

Article

# Adaptive Optimization Design of Building Energy System for Smart Elderly Care Community Based on Deep Deterministic Policy Gradient

Chunmei Liu \* and Zhe Xue

School of Economics and Management, Huizhou University, Huizhou 516007, China

\* Correspondence: lcm@hzu.edu.cn

**Abstract:** In smart elderly care communities, optimizing the design of building energy systems is crucial for improving the quality of life and health of the elderly. This study pioneers an innovative adaptive optimization design methodology for building energy systems by harnessing the cutting-edge capabilities of deep reinforcement learning. This avant-garde method initially involves modeling a myriad of energy equipment embedded within the energy ecosystem of smart elderly care community buildings, thereby extracting their energy computation formulae. In a groundbreaking progression, this study ingeniously employs the actor–critic (AC) algorithm to refine the deep deterministic policy gradient (DDPG) algorithm. The enhanced DDPG algorithm is then adeptly wielded to perform adaptive optimization of the operational states within the energy system of a smart retirement community building, signifying a trailblazing approach in this realm. Simulation experiments indicate that the proposed method has better stability and convergence compared to traditional deep Q-learning algorithms. When the environmental interaction coefficient and learning ratio is 4, the improved DDPG algorithm under the AC framework can converge after 60 iterations. The stable reward value in the convergence state is  $-996$ . When the scheduling cycle of the energy system is between 0:00 and 8:00, the photovoltaic output of the system optimized by the DDPG algorithm is 0. The wind power output fluctuates within 50 kW. This study realizes efficient operation, energy saving, and emission reduction in building energy systems in smart elderly care communities and provides new ideas and methods for research in this field. It also provides an important reference for the design and operation of building energy systems in smart elderly care communities.

**Keywords:** deep learning; smart community; architecture; energy system; adaptive; energy conservation



**Citation:** Liu, C.; Xue, Z. Adaptive Optimization Design of Building Energy System for Smart Elderly Care Community Based on Deep Deterministic Policy Gradient. *Processes* **2023**, *11*, 2155. <https://doi.org/10.3390/pr11072155>

Academic Editors: Jubair Shamim, Rajagopalan Parameshwaran, Palash Kumar Bhowmik and Davide Papurello

Received: 25 May 2023  
Revised: 6 July 2023  
Accepted: 14 July 2023  
Published: 19 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Smart elderly care communities provide a safer, more comfortable, and convenient living environment for the elderly through the use of Internet of Things (IT), cloud computing (CC), artificial intelligence (AI), and other means. Smart elderly care community buildings typically include technologies such as smart homes, remote monitoring, the IT, CC, and AI as well as supporting services and management systems. In smart elderly care communities, relevant staff can monitor the physiological health of the elderly by installing sensors, cameras, smart speakers, and other devices. Meanwhile, the communities can also provide better living security by connecting elderly families and community services through IoT technology. In addition, CC technology can be used to store the life information and data of the elderly to provide more intelligent services. The purpose of smart elderly care communities is to provide a safer, more comfortable, and convenient living environment for the elderly. They also aim to improve the level of community services and enable the elderly to enjoy a better quality of life [1,2].

In the construction of a smart elderly care community (SECC), the energy system design of the entire community building is the most crucial part. The building energy system refers to a comprehensive management system that provides energy and resources

for buildings. This includes the supply and control of various energy sources, such as electricity, gas, and hot water, and their related energy saving measures [3]. The purpose of designing a SECC building energy system (ES) is to ensure sufficient energy supply inside the building to meet the daily energy consumption of elderly people in the elderly care community. It also aims to reduce energy consumption and environmental pollution. Currently, many communities do not consider maximizing energy efficiency targets in their energy design. With the emission of a large amount of greenhouse gases, the climate change situation is becoming increasingly severe, and sustainable low-carbon development has become a consensus among countries around the world to address climate change. To achieve low-carbon living, many experts have conducted a series of observations and studies on the building industry, which is the main source of greenhouse gas emissions. Traditional methods for optimizing building energy, such as the use of passive technology, have some limitations. This involves installing solar panels and other heat-absorbing devices on the roof to provide as much energy as possible. These systems typically cannot meet the dynamic energy needs within buildings and may require extensive maintenance and upkeep. Traditional building energy optimization requires on-site operations by professionals and may not be fully automated. In addition, traditional energy optimization methods have drawbacks such as high maintenance costs and low flexibility [4]. This study focuses on an integrated energy system for smart retirement community buildings that is composed of various energy devices that work in conjunction with each other. Through cooperative control and energy coupling, this integrated system is expected to improve the efficiency of energy utilization and promote the use of renewable energy. To achieve adaptive optimization of building energy systems, this study employs reinforcement learning to construct relevant models. Reinforcement learning achieves optimization by allowing an intelligent body to interact with the external environment and learn optimal action strategies. To address the limitations of the Markov process in dealing with high-dimensional states and continuous space problems, this study introduces the deep deterministic policy gradient (DDPG) algorithm to compensate for the shortcomings of traditional methods [5,6]. Through these novel approaches and analyses, this study provides valuable insights and methods for optimizing energy systems in smart senior living community buildings, contributing to improved energy use efficiency and quality of life for users. This study is mainly divided into five parts. The first part is an overview of the overall research. The second part is a review of research by relevant scholars, both domestically and internationally. The third part is divided into two sections. In the first section, the construction of a smart elderly care community building energy system model is discussed, while the second section outlines research on adaptive optimization of smart building energy systems based on DDPG. In the fourth part, the results of experimental verification conducted using the DDPG algorithm and adaptive optimization of building energy systems are discussed. The fifth part is a summary of this study and prospects for future research. The present study provides specific adaptive optimization analysis of energy systems by building mathematical models and using Markov decision-making with DDPG algorithms, thus providing insight and specific practical solutions that can help promote the development of energy systems in smart buildings.

## 2. Related Work

Rao and other scholars [7] proposed a deep reinforcement learning optimization method for the DDPG algorithm to allocate communication resources. This can solve the problem of uneven resource allocation in communication networks. The study introduced an adaptively adjusted entropy coefficient through the principle of maximum strategy entropy. Then, research was conducted to enhance the search for resource allocation strategies and accelerate convergence. In addition, the research transformed resource allocation problems into Markov decision processes and output allocation plans by constructing an allocation strategy network (SN). Experiments showed that the improved DDPG algorithm could significantly address the problem of source allocation in relevant networks. Relative

to other methods, this algorithm has faster speed and more stable training process. Li and Yu [8] proposed an adaptive controller for better controlling the air flow rate of fuel cells. This controller is used to control the air flow. Based on the DDPG algorithm, a multirole exploration optimization DDPG algorithm was proposed and named MESD-DDPG. The explorers in this algorithm have different exploration principles to improve exploration efficiency. The presented MESD-DDPG algorithm also treats the PI coefficient in the control target. In the experimental results, compared to traditional DDPG controllers, MESD-DDPG controllers showed higher performance and better applicability. Yang et al. [9] proposed a new reconfigurable intelligent surface assisted downlink transmission framework. The research introduced an object migration automation algorithm for the long-term stochastic optimization problem. This algorithm divides users into equal-sized clusters. In addition, to optimize the matrix, the DDPG algorithm was used to synergistically control multiple reflection elements. Unlike traditional algorithmic models, a long-term self-tuning learning model was designed to automatically explore and learn the best actions for each state. A large amount of experimental simulation data showed that the reconfigurable intelligent surface-assisted downlink transmission framework proposed by the research institute increased the data transmission rate compared to traditional orthogonal multiple access frameworks. In addition, the DDPG algorithm proposed in the study enables the model to dynamically allocate resources in a long-term learning manner. Parvin et al. [10] found that the meaning of building energy management (EM) becomes gradually essential in global emissions and climate change issues. Building EM can reduce overall electricity bills and carbon emissions for residents and has a positive impact on both residents and the environment. The study reviewed traditional methods in this area, highlighting their advantages and disadvantages. It also evaluated aspects such as comfort management and other related issues, outlined the application of different relevant algorithms in constructing EM, and discussed the roles of controllers and related algorithms in constructing EM. This study discussed existing methods to determine future research directions. The study provides a reference for optimizing building EM. Feng et al. [11] used an improved algorithm to build an optimization model to optimize the structural characteristics and building shape of the furnace. The energy simulation of the entire architectural model was completed using RIUSKA software, and the Manta Ray Forging Optimizer algorithm was utilized to implement adaptive optimization. The simulation indicated that the method had more excellent simulation performance compared to traditional particle swarm optimization algorithms. In addition, the trapezoidal and matrix structures in building structures could be optimized to minimize energy calculation errors. Lal et al. [12] proposed the use of the autonomous particle swarm optimization algorithm to optimize the stability of wind power generation energy systems. The stability problem was transformed into the optimization problem in the autonomous particle swarm optimization algorithm. The eigenvalues of target work such as damping ratio were established and visualized. The research compared the control performance of the autonomous particle swarm optimization algorithm and the particle swarm optimization algorithm and verified the feasibility of the method.

