuitable for high-dimensional states because it calculates the integrated quadratic cost for every step. Vibration problems in the civil engineering field are typically high-dimensional considering the observation states, adding difficulty to traditional policy iteration, and value iteration algorithms, such as step-by-step least-squares calculation. Hence, conventional RL algorithms are unsuitable for such problems. Consequently, an urgent need arises to develop a novel approach that can utilize the learning ability of NN controllers while attaining performance close to or even better than classical optimal control methods.

The development of deep learning in the past decade has dramatically enhanced the ability of RL by involving deep NNs, known as deep RL (DRL). A well-known example is the remarkable success of Alpha GO [29–31]. DRL provides a new method for directly determining the control policy by interacting with the environment without the need to solve ARE. Novel DRL algorithms, such as deep Q-network (DQN) [32], demonstrate the possibilities of solving complex human-level problems by using DRL strategies and thus establishes the basis of DRL. However, DQN uses discrete action and is unsuitable for continuous control problems. Newly developed algorithms, such as deterministic policy gradient (DPG) [33], deep DPG (DDPG) [34], trust region policy optimization (TRPO) [35], and proximal policy optimization (PPO), [36] have exhibited notable success in robotics control tasks for the action-state domain. Radmard Rahmani et al. [37] used a modified DQN in structural vibration control by optimizing a simple reward function. However, the control signal was discrete and could not be precisely determined. Khalatbarisoltani et al. [38] used Q-learning to tune a fuzzy logic-based AMD. Duan et al. [39] reviewed a couple of benchmarks for DRL continuous control tasks, such as cart-pole balancing, mountain car, double-inverted pendulum balancing, and several locomotion tasks. Their results showed that TRPO and DDPG are effective methods for training deep NN policies, while DDPG converged significantly faster due to its better sample efficiency. The DDPG algorithm improves the policy continuously compared with batch algorithms by exploring the environment. The deterministic policy network allows the algorithm to provide continuous control, and it has been applied to robotics, quadrotors, and autonomous vehicle fields.

Research on achieving optimal structural vibration control performance using a model-free method is still absent in the review mentioned above. Existing methods exhibit at least one of the following limitations: (1) the control performance is non-optimal, (2) the method cannot be applied to multi-degree-of-freedom (MDOF) systems, or (3) the method cannot be extended to partially observed systems. This paper proposes a novel active vibration control strategy based on DRL to fill these existing gaps. This control strategy can be applied to various linear vibration control problems because its model-free nature does not require any knowledge of system dynamics. Compared with traditional active control based on NN, the current study has demonstrated the ability to find optimal strategies that traditional NN methods cannot achieve. The major contributions of this work are summarized as follows:

(a) In fully observable SDOF and MDOF systems, the proposed model-free NN controller is compared with a model-based optimal LQR controller. The control performance under free and random vibrations and the robustness against measurement noise are examined.

(b) The same strategy can be directly applied to a partially observed system. The corresponding control performance is compared with a full-state controller and an output feedback controller under different excitations.

The remainder of this paper is organized as follows: The vibration problem is formulated in Section 2. Then, the settings of the traditional model-based methods, namely, LQR and output feedback controllers, are presented as the baseline cases for comparison. The RL setup and the proposed DRL controller are presented in Section 3. The details of the implementation and simulation results of various test conditions are shown in Section 4. Finally, Section 5 summarizes the conclusion of this work.

2. Control Problem Formulation

The vibration control of a linear continuous-time system is examined in this study. The state-space representation can be written as

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{B}_w\mathbf{w}, \tag{1a}$$

$$\mathbf{y} = \mathbf{C}\mathbf{x}, \tag{1b}$$

where $\mathbf{x}$ is the system state vector, $\mathbf{y}$ is the system output, $\mathbf{u}$ is the control action, $\mathbf{w}$ is the excitation, $\mathbf{A}$ is the system matrix, $\mathbf{B}$ is the input matrix, $\mathbf{B}_w$ is the excitation input matrix, and $\mathbf{C}$ is the output matrix. We usually assume that pair ($\mathbf{A}$, $\mathbf{B}$) is controllable, and pair ($\mathbf{A}$, $\mathbf{C}$) is observable.

Two scenarios of the feedback control problem are considered. In the first scenario, the full knowledge of the system state vector $\mathbf{x}$ is measurable. In the second scenario, only a part of the state vector is observed. Two model-based controllers are selected to compare with the proposed model-free controller. The block diagrams of the traditional optimal controllers in these two scenarios are illustrated in Figure 1, in which the control forces are determined based on the state vector $\mathbf{x}$ and output vector $\mathbf{y}$, respectively, in these two controllers.

2.1. Full-State Feedback Control. For the full-state observable system, the classical active LQR controller is selected for comparison with the proposed novel DRL-based controller. The control action $\mathbf{u}$ is given as

$$\mathbf{u} = -\mathbf{K}_{lqr}\mathbf{x}, \tag{2}$$

$\mathbf{K}_{lqr}$ is the optimal feedback gain that minimizes the quadratic cost

$$J_x = \int_0^\infty \left(\mathbf{x}^T\mathbf{Q}_x\mathbf{x} + \mathbf{u}^T\mathbf{R}\mathbf{u}\right)\mathrm{d}t, \tag{3}$$

where $\mathbf{Q}_x$ and $\mathbf{R}$ are the state and control force weight matrices, respectively; $\mathbf{x}^T\mathbf{Q}_x\mathbf{x}$ stands for the cost of states; and $\mathbf{u}^T\mathbf{R}\mathbf{u}$ is the force cost. The optimal feedback gain $\mathbf{K}_{lqr}$ is determined by

$$\mathbf{K}_{lqr} = \mathbf{R}^{-1}\mathbf{B}^T\mathbf{P}, \tag{4}$$

where $\mathbf{P}$ is the solution for ARE.

$$\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} - \mathbf{P}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{P} + \mathbf{Q}_x = \mathbf{0}. \tag{5}$$

$\mathbf{P}$ can be determined by solving the ARE matrix. Using Equation (4), we obtain the optimal gain matrix $\mathbf{K}_{lqr}$. The system response with the LQR controller can now be described by

$$\dot{\mathbf{x}} = \left(\mathbf{A} - \mathbf{B}\mathbf{K}_{lqr}\right)\mathbf{x}. \tag{6}$$

If the quadratic cost function is calculated on the basis of the output vector $\mathbf{y}$, instead of the vector $\mathbf{x}$, then

$$J_y = \int_0^\infty \left(\mathbf{y}^T\mathbf{Q}_y\mathbf{y} + \mathbf{u}^T\mathbf{R}\mathbf{u}\right)\mathrm{d}t = \int_0^\infty \left(\mathbf{x}^T\mathbf{C}^\mathbf{T}\mathbf{Q}_y\mathbf{C}\mathbf{x} + \mathbf{u}^T\mathbf{R}\mathbf{u}\right)\mathrm{d}t. \tag{7}$$

The corresponding algorithm is often termed as LQRy. Note that the control action $\mathbf{u} = -\mathbf{K}_{lqry}\mathbf{x}$ is still based on the full-state vector, where $\mathbf{K}_{lqry}$ is the optimal feedback gain that minimizes the quadratic cost in Equation (7). Accordingly, the ARE is transformed into

$$\mathbf{A}^T\mathbf{P} + \mathbf{P}\mathbf{A} - \mathbf{P}\mathbf{B}\mathbf{R}^{-1}\mathbf{B}^T\mathbf{P} + \mathbf{C}^T\mathbf{Q}_y\mathbf{C} = 0. \tag{8}$$

The block diagram corresponding to the LQR/LQRy controller is presented in Figure 1(a).