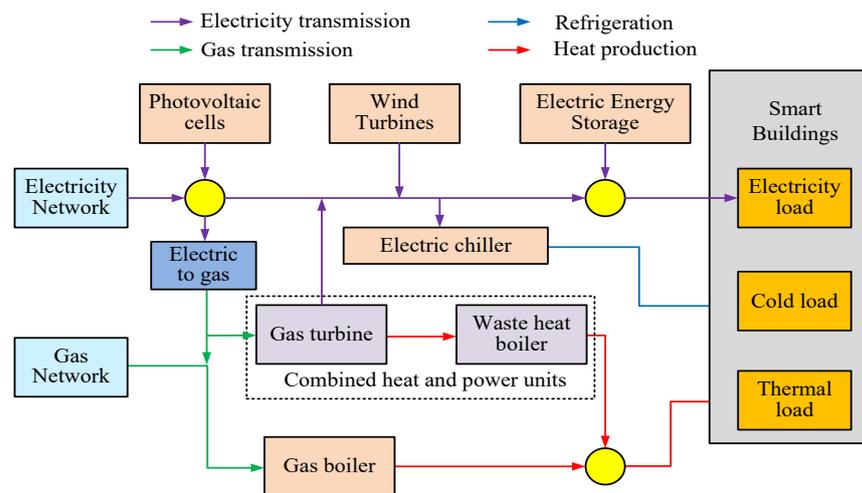
In the literature provided, various studies have focused on optimization techniques in different domains, such as communication resources, airflow control in fuel cells, downlink transmission frameworks, building energy management, and wind energy systems. Notably, there has been an emphasis on deep reinforcement learning, particularly using DDPG algorithms, and autonomous particle swarm optimization algorithms. However, there seems to be a gap in the application of these advanced optimization techniques by specifically tailoring to the technical aspects of smart elderly care communities (SECCs) and energy management systems (EMSs). Most studies have focused on their respective domains without an explicit application or consideration of the unique requirements and challenges in optimizing energy management in smart elderly care environments. This highlights a need for research that is dedicated to leveraging these optimization techniques to address the specific constraints and requirements in SECCs and EMSs. In summary, many studies have shown that the DDPG algorithm has the characteristics of good adapt-

ability, superior performance, simple operation, and easy control. Therefore, compared with other deep learning algorithms, the DDPG algorithm has been widely used in various fields of adaptive optimization problems [13,14]. Adaptive optimization of the ES is crucial for achieving low-carbon living and reducing greenhouse effects. In current ES research, most scholars have limited themselves to optimizing the mechanical structure of the system. Therefore, they have overlooked the fact that adaptive optimization can also introduce new intelligent algorithms for enhancing the intelligence level of the ES and better meet the needs of users. Based on this background, this study will improve the DDPG algorithm in deep learning models and apply the improved model for adaptive optimization of intelligent building ESs. This is expected to provide new ideas for energy saving design of ESs.

### 3. Smart Elderly Care Community Building Energy System Design

#### 3.1. Building Energy System Model Construction for Smart Elderly Care Community

As comprehensive buildings, SECC buildings also have a comprehensive ES composition. In the integrated ES of intelligent buildings, various energy devices often achieve collaborative control and energy coupling between various energy sources through mutual cooperation. This further improves energy utilization efficiency and promotes renewable energy utilization [15,16]. The internal composition of the ES in SECC buildings is showcased in Figure 1.



**Figure 1.** Integrated energy system in smart elderly care community buildings.

Figure 1 illustrates the comprehensive ES in SECC buildings. The entire comprehensive ES is separated into four parts: energy production (EP), energy conversion (EC), energy storage (EST), and energy consumption [17]. The EP part mainly consists of the power grid and gas grid. The power grid includes renewable energy installations, such as photovoltaic cells (PCs) and wind turbines, which can generate various types of energy. For example, natural gas can generate heat and electricity, while electricity can generate cold and gas energy. The EC part mainly involves the conversion of various sources through various electrical equipment [18]. For example, electrical energy can be converted into natural gas through an electric to gas conversion device (P2G). Electric energy can also be converted into cold energy through an electric refrigerant. In a complex building system, it is necessary to use equipment such as electric to gas (P2G), electric refrigeration (ER), gas turbine (GT), gas boiler (GB), waste heat boiler (WHB), etc. to complete system power supply and EC. This can achieve energy complementarity and tiered utilization of multiple intelligent building ES. Recently, intelligent building energy systems have been equipped with multiple energy storage systems. To facilitate energy calculation, this study only considers the energy storage scenario of electric energy storage (EES) and is therefore not suitable for distributed storage [19]. Electric EST, as a backup energy source for the electrical ES, can not only absorb

excess energy in the building ES but also provide timely energy supply when the system needs energy. The energy consumption part mainly refers to the energy consumption demand of various electrical and energy-consuming equipment within the building system.

In real life, the ES of intelligent buildings often have complex structures and numerous parameters. Therefore, there are also significant fluctuations in load demand at different times. Therefore, to achieve adaptive optimization of the final building ES, this study first elaborates on the mathematical models of various energy supply and EST equipment in the building ES. As one of the energy supply devices, the power output formula of PCs is shown in Equation (1) [20].

$$P_t^{PV} = P_{STC} \frac{G_T}{G_{STC}} (1 - K(\theta_C - \theta_\tau)) \quad (1)$$

In Equation (1),  $P_t^{PV}$  represents the output power (OP) of the PC at different times, and  $P_{STC}$  represents the maximum OP of the PC under standard test conditions. The standard test conditions refer to the ambient temperature of 25 °C and solar radiant intensity of 1000 W/m<sup>2</sup>.  $G_{STC}$  and  $G_T$  represent standard light intensity and solar radiant intensity, respectively;  $K$  represents the coefficient of power temperature change, with a fixed value of  $-0.47\%/^{\circ}\text{C}$  taken in this study; and  $\theta_C$  and  $\theta_\tau$  represent the surface temperature and standard temperature of the PC, respectively.

$$\theta_C = \theta_H + 30 \frac{G_T}{1000} \quad (2)$$

Equation (2) is the calculation formula for the surface temperature of PCs.  $\theta_H$  represents the ambient temperature.

$$P_t^{WT} = \begin{cases} 0 & v_{co} < v < v_{ci} \\ 0.5\rho Av^3 C_p & v_{ci} \leq v < v_r \\ P_{max}^{WT} & v_r \leq v \leq v_{co} \end{cases} \quad (3)$$

Equation (3) is the calculation formula for the OP of the fan.  $P_t^{WT}$  represents the OP of the fan at time  $t$ ;  $v_{co}$ ,  $v_r$ , and  $v_{ci}$  represent the cut-out, rated, and cut-in wind speeds of the fan, respectively;  $A$  represents the swept area; and  $C_p$  represents the wind EC coefficient with a value of 0.4.  $\rho$  is the density of air, and the density value under standard environmental conditions is 1.29 kg/m<sup>3</sup>.

$$V_t^{P2G} = \frac{\eta_{P2G} P_t^{P2G}}{H_{gas}} \quad (4)$$

Equation (4) is the calculation formula for the amount of natural gas generated by the electric to gas equipment.  $V_t^{P2G}$  represents the natural gas content generated by the electric to gas conversion equipment at time  $t$ ,  $P_t^{P2G}$  represents the relevant power consumed by the electric to gas equipment at time  $t$ ,  $\eta_{P2G}$  is the EC efficiency of the electric to gas equipment,  $H_{gas}$  represents the thermal energy generated by a unit of natural gas with a value of 9.7 kWh/m<sup>3</sup>.

$$H_t^{GB} = V_t^{GB} \eta_{GB} H_{gas} \quad (5)$$

Equation (5) is the calculation formula for the heat generation power (HGP) of GB.  $H_t^{GB}$  represents the HGP of the GB at time  $t$ ,  $V_t^{GB}$  represents the natural relevant consumption of the GB at time  $t$ , and  $\eta_{GB}$  is the heating efficiency.

$$Q_t^{EC} = P_t^{EC} \eta_{EC} \quad (6)$$

Equation (6) is the calculation formula for the cooling power of an electric refrigerator (ER).  $Q_t^{EC}$  represents the ER's cooling power at time  $t$ ,  $P_t^{EC}$  represents the power consump-

tion of the ER at time  $t$ , and  $\eta_{EC}$  is the cooling efficiency. Gas-fired boilers (GB) can be combined with GT to form a combined thermal power plant (CHP). The cogeneration unit can meet the electrical and thermal load requirements of the entire building ES.

$$P_t^{GT} = V_t^{GT} H_{gas} \eta_{GT} \quad (7)$$

Equation (7) is the calculation formula for the power generated by gas turbines (GT).  $P_t^{GT}$  represents the power generated by the GT at time  $t$ ,  $V_t^{GT}$  represents the natural gas consumption of the GT at time  $t$ , and  $\eta_{GT}$  represents the power generation efficiency.

$$H_t^{GT} = P_t^{GT} \frac{1 - \eta_{GT} - \varepsilon_{GT}}{\eta_{GT}} \quad (8)$$

Equation (8) is the thermal energy calculation formula for GT.  $H_t^{GT}$  represents the thermal energy generated by the combustion of natural gas for power supply by the GT at time  $t$ , and  $\varepsilon_{GT}$  serves as the heat loss coefficient.

$$H_t^{WHB} = H_t^{GT} \eta_{WHB} \quad (9)$$

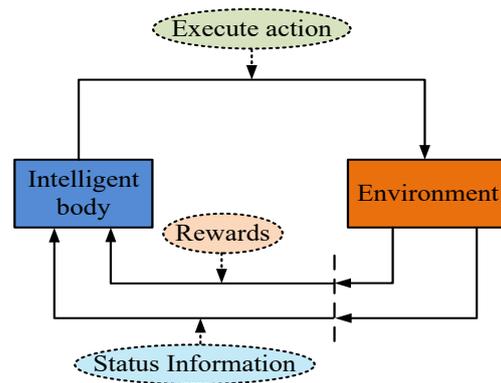
Equation (9) is the calculation formula for heat generation of waste heat boilers.  $H_t^{WHB}$  represents the heat generated by the waste heat boiler using waste heat at time  $t$ , and  $\eta_{WHB}$  represents the heating efficiency of the waste heat boiler.

$$E_t^{EES} = (1 - a_{EES}) E_{t-1}^{EES} + \left( \eta_{in}^{EES} P_{t,in}^{EES} - \frac{P_{t,out}^{EES}}{\eta_{out}^{EES}} \right) \Delta t \quad (10)$$

Equation (10) is the mathematical model of electric EST equipment [21].  $E_t^{EES}$  serves as the electrical energy stored in the electrical EST device at time  $t$ ;  $\eta_{in}^{EES}$  and  $\eta_{out}^{EES}$  represent charging and discharging efficiency, respectively;  $P_{t,in}^{EES}$  and  $P_{t,out}^{EES}$  represent the charging and discharging power of the electric EST device at time  $t$ , respectively; and  $a_{EES}$  represents the electrical energy loss coefficient.

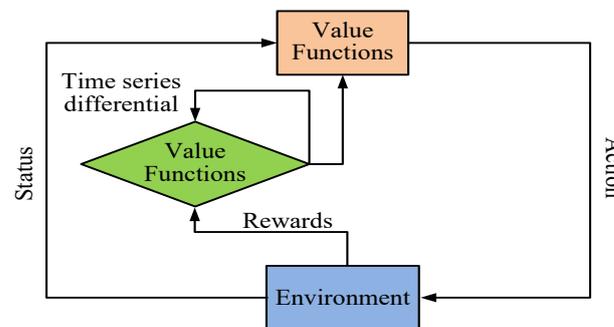
Through the above mathematical model, a complete intelligent elderly care community building ES can be built [22]. In the control and optimization of energy systems, model predictive control (MPC) and RL are both popular methods. However, in smart retirement community building energy systems, RL has better adaptability and robustness due to the uncertainty and dynamics of the processing environment. Therefore, this study uses reinforcement learning to build a relevant model in order to perform adaptive optimization of the energy system. The basic idea of reinforcement learning is to acquire optimal action strategies through mutual learning between intelligent agents and the external environment. In reinforcement learning, if the state of an agent at a certain moment is only related to the state of the previous moment, the corresponding learning state at that moment is considered to satisfy Markov properties. If the entire learning process satisfies Markov properties at all times, then the learning process is called a Markov process.

Figure 2 indicates the interaction in the intelligent agent and the external environment in the Markov process. The goal of intelligent agents interacting with the external environment is to obtain the maximum cumulative reward. Therefore, in general, actions at a certain moment do not affect the final cumulative result, and the model can modify long-term rewards by changing subsequent strategies [23]. Markov processes can describe almost all problems and have good universality. However, as the dimension of the state space is high, intelligent algorithms still need to be introduced to compensate for the shortcomings of this method in dealing with high dimensional state and continuous space problems.



**Figure 2.** The interaction process between intelligent agents and the external environment in Markov processes.

Figure 3 indicates the general framework of the actor–critic (AC) algorithm. This algorithm has fast training speed, small estimation variance, and high sample utilization. This study takes it as a general framework and ultimately uses the DDPG algorithm under the AC framework for addressing the model.



**Figure 3.** General framework flowchart of the actor–critic algorithm.

### 3.2. Research on Adaptive Optimization of Intelligent Building Energy Systems Based on the DDPG Algorithm

Currently, most of the optimization methods proposed for the uncertainty of dual supply and demand in intelligent buildings rely on the accuracy of modeling the uncertainty of renewable energy OP and load. It is difficult to accurately account for the load fluctuations in actual situations, and it does not take into account the thermal comfort of internal users in the building. Based on this current situation, this study considers the relationship between more random variables in the ES of intelligent buildings and the thermal comfort needs of internal users. This study uses Markov decision processes to jointly model various energy and thermal comfort demand problems and solves the model using the DDPG algorithm under the AC framework [24].

This research uses the DDPG algorithm of deep reinforcement learning to solve the optimization problem in the energy system of intelligent buildings, especially considering the uncertainty of bilateral supply and demand and the thermal comfort of users inside the building. The decision variable type is a continuous real variable, which includes environmental state inputs such as photovoltaic output, wind power output, and electricity load demand. The constraints include boundary constraints, such as upper and lower limits for thermal and cold energy outputs, and inequality constraints. In addition, this study also takes into account the operating efficiency constraints of the equipment, user thermal comfort requirements, and power quality limitations [25]. To simplify the computational complexity of the ES, this study assumes that the thermal energy in intelligent buildings is provided by gas turbines (GT) and waste heat boilers (WHB), while the cold energy is provided by ER. Both thermal and cold energy are only used to regulate the

indoor temperature of buildings to ensure thermal comfort for users, and there is no other consumption of thermal and cold energy. The above assumption is to ensure that there is only a single heating or cooling load demand in intelligent buildings.

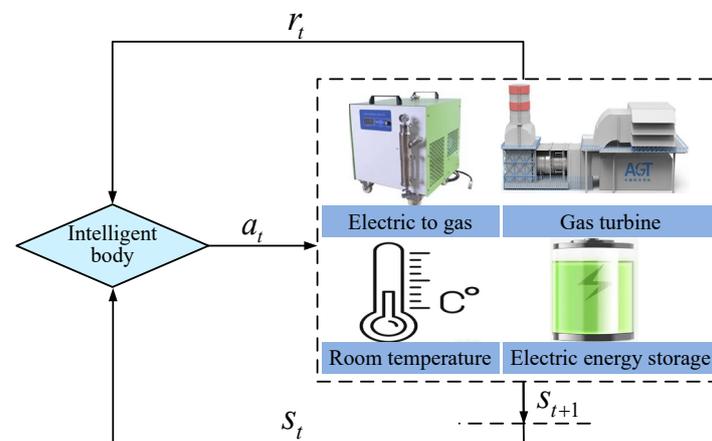
$$T_t^{in} = T_{t-1}^{in} e^{-\frac{1}{RC_{air}}} + \left( RE_t^{air} + T_t^{out} \right) \left( 1 - e^{-\frac{1}{RC_{air}}} \right) \quad (11)$$

Equation (11) represents the simplified thermodynamic mathematical model of intelligent buildings [26].  $T_t^{in}$  and  $T_t^{out}$  represent the indoor and outdoor temperatures of the building at time  $t$ , respectively;  $R$  serves as the equivalent thermal resistance of materials in a certain area inside the building; and  $C_{air}$  represents equivalent heat capacity;  $E_t^{air}$  serves as the heat supply of the area at time  $t$ . If  $E_t^{air} > 0$ , it indicates the input of thermal energy and the temperature of the area will increase accordingly. If  $E_t^{air} < 0$ , it indicates that the input of cold energy will cause the temperature of the area to decrease accordingly. By increasing or decreasing the temperature, the user's thermal comfort is ensured.