2.2. Output Feedback Control. It is usually impractical to observe a full-state vector in a large-scale structure. Therefore, the second scenario represents a practical condition when the state vector is only partially observable. The output feedback controller is adopted, in which the control action is determined directly by the output measurement,

$$\mathbf{u} = -\mathbf{K}_{of}\mathbf{y}. \tag{9}$$

The gain matrix $\mathbf{K}_{of}$ can be calculated as [14]

$$\mathbf{K}_{of} = \mathbf{R}^{-1}\mathbf{B}^T\mathbf{P}\mathbf{S}\mathbf{C}^T\left(\mathbf{C}\mathbf{S}\mathbf{C}^T\right)^{-1}, \tag{10}$$

where matrices $\mathbf{P}$ and $\mathbf{S}$ are the solutions for the following Lyapunov equations:

$$\begin{aligned}&\left(\mathbf{A} - \mathbf{B}\mathbf{K}_{of}\mathbf{C}\right)^T\mathbf{P} + \mathbf{P}\left(\mathbf{A} - \mathbf{B}\mathbf{K}_{of}\mathbf{C}\right)\\ &+\mathbf{C}^T{\mathbf{K}_{of}}^T\mathbf{R}\mathbf{K}_{of}\mathbf{C} + \mathbf{C}^T\mathbf{Q}_y\mathbf{C} = 0,\end{aligned} \tag{11a}$$

$$\left(\mathbf{A} - \mathbf{B}\mathbf{K}_{of}\mathbf{C}\right)^T\mathbf{S} + \mathbf{S}\left(\mathbf{A} - \mathbf{B}\mathbf{K}_{of}\mathbf{C}\right) + \mathbf{X} = 0. \tag{11b}$$

$\mathbf{X}$ is the expected value, i.e.,

$$\mathbf{X} = {}^{\mathrm{E}}\left[\mathbf{x}(0)\mathbf{x}^T(0)\right]. \tag{12}$$

Considering that an initial gain matrix iterative solution algorithm can be used to search for the optimal gain matrix $\mathbf{K}_{of}$, the system response under output feedback control is written as

$$\dot{\mathbf{x}} = \left(\mathbf{A} - \mathbf{B}\mathbf{K}_{of}\mathbf{C}\right)\mathbf{x}. \tag{13}$$

The block diagram of the output feedback control is presented in Figure 1(b). Note that the LQR, LQRy, and output feedback controllers represent the classic model-based control algorithms that require the full knowledge of the state space models.

3. Vibration Control Based on DRL

3.1. RL for Vibration Control. The RL problem includes five key components: environment, agent, state, action, and reward. When applied to a feedback control system, the environment is equivalent to the plant model; the agent acts as the controller; the state is equal to the complete observation of the environment, but the observation can also be partial; the action is the same as that in the control scheme; and the agent interacts directly with the environment by observing the state and performing the action. [40] In addition, a reward is generated in each step to measure the success or failure of the agent's action, and the agent can learn the action policy through experience. The agent aims to find the optimal policy μ^* to maximize the total discounted reward. If a model-free RL algorithm is adopted, the optimal controllers can be determined without knowledge of system

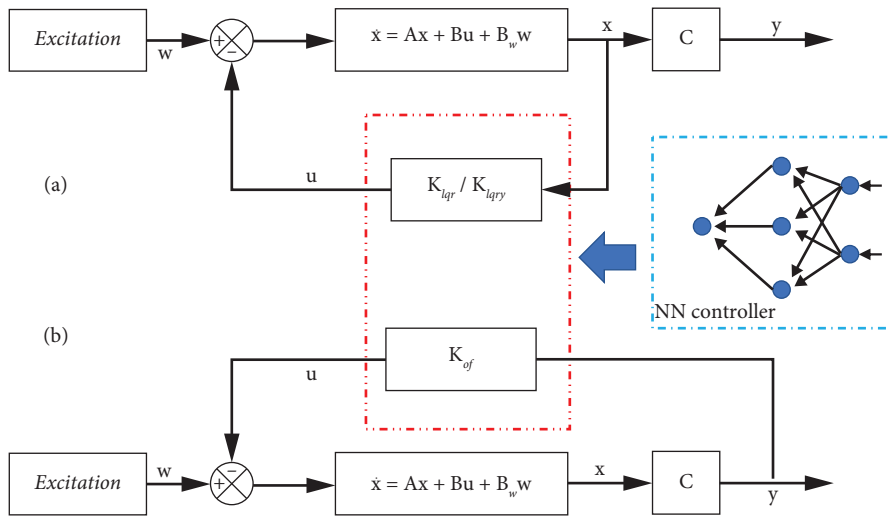

FIGURE 1: Block diagrams of classical control: (a) state-space model with LQR and LQRy controllers and (b) state-space model with an output feedback controller.

dynamics. Moreover, a model-free algorithm exhibits the advantage of handling model uncertainties in real-world control applications, as it does not require accurate system dynamics. The interaction between an agent and the environment in a standard RL setup is illustrated in Figure 2.

The dynamic system model is first discretized when designing an RL environment for vibration control. Then, the environment is formulated as a Markov decision process (MDP) model, that is, the future state only depends on the most current state and action, referred to as the Markov property. At each time step t, the environment experiences the state s_t, receives the action a_t based on the policy μ, and the corresponding reward is r_t. Then, it ends in a new state s_{t+1}. The total discounted reward from time t is defined as the return G_t, with the discount factor $\gamma \in (0, 1)$.

$$G_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \ldots = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1}. \tag{14}$$

If the state is partially observed, then the system is regarded as a partially observable Markov decision process (POMDP). The time horizon can also be set to finite to achieve the optimal reward in the given steps. The optimal policy μ^* shall maximize the discounted reward.

$$\mu^* = \operatorname{argmax} G_t. \tag{15}$$

The objective is to suppress vibrations under excitations via the quadratic cost functions. For each step, the expected reward G_t represents future control performance. Thus, instead of obtaining an analytical solution or numerically searching for optimal strategies, the policy in RL is learned through interaction without requiring a full-dynamics model. The step reward for full-state feedback control is defined as the negative quadratic cost for the vibration control problem.

$$r_t = -\mathbf{x}_t^T \mathbf{Q}_x \mathbf{x}_t - \mathbf{u}_t^T \mathbf{R} \mathbf{u}_t. \tag{16}$$

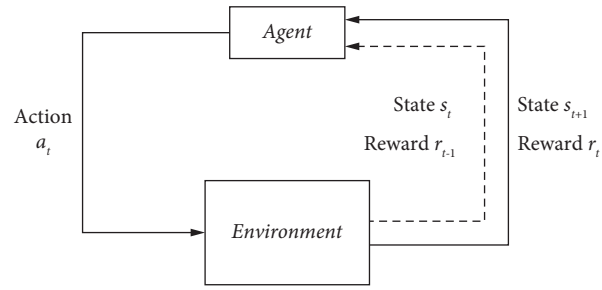

FIGURE 2: RL setup.

The step reward for partial observable state control is

$$r_t = -\mathbf{y}_t^T \mathbf{Q}_y \mathbf{y}_t - \mathbf{u}_t^T \mathbf{R} \mathbf{u}_t. \tag{17}$$

Consequently, maximizing the total reward is equivalent to minimizing the quadratic cost function.

3.2. DRL-Based Vibration Control Strategy. The current study chooses the DDPG algorithm to train the NN controller. DDPG is a model-free, off-policy, online RL algorithm. The DDPG algorithm uses a deterministic policy instead of the traditional stochastic policy. Thus, the DDPG agent outputs a deterministic action for each observation state. Furthermore, in contrast to the well-known DQN algorithm, which can only handle discrete actions, the policy of DDPG is modeled as a deep NN. Consequently, it can work with continuous action spaces, and this capability is critical for high-dimensional space vibration control.

The DDPG algorithm adopted in this work is a combination of DPG and DQN presented in Ref. [34]. The actor-critic structure in DPG is represented by deep NNs in the DDPG agent. The actor outputs the corresponding action to the environment, and the critic approximates the long-term reward based on the observation and action. The expectation Q^μ of a state-action pair (s_t, a_t) is

$$Q^{\mu}(s_t, a_t) = \mathrm{E}[r(s_t, a_t) + \gamma Q^{\mu}(s_{t+1}, \mu(s_{t+1}))]. \quad (18)$$

The critic network Q^{μ} parameterized by θ^{Q} is optimized by minimizing loss L as follows:

$$L = \mathrm{E}\left[\left(Q^{\mu}(s_t, a_t \mid \theta^{Q^{\mu}}) - Y_t\right)^2\right], \quad (19)$$

where Y_t is the value function target at time t, i.e.,

$$Y_t = r(s_t, a_t) + \gamma Q^{\mu}(s_{t+1}, \mu(s_{t+1}) \mid \theta^{Q^{\mu}}). \quad (20)$$

Two significant changes introduced into the Q-learning by DDPG are a replay buffer and two separate networks for critic and actor updates to increase the learning stability. The critic is learned using Bellman's equation, i.e., Equation (18), and the actor is updated through gradient descent. As an off-policy algorithm, DDPG exhibits the advantage of independent exploration from learning. Temporally correlated exploration noise is added to the action during training for better exploration. The flowchart of using DDPG to train the NN controller is presented in Figure 3, and the training follows Algorithm 1.

After the training process, the actor-network can be used as the NN controller separately to generate control forces by accepting the states as input, similar to a regular controller in an offline manner, as shown in Figure 1.

4. Simulation Results

In this section, various cases are presented to demonstrate the effectiveness of DRL agents in vibration control. Starting with a simple SDOF model (Section 4.1), the performance of the DRL-trained NN controller is compared with the LQR control under free and random vibrations. Subsequently, this strategy is extended to a six-story shear building model with an actuator installed in the first story. Full-state control (Section 4.2) and partially observed state control (Section 4.3) are examined. Moreover, the robustness of the DRL-based controller is tested by including measurement noise (Section 4.4).