In the ES of intelligent buildings, the indoor temperature at the next moment is only related to the current indoor temperature, outdoor temperature, and equipment operation. The operating state of the system at the next moment is only related to the system status and equipment operation at the previous moment. Therefore, Markov decision processes can be used to jointly model and solve user thermal comfort problems with an intelligent building ES.

$$M = (S, A, P, R, \gamma) \quad (12)$$

Equation (12) describes the Markov decision mathematical process used by the research institute [27].  $S$ ,  $A$ ,  $P$ , and  $R$  represent the environmental state set, action space set, state transition matrix, and reward function, respectively, and  $\gamma$  represents the discount factor. The relevant process in the intelligent building ES agent and the external environment is shown in Figure 4.

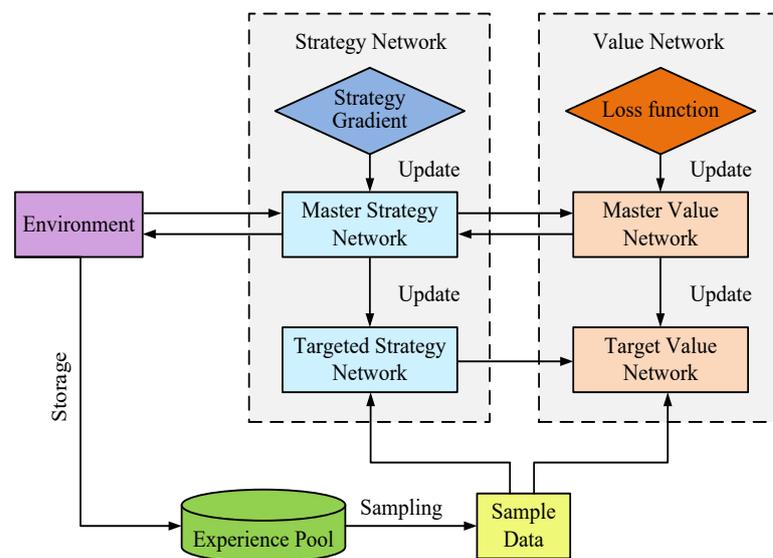


**Figure 4.** Interaction flowchart between the intelligent agent and the external environment.

In Figure 4, the agent learns the optimal learning action strategy by continuously interacting with the external environment. In this study, the intelligent building ES is the interactive environment for intelligent agents. The information that can be exchanged includes the OP of the device, the OP of renewable energy, the status of electrical EST, indoor and outdoor temperatures, load demands, etc. In Figure 4,  $t$  represents the time,  $S_t$  represents the current environmental status information received by the intelligent agent at time  $t$ ,  $a_t$  serves as the execution action at time  $t$ , and  $r_t$  serves as the reward at moment  $t$ . Through continuous learning, intelligent agents can maximize cumulative rewards and make optimal strategies.

The indoor temperature and OP of various equipment are influenced by various factors. Therefore, it is difficult to accurately describe the changing data using the state transition probability matrix. For continuous state space and action space problems, the

DDPG algorithm is studied for optimization. The relevant framework structure is indicated in Figure 5.

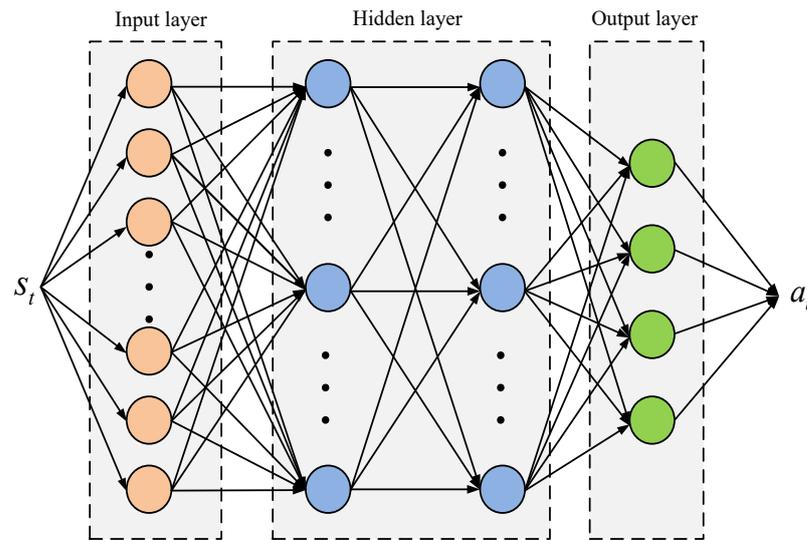


**Figure 5.** Framework structure diagram of the deep deterministic policy gradient algorithm.

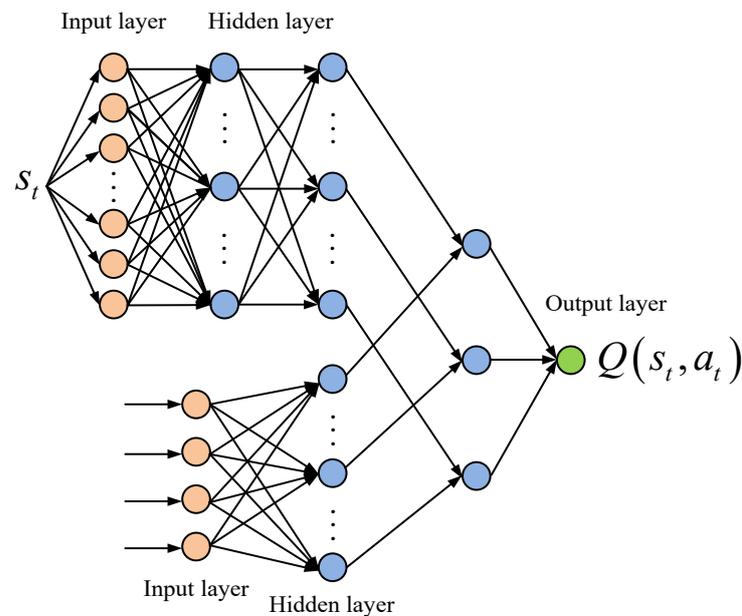
Figure 5 indicates the DDPG algorithm includes four parts: environment, experience pool, SN, and value network (VN). In the SN, there are mainly three parts: strategy gradient, main SN, and target SN. In the VN, there are mainly three parts: loss function, main VN, and target VN. In the DDPG algorithm, the SN and VN are used to approximate the strategy function and value function, respectively. To prevent fluctuations in the learning process of the network due to parameter updates, a target network was set for the SN and the VN, and the learning reward value was ultimately obtained through the target network.

Figure 6 demonstrates the related structure diagram in the DDPG algorithm. Figure 6 illustrates that the entire policy network includes one input layer (IL), two hidden layers, and one output layer. The input of the algorithm is the environmental state, including photovoltaic output, wind power output, and electrical load demand. The output is an action waiting to be executed. The input data comes from the data generated by EnergyPlus software to simulate the operation of intelligent buildings, including power load demand, outdoor temperature, wind speed, solar radiant intensity, etc. After the intelligent agent executes an action, the environment generates a state transition and receives corresponding reward values. By continuously training the SN, the final execution action can no longer add noise and ultimately achieve a predetermined effect. This study selected the environmental status of a smart elderly care community building energy system, including photovoltaic output, wind power output, and electrical load demand, for training. Environmental state parameters were collected from historical records of the energy system. By analyzing these historical data, we could understand their distribution characteristics and randomly sample them to simulate various environmental conditions.

Figure 7 showcases the structure diagram of the VN in the DDPG algorithm. Figure 7 depicts that the basic structure of the VN is similar to that of the SN, but the input terms of its IL are no longer limited to the OP and environmental parameters of each device. It also includes the execution actions of the policy network output. The output of the VN obtained through continuous training is a set composed of execution actions and environmental parameters. It uses Markov decision-making and DDPG algorithm to optimize the mathematical model of the ES built above and observes and analyzes the effectiveness of adaptive optimization of the ES.



**Figure 6.** Structure diagram of the policy network in the DDPG algorithm.



**Figure 7.** Structure diagram of the value network in the DDPG algorithm.

## 4. Analysis of Adaptive Optimization Results for Building Energy Systems Based on the DDPG Algorithm

### 4.1. Algorithm Performance Analysis

To simulate the stochasticity of the parameters of the intelligent building thermal dynamics model, this study assumed that the model parameters follow Gaussian distribution. The initial temperature and appropriate temperature were set to 21 °C, the minimum indoor temperature was 18 °C, the maximum indoor temperature was 24 °C, the thermal discomfort penalty factor was 1 CNY/kW, the environmental interaction and learning ratio was 6, the discount factor was 0.95, and the capacity of the experience pool was 5000. The remaining relevant experimental parameters are shown in Table 1.

**Table 1.** Experimental parameters and their values.

Parameter	Numerical Value
Building thermal-equivalent resistance	$N\sim(18, 0.5)$
Building thermal-equivalent capacitance	$N\sim(0.525, 0.1)$
Overlay noise	$N\sim(0.5, 0.2)$
Explore factors	$U\sim(0, 1)$
Value network learning rate	$1.00 \times 10^{-04}$
Number of hidden layer neurons of the value network	[300,600,600,300]
Policy network learning rate	$1.00 \times 10^{-04}$
Number of hidden layer neurons in the policy network	[300, 600]
Target network update factor	$1.00 \times 10^{-03}$
Initial proportion of the superimposed noise	1
Superimposed noise decay factor	0.997
Superlay noise proportion minimum	0.05

To demonstrate the merits of the DDPG algorithm, the public UMass Smart Dataset from Massachusetts Institute of Technology was selected for testing. This dataset includes various senior living community buildings, daily household electricity load requirements, indoor and outdoor temperatures, and related environmental data. In this study, it was divided into training and testing datasets, and simulation experiments were completed on the MATLAB experimental platform to analyze the performance of different algorithms in the same dataset.

Figure 8 depicts the relevant curves of different algorithms under different rewards. Figure 8a depicts the relevant curve of the traditional DDPG algorithm under single reward and average reward. As the number of episodes increases, the single reward curve and average reward curve of the DDPG algorithm show a continuous upward trend. Before 200 episodes, the algorithm experiences significant fluctuations, and as the number of episodes increases, the fluctuation range of the algorithm gradually decreases. Figure 8b demonstrates the iterative convergence curves of the traditional Deep Q Network (DQN) algorithm under single and average rewards. As the number of episodes increases, the single reward curve and average reward curve of the DQN algorithm also shows a continuous upward trend. However, relative to the DDPG algorithm, the fluctuation of the DQN algorithm is larger in amplitude and frequency. This depicts that the DDPG algorithm has better stability.

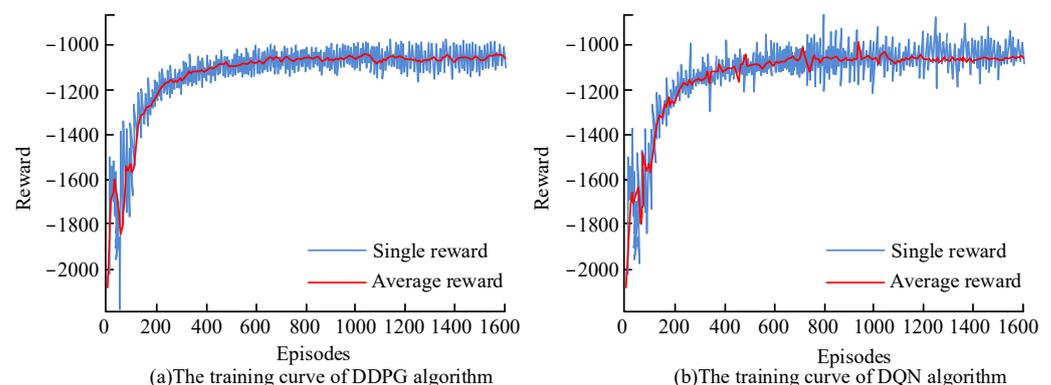
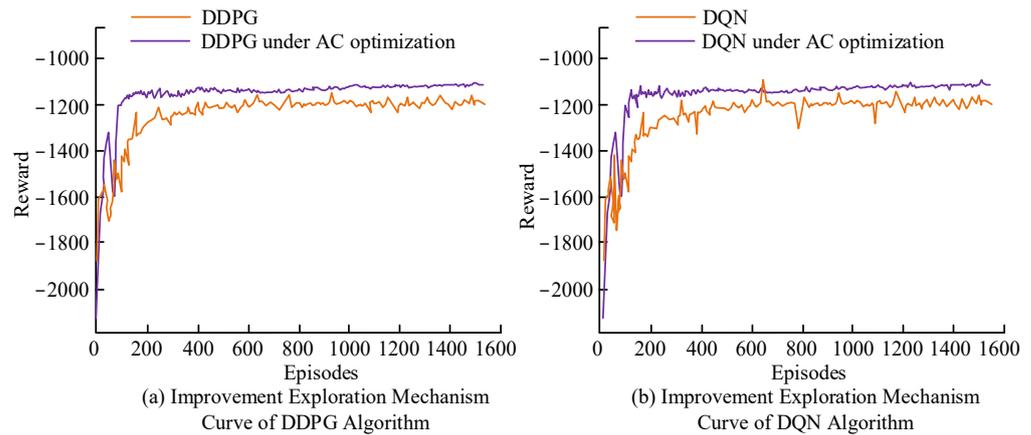
**Figure 8.** Reward convergence iterative curves of DDPG and DQN algorithms.

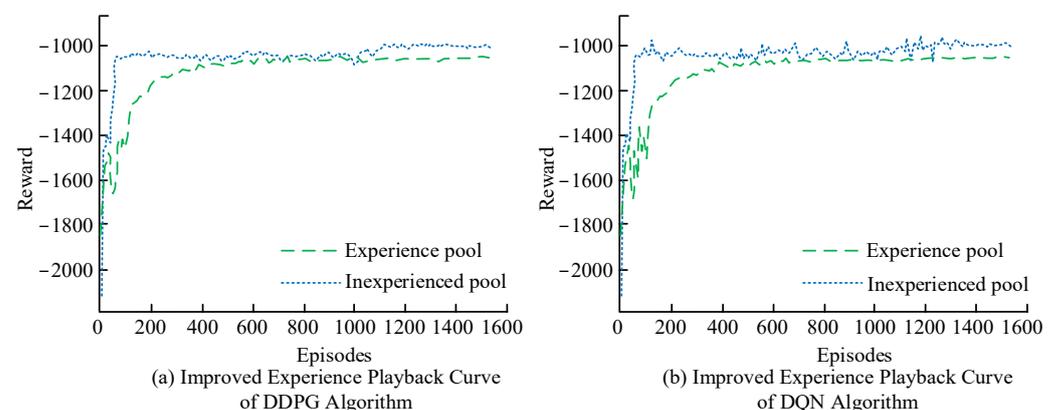
Figure 9 showcases the variation of the improved exploration mechanism curves for different algorithms. Figure 9a showcases that as the number of episodes increases, both the DDPG algorithm and the DQN algorithm under the AC framework show an upward trend. Compared to traditional DDPG algorithms, the improved DDPG algorithm based on the AC framework can achieve faster convergence and has better stability. When the number of episodes of the model is 200 times, the DDPG algorithm under the AC framework begins

to reach a stable reward state, with a reward value of  $-1186$ . Figure 9b indicates that as the number of episodes increases, both the DQN algorithm and the DQN algorithm under the AC framework show an upward trend. Compared to traditional DQN algorithms, the improved AC framework based on the DQN algorithm has better stability and convergence. When the number of episodes of the model is 200 times, the algorithm begins to converge. The reward value of the DQN algorithm under the AC framework when it reaches a stable reward state is  $-1182$ . In addition, the DQN and the DQN under the AC framework have more fluctuations and poor stability relative to the DDPG and the DDPG algorithm under the AC framework.