4.1. SDOF System. Consider an SDOF system with the system matrices as follows:

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \mathbf{B} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \quad (21)$$

The weighting matrices $\mathbf{Q}_x$ and $\mathbf{R}$ in the cost function are selected to be

$$\mathbf{Q}_x = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \mathbf{R} = 1. \quad (22)$$

Given an initial displacement and velocity, the SDOF system undergoes free vibration with the control action. 10-s free vibration records under the current agent are used for each training episode. The actor-network has one hidden layer with a dimension of 64, and the critic network has two hidden layers, each with a dimension of 64. Both networks are composed of rectified linear unit (ReLU) activations. The training progress is shown in Figure 4, and the agent reaches the total reward of the LQR controller in less than 500 episodes.

The performance of the LQR controller (model-based) and the DDPG agent (model-free) under free vibrations is compared in Figures 5(a) and 5(b). Given the initial state (1, 1), the displacement and velocity trajectories of the SDOF system controlled by the DDPG agent and the LQR controller are nearly the same.

Although the DDPG agent is only trained on the basis of free vibration, white noise random excitations are also applied to test the performance. Random excitations of three different levels (low, middle, and high) are applied to examine if the DDPG agent produces any nonlinear effect that would affect the control performance. The cumulative cost for each simulation is provided in Table 1. The result of the DDPG controller is still very close to that of the LQR controller. Figure 6 presents the displacement comparison in one random excitation case. The performance of the two controllers is generally consistent. Although trained on free vibration only, the model-free DDPG controller can still offer an optimal control performance comparable to the classical LQR controller for the SDOF system under random situations.

4.2. Full-State Observable MDOF System. The proposed control strategy is extended from SDOF to a higher-order MDOF system, namely, a six-story shear building with an actuator in the first story as shown in Figure 7. A couple of assumptions are made for this simplified model. The floor is assumed to be rigid without rotation, and the floor mass is modeled as a lumped mass. Table 2 provides each floor's structural properties. The damping ratio is set as 5% for all modes. Notably, although the actuator is assumed to be installed in the first story of the six-story building in this simulation case, the presented control strategy can be extended to other actuator locations or even other structures.

Only the inplane vibration is considered. The corresponding equation of motion can be written as

$$\mathbf{m}\ddot{\mathbf{z}} + \mathbf{c}\dot{\mathbf{z}} + \mathbf{k}\mathbf{z} = -\mathbf{m}\mathbf{b}_w\mathbf{w} + \mathbf{b}\mathbf{u}, \quad (23)$$

where $\mathbf{m}$, $\mathbf{c}$, and $\mathbf{k}$ are the mass, damping, and stiffness matrices of the shear building, respectively; $\mathbf{z} = [z_1 \ \ldots \ z_6]^T$ includes the displacement of each floor; $\mathbf{w}$ is the ground excitation; $\mathbf{u}$ is the control force; $\mathbf{b} = (1\ 0\ 0\ 0\ 0\ 0)^T$; and $\mathbf{b}_w = (1\ 1\ 1\ 1\ 1\ 1)^T$.

This equation can transform into a state-space representation as

$$\dot{\mathbf{x}} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{B}_w\mathbf{w}, \quad (24)$$

where $\mathbf{x}$ is the full-state vector of velocity and displacements. $\mathbf{A}$ is the state matrix, i.e.,

$$\mathbf{A} = \begin{bmatrix} \mathbf{0} & \mathbf{I} \\ -\mathbf{m}^{-1}\mathbf{k} & -\mathbf{m}^{-1}\mathbf{c} \end{bmatrix}. \quad (25)$$

$\mathbf{B}_{\mathrm{w}}$ is the input matrix for the excitation force, i.e.,

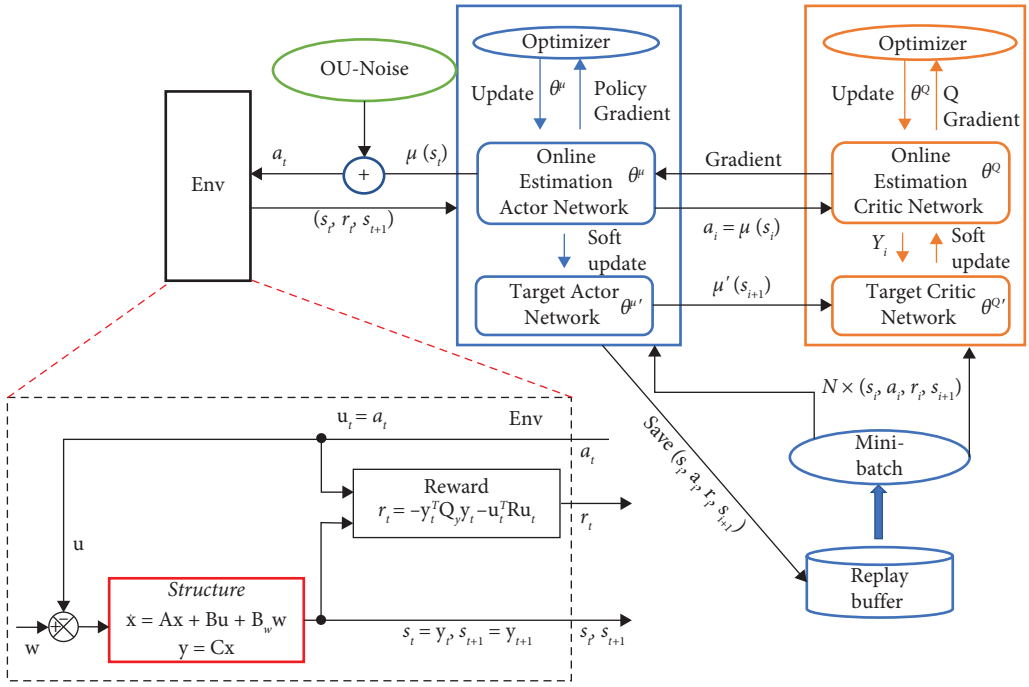

Figure 3: Flowchart of using the DDPG algorithm to train a vibration controller.

Initialization step: initial DDPG agent, initial state vector $\mathbf{x}(0)$
repeat
  Observe state s_t and get action $a_t = \mu(s_t)$
  Execute a_t in the environment
  Observe the next state
  Get the reward $r_t(s_t, a_t)$
  If s_{t+1} is terminal, then reset the initial states and start a new episode
  If it is time to update, then
    $\mu' = \text{DDPG}(\mu, [s_t, r_t(s_t, a_t), s_{t+1}])$
  end if
until convergence

Algorithm 1: DRL controller training.

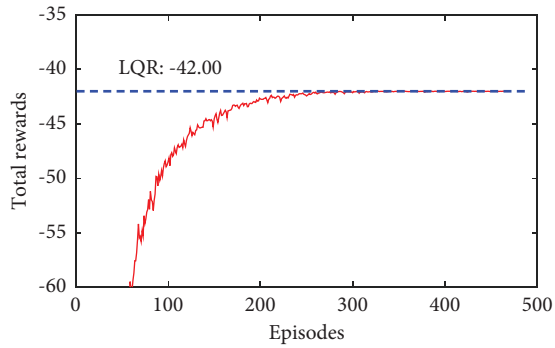

Figure 4: Episode reward of the DDPG agent during training for the SDOF system.

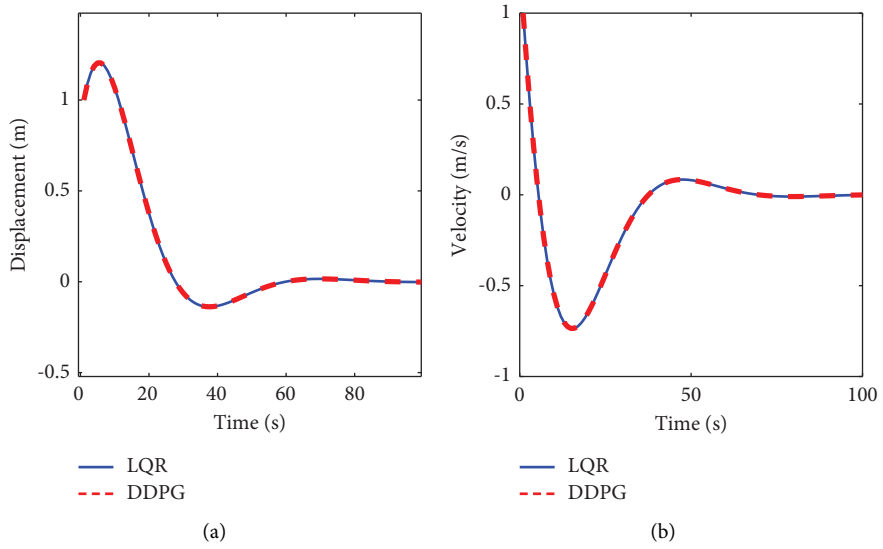

Figure 5: Comparison of SDOF free vibration: (a) displacement and (b) velocity.