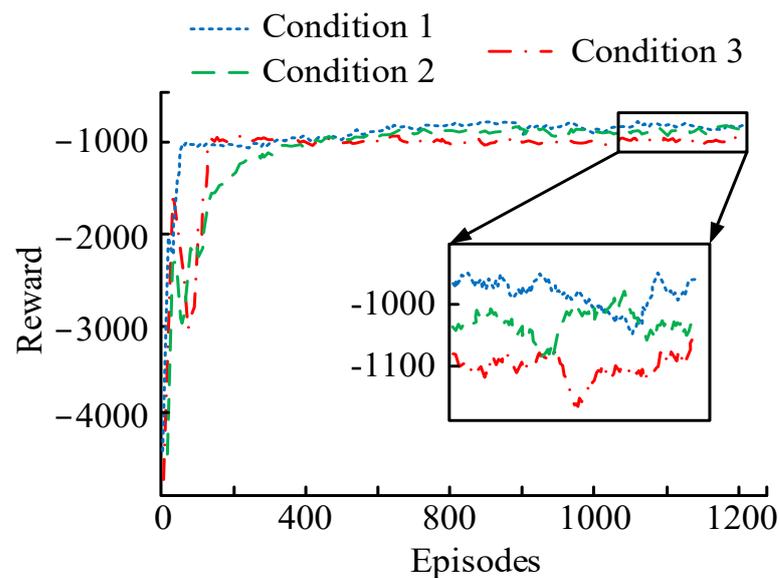
**Figure 9.** Improved exploration mechanism curves for different algorithms.

Figure 10 showcases the improved experience playback curves under different algorithms. Figure 10a showcases that with an experience pool, the DDPG algorithm can begin to stabilize after 400 episodes. When no experience pool is set, the DDPG algorithm begins to stabilize at 85 episodes. Comparing the two, it can be seen that not setting an experience pool can accelerate the convergence of the model and enable it to converge to a stable reward state as soon as possible. However, as the episodes progress, models without an experience pool begin to experience fluctuations, and their stability is poorer compared to models with an experience pool. In summary, adding an experience pool during model training can make the convergence process of the algorithm more stable. Figure 10b showcases that with an experience pool, the DQN can begin to stabilize after 600 episodes. When no experience pool is set, the DDPG algorithm begins to stabilize at 70 episodes. In an environment without an experience pool, compared to the DDPG, the fluctuation amplitude of the DQN algorithm is greater. In summary, the analysis of the results in Figures 8–10 fully demonstrates that the stability and convergence of the DDPG algorithm are better than the other commonly used deep learning algorithm DQN.



**Figure 10.** Improved empirical playback curves for different algorithms.

Figure 11 depicts the related curves of the improving DDPG algorithm under various conditions. Condition 1, 2, and 3 in Figure 11 refer to coefficients of 4, 8, and 12 for environmental interaction and learning ratio, respectively. Figure 11 depicts that when the coefficient of environmental interaction and learning ratio is small, the improved DDPG algorithm under the AC framework can achieve faster convergence and stable reward values. When the coefficient of environmental interaction and learning ratio is large, the improved DDPG algorithm under the AC framework requires more training to complete convergence. When the number of network training reaches about 1200, the final convergence of the improved DDPG algorithm with large coefficients is worse than that with small coefficients.



**Figure 11.** Convergence curves of the improved DDPG algorithm under different conditions.

#### 4.2. Application Effect Analysis

To further demonstrate that the improved DDPG algorithm under the AC framework can jointly optimize the ES of intelligent buildings and user thermal comfort, the study selected daily load data, temperature data, etc. in the above test dataset and performed optimization testing using SN and VN. Considering the significant temperature difference between different seasons, the study first divided the entire year into two parts: the heating season and the cooling season. Then, this study analyzed the load changes, wind power output (WPO), and photovoltaic output (PO) during the heating and cooling seasons.

Figure 12 depicts the load changes under different conditions. Figure 12a depicts that as the average daily usage time increases, the intelligent building system has almost no cooling load demand during the heating season, while the demand for electricity and heating loads depicts a constantly changing trend. Figure 12b depicts that as the average daily usage time increases, the intelligent building system has almost no heating load demand during the cooling season. The total demand for electricity and heating loads shows a continuous upward trend.

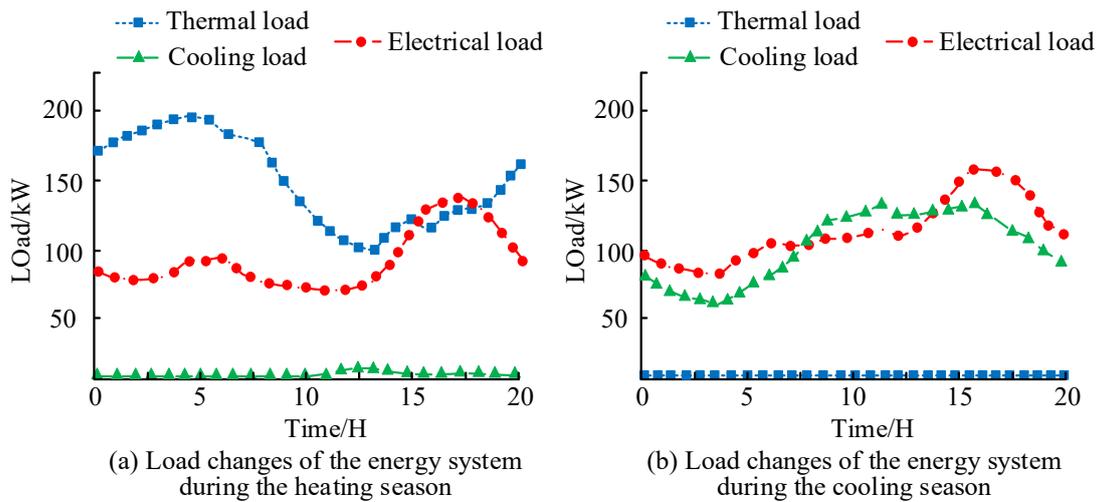


Figure 12. Variation of load under different conditions.

Figure 13 indicates the changes in WPO and PO under different conditions. Figure 13a indicates that as the average daily usage time increases, the WPO of the intelligent building system during the heating season indicates a continuous upward trend, while its PO indicates a continuous downward trend. Figure 13b indicates that as the average daily usage time increases, the WPO and PO of the intelligent building system change during the cooling season. The PO is 0 both before and after 5 h of daily use but indicates a trend that first increases and then decreases in the middle. The WPO fluctuates between 20 and 70 kW. From Figures 12 and 13, it can be observed that seasonal changes have an essential influence on the ES of intelligent buildings. The electricity load, cooling load, heating load, PO, and WPO vary in different seasons.

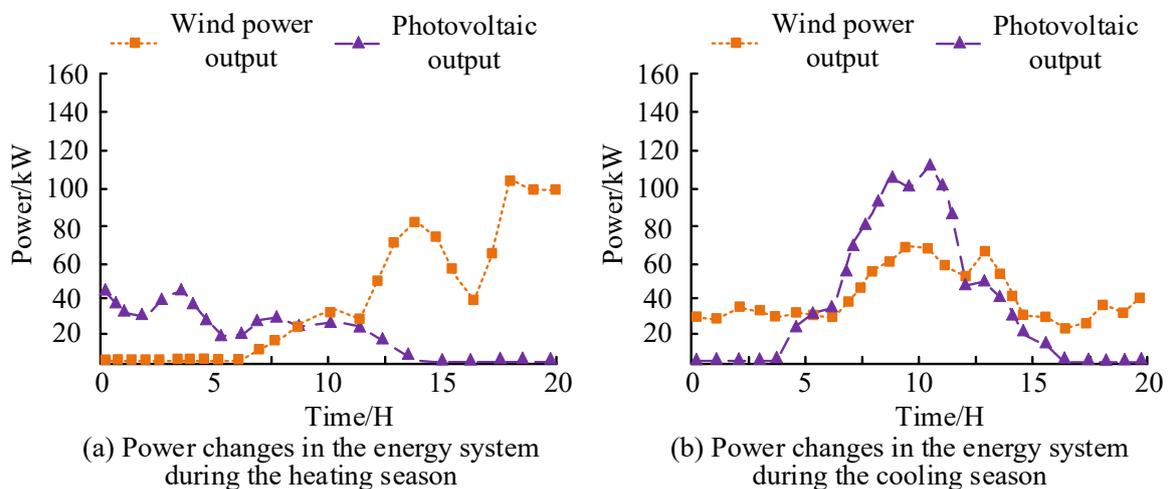
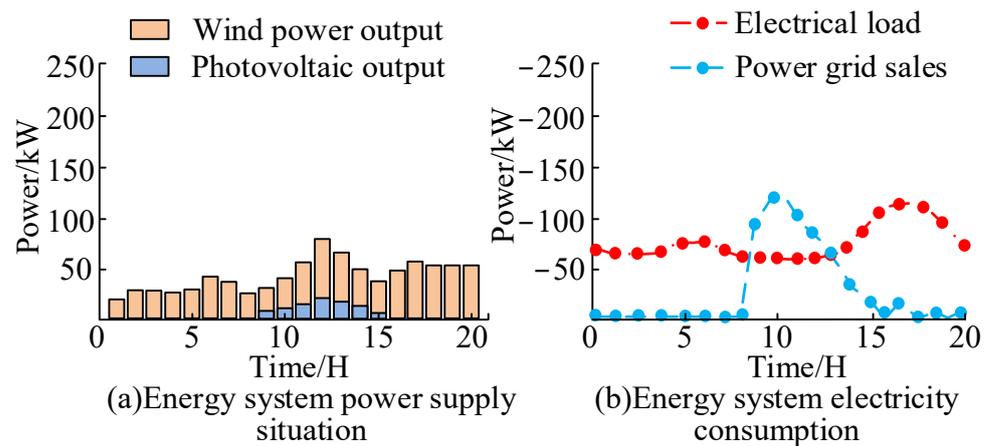


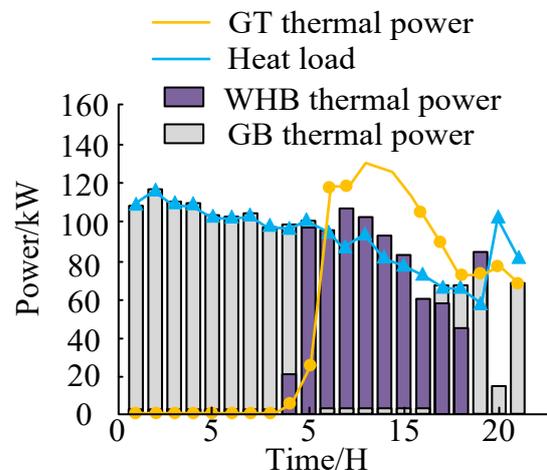
Figure 13. Variation of wind power output and PV output under different conditions.

Figure 14 illustrates the power supply and consumption of the intelligent building ES optimized by the improved DDPG algorithm. Figure 14a illustrates that when the scheduling cycle of the ES is between 0:00 and 8:00, the PO of the system is 0, and only the WPO is used to provide the energy required by users. When the scheduling cycle of the ES is between 9:00 and 15:00, the energy supply of the system is jointly provided by PO and WPO. After 15:00, the system is powered by WPO. Figure 14b illustrates that when the scheduling period of the ES is between 0:00 and 20:00, the electrical load demand of the system is between  $-50$  and  $-150$  kW. The electrical energy provided by the system at each stage can basically meet the thermal comfort needs of users.



**Figure 14.** Intelligent building energy system power supply and power consumption.

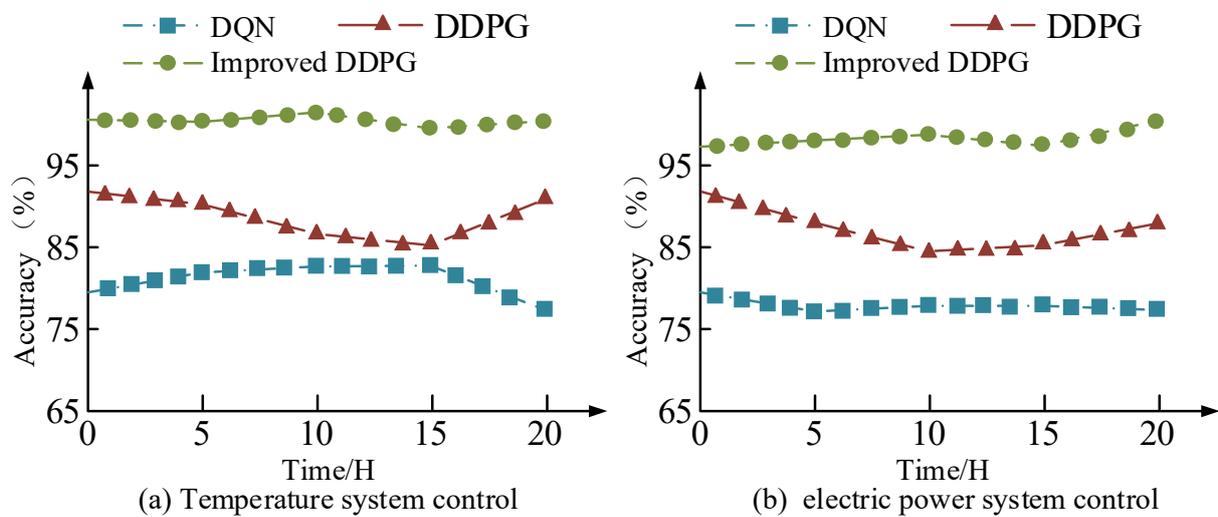
Figure 15 illustrates the heating and heat consumption situation of the intelligent building ES optimized by the improved DDPG algorithm. Figure 15a illustrates that when the scheduling period of the ES is between 0:00 and 9:00, the user's heat load demand is basically met by the GB. Waste heat boiler (WHB) produces almost no new thermal energy. When the scheduling cycle of the system is after 9:00, the waste heat boiler begins to generate a large amount of thermal energy for satisfying the thermal comfort requirements of users. This is because of the continuous generation of electricity by GT, accompanied by a large amount of heat generation being recycled by WHB to satisfy the user's needs.



**Figure 15.** Intelligent building energy system heat supply and heat consumption.

The accuracy of adaptive regulation of intelligent building energy systems is shown in Figure 16. Figure 16a shows the comparison of control accuracy for temperature systems. From the graph, it can be seen that the average accuracy of DQN regulation is 81.42%, the average accuracy of DDPG regulation is 88.23%, and the average accuracy of improved DDPG regulation is 98.45%. Compared to DQN and DDPG, the improved DDPG algorithm has a significant increase in regulation accuracy, which confirms its superiority in temperature adaptive regulation of intelligent building energy systems. Figure 16b shows the comparison of regulation accuracy for the power system. From the graph, it can be seen that the average accuracy of DQN regulation is 78.11%, the average accuracy of DDPG regulation is 87.66%, and the average accuracy of improved DDPG regulation is 96.09%. Compared to DQN and DDPG, the improved DDPG algorithm has significantly improved the accuracy of power system regulation, verifying the effectiveness of this study. In summary, the improved DDPG algorithm used in this study can take into account the interaction between various parameters in the building dynamics model, thus achieving

the goal of adaptive optimization. It can also consider the thermal comfort needs of users while meeting the energy supply goals of the system.



**Figure 16.** Accuracy comparison of adaptive regulation of energy systems in intelligent buildings.

## 5. Conclusions

With the acceleration of population aging, the quality of life and health status of the elderly are increasingly receiving social attention. This research used the DDPG algorithm with deep reinforcement learning to solve the optimization problem of smart building energy systems and specifically considered the uncertainty of bilateral supply and demand as well as the thermal comfort of users inside the building. In addition, the study considered the environmental state input as well as several constraints such as the operational efficiency constraints of the equipment and the thermal comfort demand of the users so as to achieve optimization of the smart building energy system in a practical and complex scenario. The outcomes indicate that the DDPG algorithm has a large fluctuation range before 200 training sessions. As the quantity of training grows, the fluctuation range of the algorithm gradually diminishes. When the coefficient of environmental interaction and learning ratio is 4, the improved DDPG algorithm can converge by iterating 60 times, and the stable reward value in the convergence state is  $-996$ . Compared with the DQN algorithm, the DDPG algorithm has better stability and convergence. By applying the DDPG algorithm to adaptive optimization of building ES, it was found that there are differences in system load changes, WPO, and PO during the heating and cooling seasons. This study analyzed the power supply and consumption of intelligent building ES optimized by the improved DDPG algorithm. When the scheduling cycle is between 0:00 and 8:00, the PO is 0, and the WPO is within 50 kW. When the scheduling cycle is between 9:00 and 15:00, the system is powered by a combination of the two. When the scheduling period is after 15:00, the PO becomes 0 again, but the WPO is within 100 kW. When the scheduling cycle is between 0:00 and 9:00, the user's required heat load demand is supplied by GB, and the GB range is within 120 kW. When the scheduling cycle is after 9:00, GT and GB jointly supply the user's thermal demand. In summary, the DDPG algorithm used in this study has a good adaptive optimization effect in smart building energy systems. However, due to the limited conditions, the study is not a practical experiment on real buildings, so the study is only for reference purposes. Improving the reliability of the study will be the main topic in the future. Future research can explore the robustness of DDPG under various climatic conditions and design strategies to achieve the best performance in different environments. In addition, future research should also explore the application of model predictive control (MPC) and RL in the optimization of energy systems in smart buildings and further analyze the advantages and disadvantages of both.