Table 1: Cumulative cost comparison under white noise excitations.

Time (s)	Excitation level	DDPG	LQR
500	Low	720.96	720.94
	Middle	902.70	902.73
	High	7.18×10^4	7.18×10^4

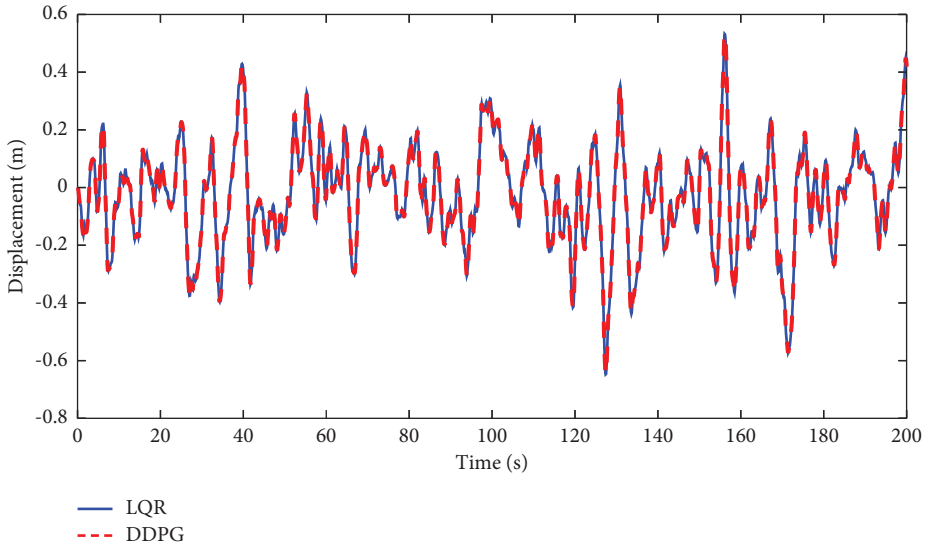

Figure 6: Displacement comparison of the SDOF system under random excitations.

$$\mathbf{B}_w = \begin{bmatrix} \mathbf{0} \\ \mathbf{1} \end{bmatrix}. \tag{26}$$

$\mathbf{B}$ is the input matrix for the control force, i.e.,

$$\mathbf{B} = \begin{bmatrix} \mathbf{0} \\ \mathbf{m}^{-1}\mathbf{b} \end{bmatrix}. \tag{27}$$

The continuous state-space model is further discretized with a sampling interval of 0.01 s. The performance indices $\mathbf{Q}_x$ and $\mathbf{R}$ in the cost function are set as $\mathbf{Q}_x = 10 \times \mathbf{I}_{12\times 12}$ and $\mathbf{R} = 10^{-4}$. The agent critic has three hidden layers, each with a dimension of 64 and ReLU activations. The actor is represented by an NN composed of tanh nonlinearities and two hidden layers, each with a dimension of 64. The training process is the same as that for the SDOF system. A step

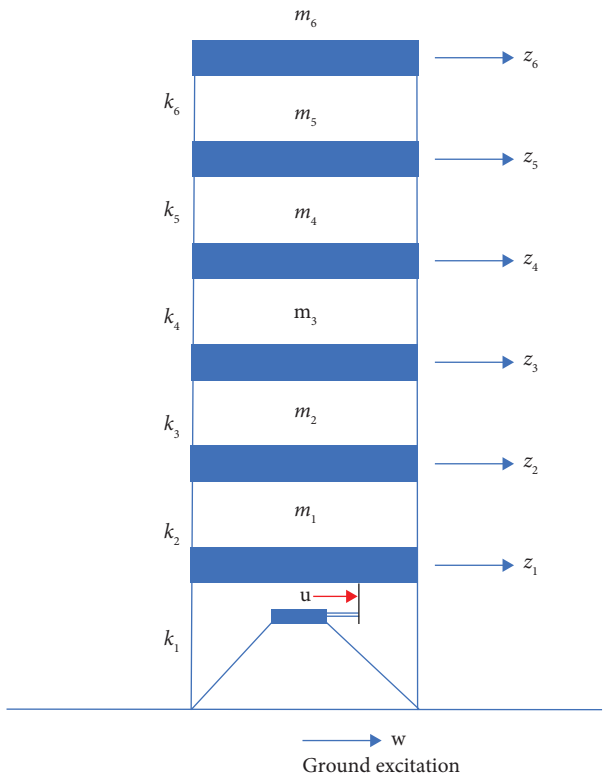

Figure 7: Six-story shear building diagram.

Table 2: Six-story shear building model properties.

Story	Mass ($\times 10^3$ kg)	Stiffness ($\times 10^6$ N/m)
1	25	6.82
2–5	20	6.06
6	15	6.06

excitation is applied at the starting point, and the building is allowed to perform under free vibration, and 5-s free vibration records are used in the training process.

The total cost of the LQR controller and the trained DDPG agent are 292.23 and 295.45. The total cost of the DDPG agent is slightly higher, but its performance is still very close to that of the LQR controller. Figures 8(a) and 8(b) show the displacement and velocity time histories of two representative floors. The vibration decay rates of the two controllers are extremely close to each other and are in the same phase.

The NN controller trained on free vibration is further tested under random excitations and sine sweep excitations (sweeping from 0.5 Hz to 10 Hz in 20 s). Figure 9(a) depicts the displacement response of the first floor under random excitations. Figure 9(b) shows the root-mean-square (RMS) displacement responses at different floors under random excitations. The RMS displacement of the DDPG agent is smaller than that of the LQR controller, indicating a slightly better control performance of the DDPG agent.

Figure 10(a) presents the displacement response of the sixth floor under sine sweep excitations. The response time histories are nearly identical for the two controllers under all frequencies. Figure 10(b) illustrates the relationship between floor displacement and control force. The two controllers demonstrate very close features. Vibration comparison under different excitations indicates that the model-free NN controller trained through the DRL method without any knowledge of system dynamics under various occasions can compete with the performance of the model-based LQR controller in full-state feedback control.

4.3. Partially Observed MDOF System. One case with a partially observed state is further investigated using the same six-story shear building model. Only six states are observed: the first-, third-, and fifth-floor displacements and velocities. The agent structure follows that of the full-state controller except for changing the nodes of the input layers. The only difference in training is the reward calculation. The weighting matrix $\mathbf{Q}_y$ is selected to be $10 \times \mathbf{I}_{6\times 6}$, while $\mathbf{R}$ is the same. The system performs training under free vibration for 6 s during each training episode.

The free vibration results are provided in Table 3. The performance of LQRy is the best among the three because it uses the fully observable state to calculate the control force. The trained DDPG agent using partially observable states can achieve performance similar to that of LQRy, and the DDPG agent can outperform the output feedback controller with the same number of observable states. From the training progress shown in Figure 11, the episode reward of the DDPG agent surpasses that of the output feedback controller after a dozen training steps. When training stops, the total reward is extremely close to the result of the full-state controller LQRy. The control performance of the output feedback controller is the weakest among the three. Figures 12(a) and 12(b) present the displacement responses of the first floor and top floor, respectively.

Random and sine sweep excitations are also applied. Figures 13(a) and 13(b) show the displacement of the first floor under two different excitations. Similarly, the control performance of the DDPG agent is close to that of the full-state LQRy controller and exceeds that of the output feedback controller. Figure 13(c) illustrates the force-displacement relationship of the three controllers, in which the negative stiffness features of the LQRy and DDPG agents are close to each other and more significant than that of the output feedback controller when selecting the same weighting matrix. The comparison of the RMS displacement of all the floors is depicted in Figure 13(d).

Figure 14 shows the components of the total cost under sine sweep excitations. The output feedback controller has the highest cost. Although the force component cost of the output feedback controller is considerably lower than the other two, its state cost is enormous. The DDPG controller has a slightly higher state cost than LQRy, leading to a higher total cost.

4.4. Full-State Observable MDOF System with Measurement Noise. Measurement noise in the observed state vector is considered to demonstrate the robustness of the DRL strategy. Thus, the full-state observation becomes

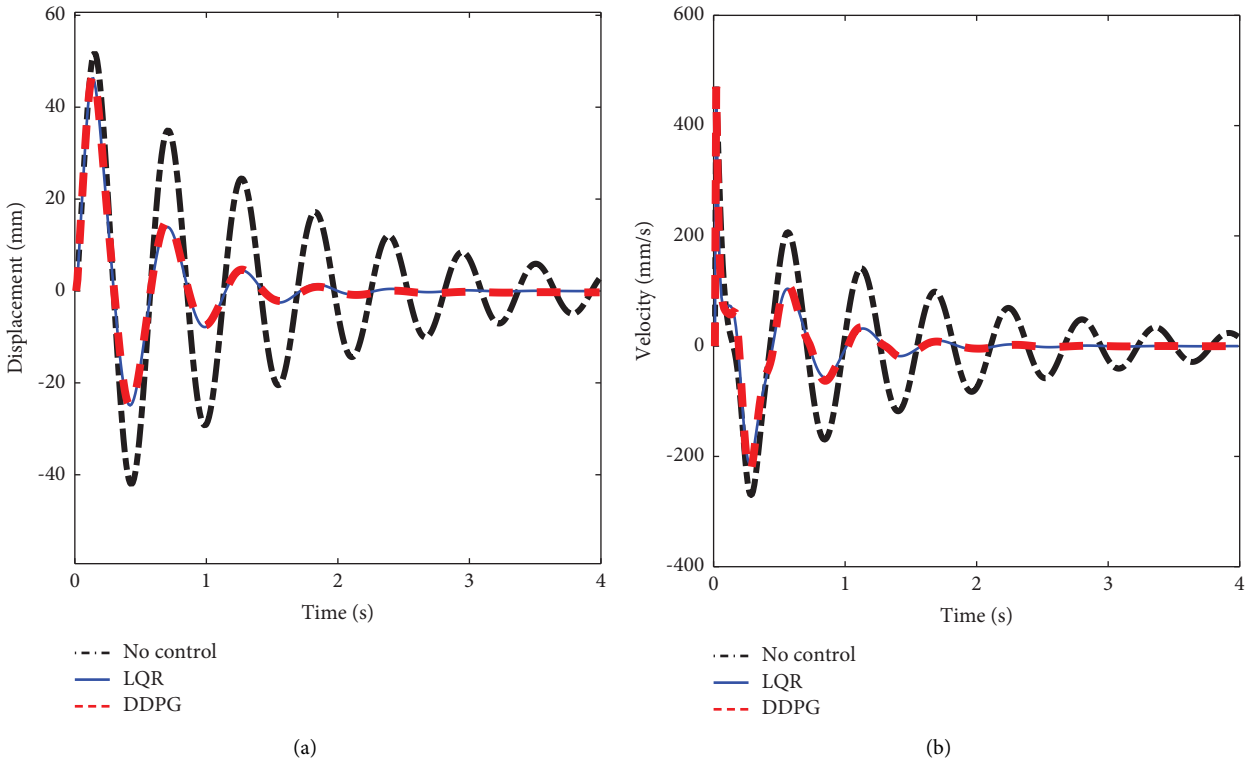

Figure 8: Free vibration of full-state observable MDOF: (a) top-floor displacement and (b) first-floor velocity.

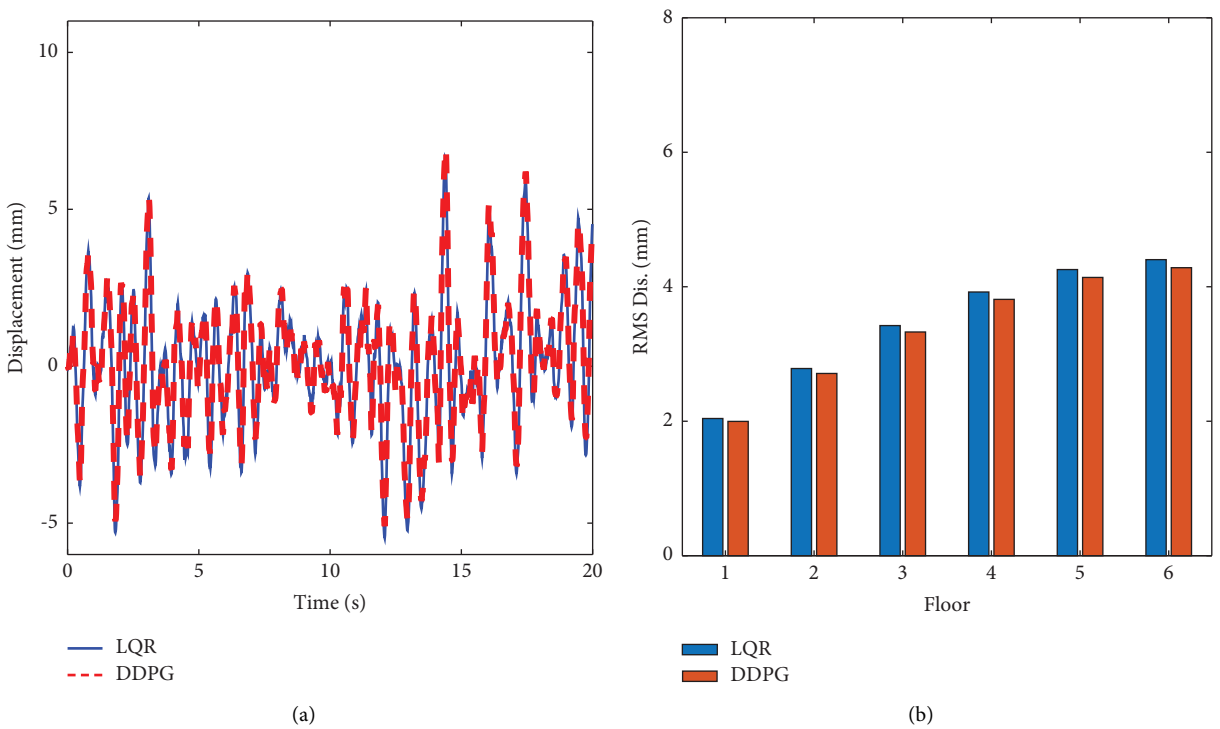

Figure 9: Full-state observable MDOF under random excitations: (a) top-floor displacement and (b) RMS displacement of different floors.

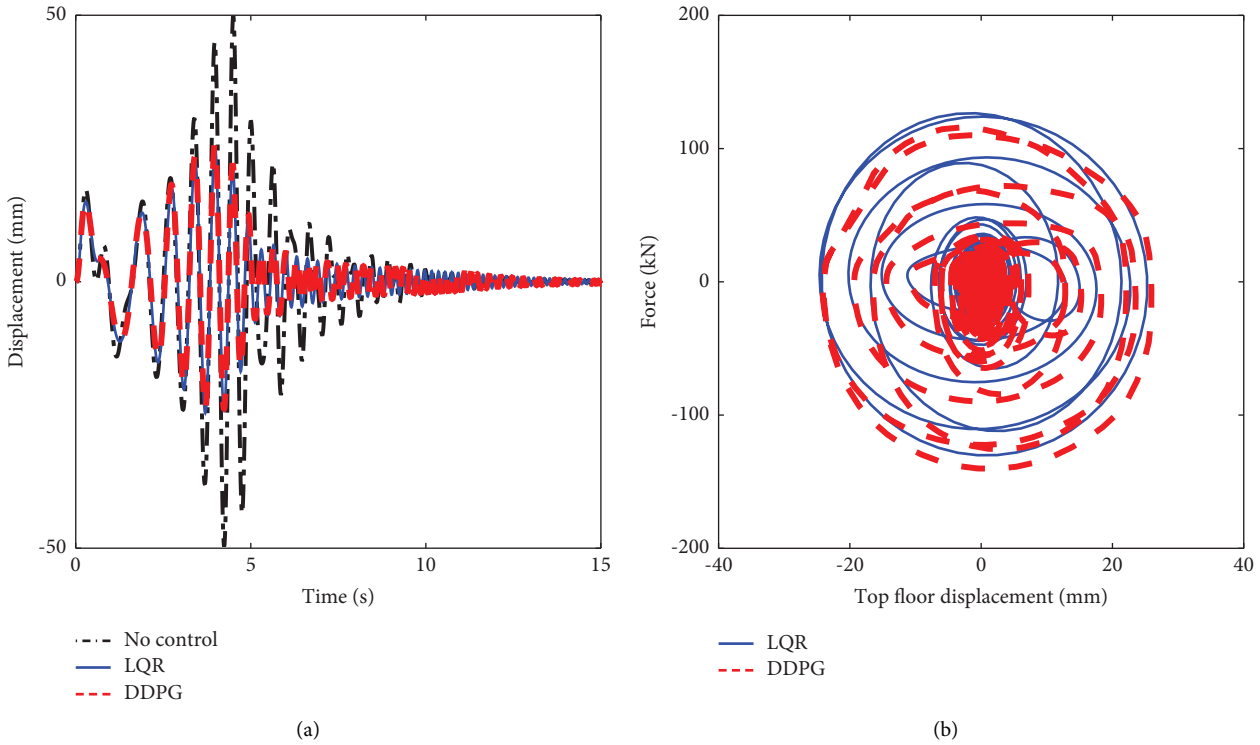

Figure 10: Full-state observable MDOF under sine sweep excitation: (a) top-floor displacement response and (b) control force vs. first-floor displacement relationship under sine sweep excitation.