**Author Contributions:** C.L. and Z.X. collected the samples. C.L. analysed the data. Z.X. conducted the experiments and analysed the results. All authors discussed the results and wrote the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** The research is supported by the Youth Project of the National Natural Science Foundation of China, “Study on the Effect of Social Elderly Care Service Supply and the Mechanism of Classified Provision Based on the Perspective of Health Quality Difference of the Elderly” (No. 71603210), and the General Project of Guangdong Province Philosophy and Social Science Planning, “Research on precise supply of intelligent elderly care services from the perspective of demand heterogeneity” (No. GD22CGL14).

**Data Availability Statement:** The data and code can be given based on the request.

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

## References

1. Wu, J.; Wei, Z.; Liu, K.; Quan, Z.; Li, Y. Battery-Involved Energy Management for Hybrid Electric Bus Based on Expert-Assistance Deep Deterministic Policy Gradient Algorithm. *IEEE Trans. Veh. Technol.* **2020**, *69*, 12786–12796. [[CrossRef](#)]
2. Mei, J.; Wang, X.; Zheng, K.; Boudreau, G.; Bin Sediq, A.; Abou-Zeid, H. Intelligent Radio Access Network Slicing for Service Provisioning in 6G: A Hierarchical Deep Reinforcement Learning Approach. *IEEE Trans. Commun.* **2021**, *69*, 6063–6078. [[CrossRef](#)]
3. Cao, B.; Dong, W.; Lv, Z.; Gu, Y.; Singh, S.; Kumar, P. Hybrid Microgrid Many-Objective Sizing Optimization with Fuzzy Decision. *IEEE Trans. Fuzzy Syst.* **2020**, *28*, 2702–2710. [[CrossRef](#)]
4. Cao, K.; Zhou, J.; Xu, G.; Wei, T.; Hu, S. Exploring Renewable-Adaptive Computation Offloading for Hierarchical QoS Optimization in Fog Computing. *IEEE Trans. Comput. Des. Integr. Circuits Syst.* **2019**, *39*, 2095–2108. [[CrossRef](#)]
5. Melgar-Dominguez, O.D.; Pourakbari-Kasmaei, M.; Mantovani, J.R.S. Adaptive Robust Short-Term Planning of Electrical Distribution Systems Considering Siting and Sizing of Renewable Energy Based DG Units. *IEEE Trans. Sustain. Energy* **2019**, *10*, 158–169. [[CrossRef](#)]
6. Liu, Y.; Wang, X.; Mei, J.; Boudreau, G.; Abou-Zeid, H.; Bin Sediq, A. Situation-Aware Resource Allocation for Multi-Dimensional Intelligent Multiple Access: A Proactive Deep Learning Framework. *IEEE J. Sel. Areas Commun.* **2020**, *39*, 116–130. [[CrossRef](#)]
7. Rao, N.; Xu, H.; Qi, Z.; Song, B.; Shi, Y. Allocation method of communication interference resource based on deep reinforcement learning of maximum policy entropy. *Xibei Gongye Daxue Xuebao/J. Northwestern Polytech. Univ.* **2021**, *39*, 1077–1086. [[CrossRef](#)]
8. Li, J.; Yu, T. A new adaptive controller based on distributed deep reinforcement learning for PEMFC air supply system. *Energy Rep.* **2021**, *7*, 1267–1279. [[CrossRef](#)]
9. Yang, Z.; Liu, Y.; Chen, Y.; Al-Dhahir, N. Machine Learning for User Partitioning and Phase Shifters Design in RIS-Aided NOMA Networks. *IEEE Trans. Commun.* **2021**, *69*, 7414–7428. [[CrossRef](#)]
10. Parvin, K.; Lipu, M.S.H.; Hannan, M.A.; Abdullah, M.A.; Jern, K.P.; Begum, R.A.; Mansur, M.; Muttaqi, K.M.; Mahlia, T.M.I.; Dong, Z.Y. Intelligent Controllers and Optimization Algorithms for Building Energy Management Towards Achieving Sustainable Development: Challenges and Prospects. *IEEE Access* **2021**, *9*, 41577–41602. [[CrossRef](#)]
11. Feng, J.; Luo, X.; Gao, M.; Abbas, A.; Xu, Y.-P.; Pouramini, S. Minimization of energy consumption by building shape optimization using an improved Manta-Ray Foraging Optimization algorithm. *Energy Rep.* **2021**, *7*, 1068–1078. [[CrossRef](#)]
12. Lal, C.; Balachander, K. Design of power system stabilizer for multi-machine systems using modified swarm optimization algorithm with wind energy generation. *J. Green Eng.* **2021**, *111*, 156–178.
13. Senthilkumar, G.; Chitra, M. An ensemble dynamic optimization based inverse adaptive heuristic critic in IaaS cloud computing for resource allocation. *J. Intell. Fuzzy Syst.* **2020**, *39*, 7521–7535. [[CrossRef](#)]
14. Gheisarnejad, M.; Khooban, M.H. An Intelligent Non-Integer PID Controller-Based Deep Reinforcement Learning: Implementation and Experimental Results. *IEEE Trans. Ind. Electron.* **2021**, *68*, 3609–3618. [[CrossRef](#)]
15. Xie, B.; Yuan, L.; Li, S.; Liu, C.H.; Cheng, X.; Wang, G. Active learning for domain adaptation: An energy-based approach. *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 8708–8716. [[CrossRef](#)]
16. Chen, C.; Cui, M.; Li, F.F.; Yin, S.; Wang, X. Model-Free Emergency Frequency Control Based on Reinforcement Learning. *IEEE Trans. Ind. Inform.* **2020**, *17*, 2336–2346. [[CrossRef](#)]
17. Min, C.; Pan, Y.; Dai, W.; Kawsar, I.; Li, Z.; Wang, G. Trajectory optimization of an electric vehicle with minimum energy consumption using inverse dynamics model and servo constraints. *Mech. Mach. Theory* **2023**, *181*, 105185. [[CrossRef](#)]
18. Chen, C.; Lu, J. Nonlinear Multivariable Control of a Dividing Wall Column Using a Different-Factor Full-Form Model-Free Adaptive Controller. *Ind. Eng. Chem. Res.* **2022**, *61*, 1897–1911. [[CrossRef](#)]
19. Venkatesan, K.; Govindarajan, U. Optimal power flow control of hybrid renewable energy system with energy storage: A WOANN strategy. *J. Renew. Sustain. Energy* **2019**, *11*, 015501. [[CrossRef](#)]
20. Li, J.; Xu, G.; Wang, Z.; Wang, Z. Modeling and optimization of urban rail transit scheduling with adaptive fruit fly optimization algorithm. *Open Phys.* **2019**, *17*, 888–896. [[CrossRef](#)]
21. Gao, A.; Geng, T.; Ng, S.X.; Liang, W. A Continuous Policy Learning Approach for Hybrid Offloading in Backscatter Communication. *IEEE Commun. Lett.* **2020**, *25*, 523–527. [[CrossRef](#)]

22. Barma, M.; Modibbo, U.M. Multiobjective mathematical optimization model for municipal solid waste management with economic analysis of reuse/recycling recovered waste materials. *J. Comput. Cogn. Eng.* **2022**, *1*, 122–137. [[CrossRef](#)]
23. Sun, X.; Qiu, J. A Customized Voltage Control Strategy for Electric Vehicles in Distribution Networks with Reinforcement Learning Method. *IEEE Trans. Ind. Inform.* **2021**, *17*, 6852–6863. [[CrossRef](#)]
24. Gong, S.; Zou, Y.; Xu, J.; Hoang, D.T.; Lyu, B.; Niyato, D. Optimization-Driven Hierarchical Learning Framework for Wireless Powered Backscatter-Aided Relay Communications. *IEEE Trans. Wirel. Commun.* **2021**, *21*, 1378–1391. [[CrossRef](#)]
25. Scarabaggio, P.; Carli, R.; Dotoli, M. Noncooperative Equilibrium-Seeking in Distributed Energy Systems Under AC Power Flow Nonlinear Constraints. *IEEE Trans. Control. Netw. Syst.* **2022**, *9*, 1731–1742. [[CrossRef](#)]
26. Choudhuri, S.; Adeniye, S.; Sen, A. Distribution alignment using complement entropy objective and adaptive consensus-based label refinement for partial domain adaptation. *Artif. Intell. Appl.* **2023**, *1*, 43–51. [[CrossRef](#)]
27. Kiritat, A.; Krejcar, O.; Ekici, B.; Tasgetiren, M.F. Multi-objective energy and daylight optimization of amorphous shading devices in buildings. *Sol. Energy* **2019**, *185*, 100–111. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.