Table 3: Total cost for free vibration of partially observed MDOF system.

Controller	No. of observable states	Total cost
LQRy	12	164.56
Output feedback	6	203.56
DDPG	6	168.80

$$\mathbf{y}_n = \mathbf{y} + \mathrm{w}_n. \tag{28}$$

The training progress stays the same as that in Section 4.2 except for adding the measurement noise, and the trained controller is compared with the LQR controller. The total cost after the training is summarized in Table 4. Similar to the case without noise, the total cost of LQR is still slightly lower than that of DDPG when the measurement noise is considered. In both cases without and with noise, the differences in the total cost are less than 2%. The displacement time histories of the first and top floors are shown in Figures 15(a) and 15(b), respectively. The vibration decay rates are nearly the same in consideration of measurement noise in the observation function. The control force–displacement relationship is presented in Figure 16. The maximum force provided by the DDPG agent is lower than that of LQR, but the overall trend is highly similar.

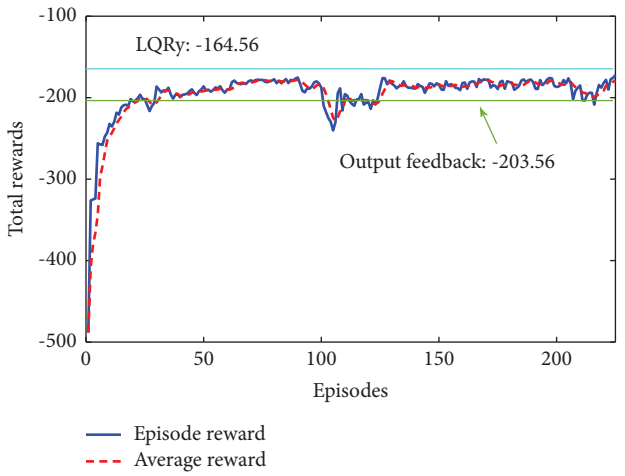

Figure 11: Episode and average reward of DDPG during training of the partially observed MDOF system.

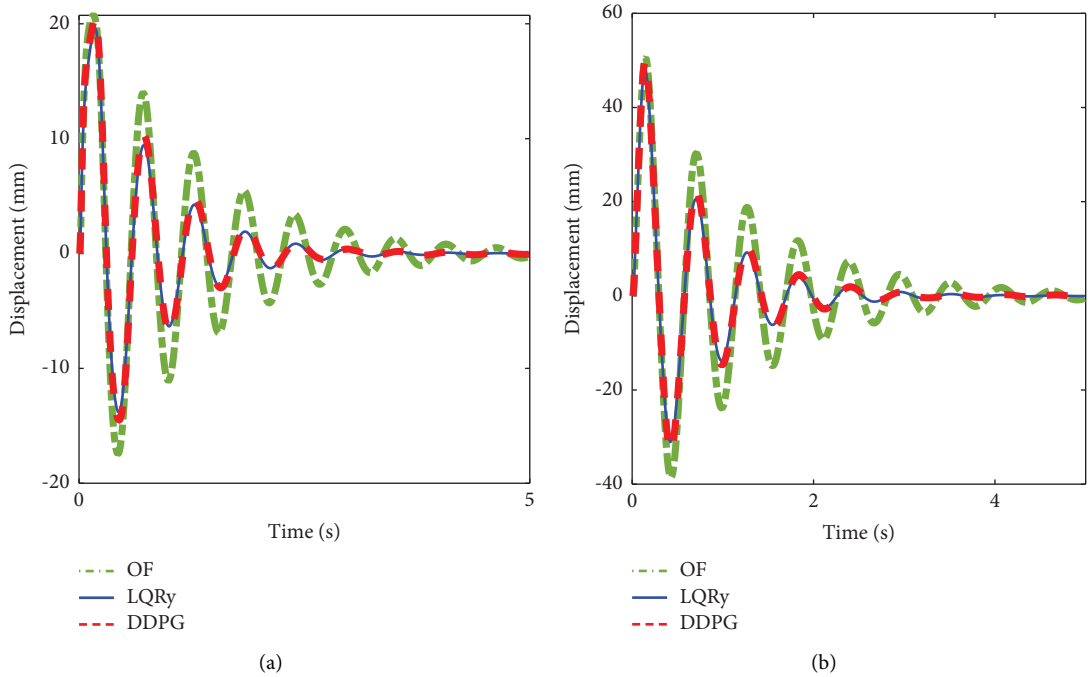

Figure 12: Displacement response of the partially observed MDOF system under free vibration: (a) first floor and (b) top floor.

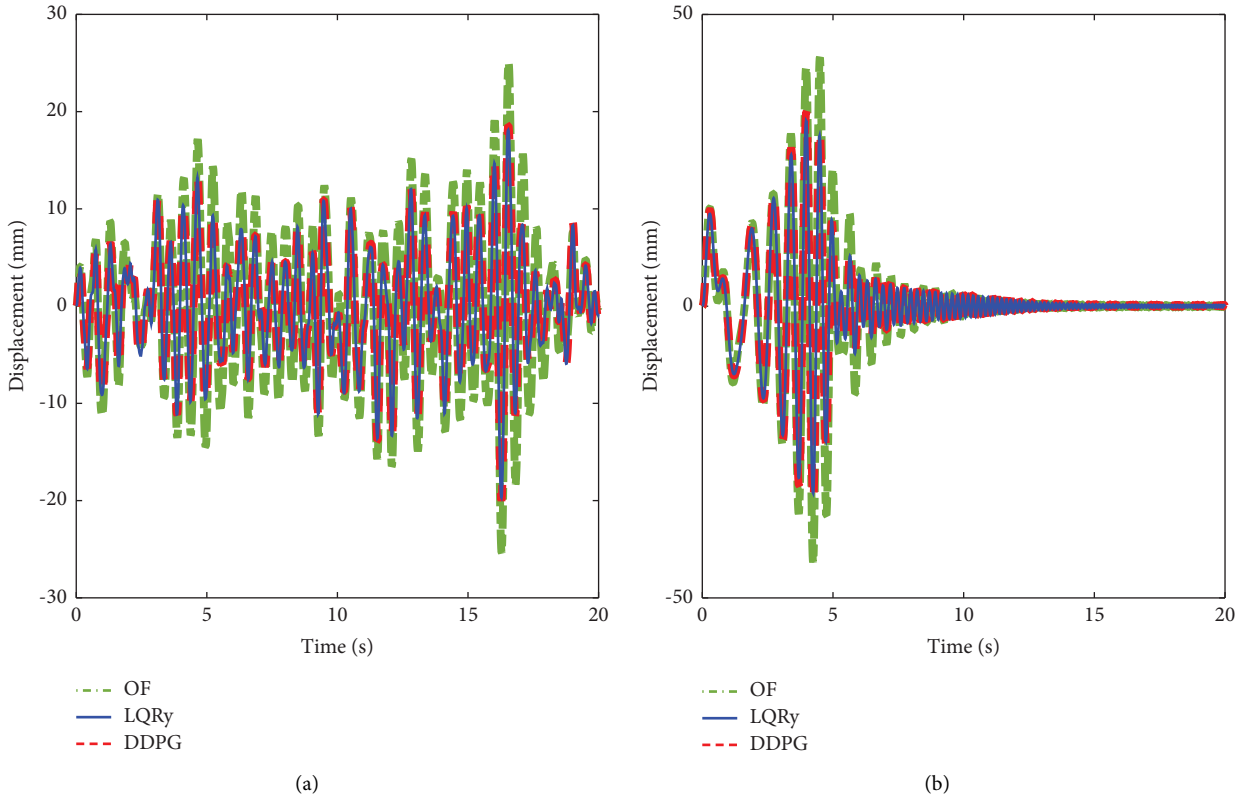

Figure 13: Continued.

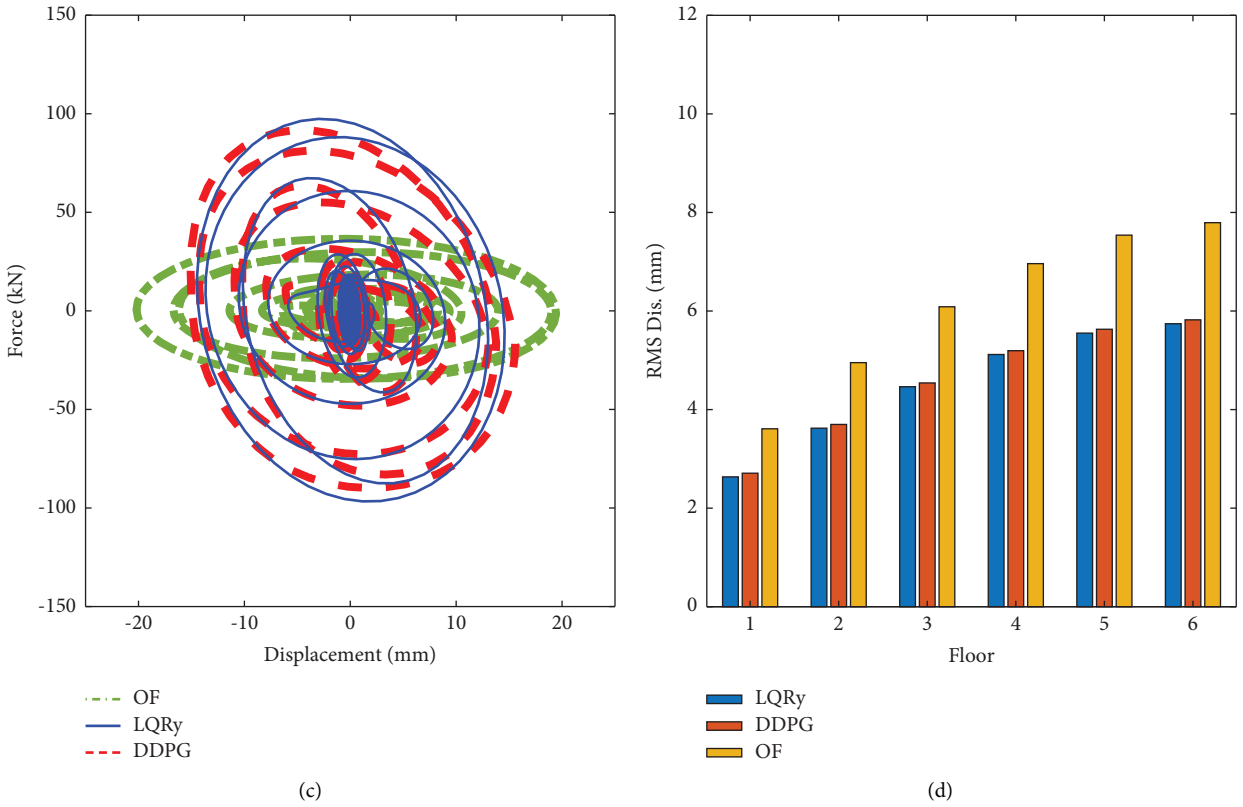

Figure 13: Partially observed MDOF under different excitations: (a) top floor displacement under random excitations, (b) top floor displacement under sine sweep excitations, (c) control force vs. first-floor displacement relationship under sine sweep excitation, and (d) RMS displacement at different floors under random excitations.

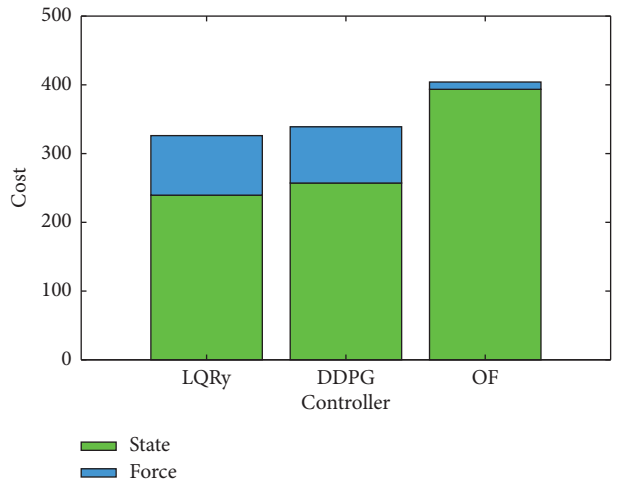

Figure 14: Components of total cost under sweep signal excitations.

Table 4: Total cost of MDOF under free vibration with and without noise.

Noise	LQR	DDPG	Difference (%)
Without noise	292.23	296.96	1.59
With noise	293.42	298.26	1.62

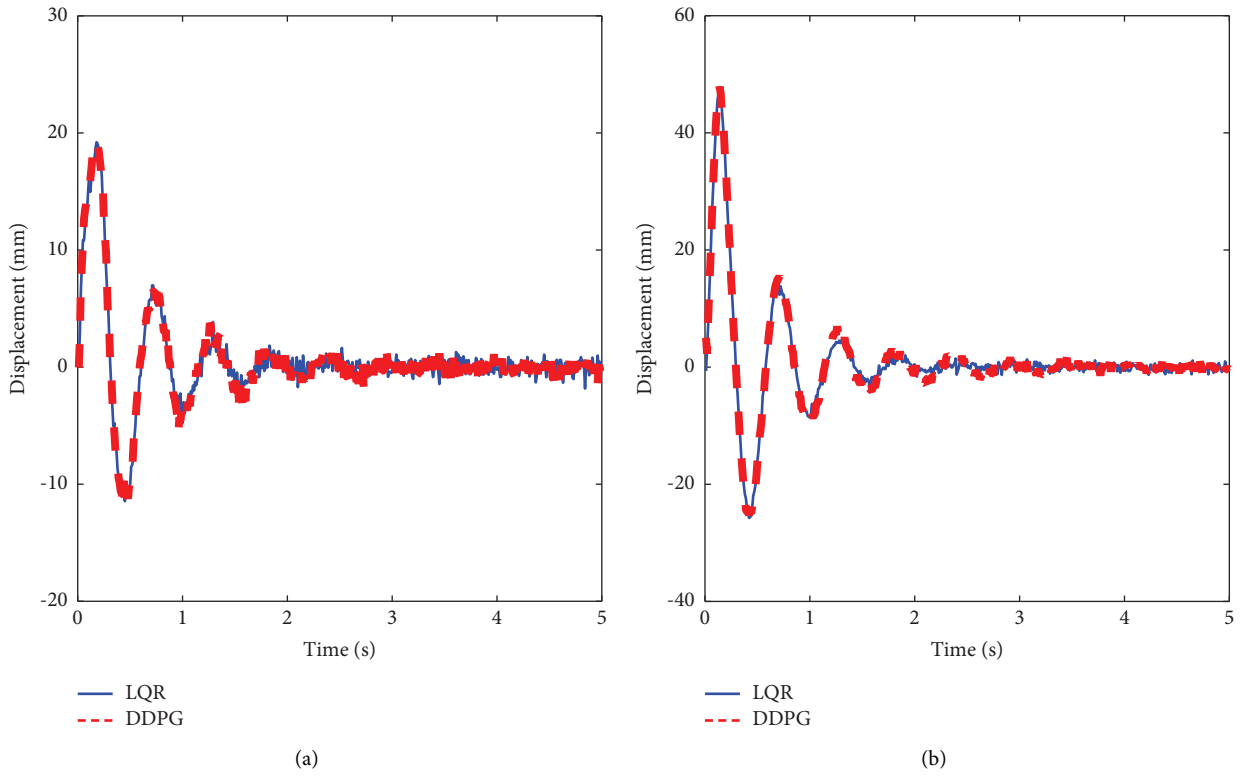

Figure 15: Displacement response of MDOF with measurement noise under free vibration: (a) first floor and (b) top floor.

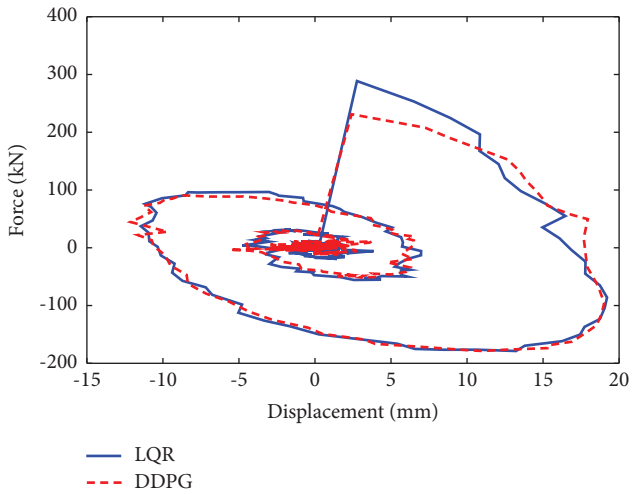

Figure 16: Control force vs. first-floor displacement relationship of MDOF with measurement noise.

5. Conclusion

A novel DRL-based vibration control strategy is presented in this work, which, to the best of the authors' knowledge, presents the first study that introduces DRL into the optimal structure control domain. An off-policy model-free algorithm, i.e., DDPG, is used to train the NN controller, and its effectiveness is tested under various conditions.

The proposed DRL-based controller does not require any information regarding structural dynamic models but can still achieve control performance comparable to that of the traditional model-based LQR controllers. The proposed method is tested in an SDOF system and a six-story shear building MDOF system. The agents are always trained under free vibrations and then tested under different levels of random excitations. For all the cases, the total costs and displacement levels are nearly the same as those of LQR. In addition, the control performance is relatively consistent with measurement noise.

When feedback is not in a full state, the DRL controller is considerably better than the output feedback controller with the same number of observed states. The control performance of the DRL controller does not degrade too much compared with the full-state feedback controller LQR*y*, although the former uses fewer states in the feedback than the latter.

Implementing this proposed strategy in nonlinear systems needs to be investigated in future studies, in which the strength of deep learning is expected to be better utilized.

Data Availability

The data used to support the findings of the study can be obtained from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This study was supported by the Research Grants Council of Hong Kong (Project Nos. 15214620, 15213122, and PolyU R5020-18), the Hong Kong Branch of the National Rail Transit Electrification and Automation Engineering Technology Research Center (K-BBY1), and The Hong Kong Polytechnic University (Project Nos. ZE2L and ZVX6).

References

[1] G. W. Housner, L. A. Bergman, T. K. Caughey et al., "Structural control: past, present, and future," *Journal of Engineering Mechanics*, vol. 123, no. 9, pp. 897–971, 1997.

[2] Q. Cai and S. Zhu, "The nexus between vibration-based energy harvesting and structural vibration control: a comprehensive review," *Renewable and Sustainable Energy Reviews*, vol. 155, Article ID 111920, 2022.

[3] B. F. Spencer and S. Nagarajaiah, "State of the art of structural control," *Journal of Structural Engineering*, vol. 129, no. 7, pp. 845–856, 2003.

[4] F. Y. Cheng, H. Jiang, and K. Lou, *Smart Structures: Innovative Systems for Seismic Response Control*, CRC Press, Boca Raton, FL, USA, 2008.

[5] S. J. Dyke, B. F. Spencer, P. Quast, D. C. Kaspari, and M. K. Sain, "Implementation of an active mass driver using acceleration feedback control," *Computer-Aided Civil and Infrastructure Engineering*, vol. 11, no. 5, pp. 305–323, 1996.

[6] F. Bossens and A. Preumont, "Active tendon control of cable-stayed bridges: a large-scale demonstration," *Earthquake Engineering & Structural Dynamics*, vol. 30, no. 7, pp. 961–979, 2001.

[7] J. Rodellar, V. Mañosa, and C. Monroy, "An active tendon control scheme for cable-stayed bridges with model uncertainties and seismic excitation," *Journal of Structural Control*, vol. 9, no. 1, pp. 75–94, 2002.

[8] A. M. Reinhorn, T. T. Soong, R. Lin et al., "Active bracing system: a full scale implementation of active control," *National Center for Earthquake Engineering Research*, vol. 14, 1992.

[9] T. Datta, "A state-of-the-art review on active control of structures," *ISET Journal of Earthquake Technology*, vol. 40, no. 1, pp. 1–17, 2003.

[10] J.-N. Yang, "Application of optimal control theory to civil engineering structures," *Journal of the Engineering Mechanics Division*, vol. 101, no. 6, pp. 819–838, 1975.

[11] M. Abdel-Rohman and H. H. Leipholz, "Active control of flexible structures," *Journal of the Structural Division*, vol. 104, no. 8, pp. 1251–1266, 1978.

[12] G. Song and H. Gu, "Active vibration suppression of a smart flexible beam using a sliding mode based controller," *Journal of Vibration and Control*, vol. 13, no. 8, pp. 1095–1107, 2007.

[13] F. Casciati, J. Rodellar, and U. Yildirim, "Active and semi-active control of structures - theory and applications: a review of recent advances," *Journal of Intelligent Material Systems and Structures*, vol. 23, no. 11, pp. 1181–1195, Jul 2012.

[14] F. L. Lewis, D. Vrabie, and V. L. Syrmos, *Optimal Control*, John Wiley & Sons, New York, NY, USA, 2012.

[15] Z. G. Ying and Y. Q. Ni, "Optimal control for vibration peak reduction via minimizing large responses," *Structural Control and Health Monitoring*, vol. 22, no. 5, pp. 826–846, 2015.

[16] Y. Z. Xie, M. Ebad Sichani, J. E. Padgett, and R. DesRoches, "The promise of implementing machine learning in earthquake engineering: a state-of-the-art review," *Earthquake Spectra*, vol. 36, no. 4, pp. 1769–1801, 2020.

[17] H. Khodabandehlou, G. Pekcan, M. Sami Fadali, and M. M. Salem, "Active neural predictive control of seismically isolated structures," *Structural Control and Health Monitoring*, vol. 25, no. 7, Article ID e2201, 2018.

[18] N. Siddique and H. Adeli, *Computational Intelligence: Synergies of Fuzzy Logic, Neural Networks and Evolutionary Computing*, John Wiley & Sons, New York, NY, USA, 2013.

[19] K. Bani-Hani and J. Ghaboussi, "Nonlinear structural control using neural networks," *Journal of Engineering Mechanics*, vol. 124, no. 3, pp. 319–327, 1998.

[20] Y. A. He and J. J. Wu, "Control of structural seismic response by self-recurrent neural network (SRNN). *Earthquake Engineering & Structural Dynamics*," *Earthquake Engineering & Structural Dynamics*, vol. 27, no. 7, pp. 641–648, 1998.

[21] Y. Tang, "Active control of SDF systems using artificial neural networks," *Computers & Structures*, vol. 60, no. 5, pp. 695–703, Jul 10 1996.

[22] H. M. Chen, K. H. Tsai, G. Z. Qi, J. C. S. Yang, and F. Amini, "Neural-network for structure control," *Journal of Computing in Civil Engineering*, vol. 9, no. 2, pp. 168–176, 1995.

[23] J. Ghaboussi and A. Joghataie, "Active control of structures using neural networks," *Journal of Engineering Mechanics*, vol. 121, no. 4, pp. 555–567, 1995.

[24] J. T. Kim, H. J. Jung, and I. W. Lee, "Optimal structural control using neural networks. *Journal of Engineering Mechanics*," *Journal of Engineering Mechanics*, vol. 126, no. 2, pp. 201–205, 2000.

[25] D. Vrabie, O. Pastravanu, M. Abu-Khalaf, and F. L. Lewis, "Adaptive optimal control for continuous-time linear systems based on policy iteration," *Automatica*, vol. 45, no. 2, pp. 477–484, 2009.

[26] L. M. M. Zhu, H. Modares, G. O. Peen, F. L. Lewis, and B. Z. Yue, "Adaptive suboptimal output-feedback control for linear systems using integral reinforcement learning," *IEEE Transactions on Control Systems Technology*, vol. 23, no. 1, pp. 264–273, 2015.

[27] H. Modares and F. L. Lewis, "Linear quadratic tracking control of partially-unknown continuous-time systems using reinforcement learning," *IEEE Transactions on Automatic Control*, vol. 59, no. 11, pp. 3051–3056, 2014.

[28] R. Moghadam and F. L. Lewis, "Output-feedback H∞quadratic tracking control of linear systems using reinforcement learning," *International Journal of Adaptive Control and Signal Processing*, vol. 33, no. 2, pp. 300–314, 2019.

[29] D. Silver, J. Schrittwieser, K. Simonyan et al., "Mastering the game of go without human knowledge," *Nature*, vol. 550, no. 7676, pp. 354–359, 2017.

[30] Y. Li, "Deep reinforcement learning: an overview," 2017, https://arxiv.org/abs/1701.07274.

[31] V. Mnih, K. Kavukcuoglu, D. Silver et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[32] V. Mnih, K. Kavukcuoglu, D. Silver et al., "Playing atari with deep reinforcement learning," 2013, https://arxiv.org/abs/1312.5602.

[33] D. Silver, G. Lever, N. Heess, T. Degris, D. Wierstra, and M. Riedmiller, "Deterministic policy gradient algorithms," in *Proceedings of the 31st international conference on machine learning*, Beijing China, June 2014.

[34] T. P. Lillicrap, J. J. Hunt, A. Pritzel et al., "Continuous control with deep reinforcement learning," 2015, https://arxiv.org/abs/1509.02971.

[35] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," pp. 1889–1897, 2015, https://arxiv.org/abs/1502.05477.

[36] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," 2017, https://arxiv.org/abs/1707.06347.

[37] H. Radmard Rahmani, G. Chase, M. Wiering, and C. Könke, "A framework for brain learning-based control of smart structures," *Advanced Engineering Informatics*, vol. 42, Article ID 100986, 2019.

[38] A. Khalatbarisoltani, M. Soleymani, and M. Khodadadi, "Online control of an active seismic system via reinforcement learning," *Structural Control and Health Monitoring*, vol. 26, no. 3, Article ID e2298, 2019.

[39] Y. Duan, X. Chen, R. Houthooft, J. Schulman, and P. Abbeel, *Benchmarking Deep Reinforcement Learning for Continuous Control*, in *Proceedings of the 33rd International Conference on International Conference on Machine Learning*, pp. 1329–1338, New York, NY, USA, June 2016.

[40] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, MIT press, Cambridge, MA, USA, 2018.