

Optimal Dispatch of Integrated Energy Systems Based on Deep Reinforcement Learning

Daqing Kuang¹, Yingjun Ruan^{1*}, Hua Meng¹, Tingting Xu¹, Yuting Yao¹, Chaoliang Wang², Wei Liu²

¹ College of Mechanical Engineering, Tongji University, Shanghai, 201800, China

² State Grid Zhejiang Marketing Service Centre, Hangzhou, 310014, China

(Corresponding Author: 08156@tongji.edu.cn)

ABSTRACT

Integrated Energy Systems (IES) play a crucial role in promoting multi-energy complementarity and enhancing overall energy utilization efficiency. However, the intermittency of renewable energy sources and the stochastic nature of load demand pose significant challenges to system dispatch. To address these issues, this study proposes an economic dispatch method for renewable-integrated IES based on Deep Reinforcement Learning (DRL). The system components, including internal combustion engines, absorption chillers, and battery energy storage systems, are modeled, and the dispatch problem is formulated as a Markov Decision Process (MDP). Two DRL algorithms, Deep Q-Network (DQN) and Twin Delayed Deep Deterministic Policy Gradient (TD3), are employed to train optimal control strategies. Experimental results demonstrate that both algorithms achieve efficient system dispatch, with TD3 exhibiting superior convergence speed and overall performance. The proposed approach does not rely on explicit uncertainty modeling and can adaptively respond to fluctuations in renewable generation and load demand, thereby improving the economic efficiency and robustness of system operation.

Keywords: integrated energy system, DQN, TD3, economic dispatch

1. INTRODUCTION

Amid the accelerating global transition toward low-carbon and intelligent energy systems, Integrated Energy Systems (IES) have emerged as a key architecture for next-generation energy infrastructure, owing to their advantages in multi-energy coupling and coordinated control [1]. By integrating multiple energy carriers—such as electricity, heating, and cooling—with technologies including photovoltaics, internal combustion engines,

and energy storage, IES enable synergistic energy flow management, thereby improving efficiency and reducing carbon emissions. However, the intermittency of renewable energy and the stochastic nature of load demand render the dispatch problem high-dimensional and complex. Conventional optimization methods such as Mixed-Integer Linear Programming (MILP) [2], Dynamic Programming (DP) [3], and Model Predictive Control (MPC) [4] can yield near-optimal solutions under deterministic conditions but rely on precise modeling, involve high computational costs, and lack adaptability to dynamic environments. Therefore, data-driven, model-free, and self-learning approaches are urgently needed. Reinforcement Learning (RL), with its adaptive and real-time decision-making capability under uncertainty, offers a promising alternative. In recent years, Deep Reinforcement Learning (DRL) has been widely applied to energy system optimization [5,6]. For instance, Li et al. [7] proposed a DRL-based approach for renewable energy dispatch, while Cardo-Miota et al. [8] employed the TD3 algorithm to maximize the joint revenue of photovoltaic and energy storage systems.

Reinforcement learning has also been widely applied to the optimization of combined cooling, heating, and power (CCHP) systems and demand response. Ruan et al. [9] employed the DDPG and TD3 algorithms for the dispatch of a multi-energy system integrating CCHP units, photovoltaics, and energy storage, with TD3 demonstrating superior performance close to the theoretical optimum, thereby validating the effectiveness of DRL in complex system scenarios. Lu et al. [10] proposed a Q-learning – based dynamic pricing mechanism that enables online optimization of retail electricity prices, reducing user costs and enhancing system reliability. Furthermore, some studies have integrated deep learning – based load forecasting with reinforcement learning – based dispatch optimization,

This is a paper for the 11th Applied Energy Symposium: Low Carbon Cities & Urban Energy Systems (CUE2025), July 18-22, 2024, Kitakyushu, Japan.

further improving the intelligence and operational efficiency of energy systems [11].

In summary, reinforcement learning demonstrates strong applicability in energy system dispatch and significant potential for intelligent decision-making under uncertainty. However, existing studies mainly focus on the optimization of local subsystems, such as energy storage or combined cooling, heating, and power (CCHP) units, while research on system-level optimization under multi-energy coordination remains limited. Moreover, comparative analyses of different algorithms within integrated energy systems (IES) are still insufficient.

To address these gaps, this study investigates two representative reinforcement learning algorithms—the value-based Deep Q-Network (DQN) and the policy-based Twin Delayed Deep Deterministic Policy Gradient (TD3)—in a typical IES scenario. A dispatch model is developed, and control strategies are learned through interactions with the environment. The two algorithms are compared in terms of dispatch performance and convergence characteristics, aiming to provide methodological insights and practical guidance for the coordinated optimization of integrated energy systems using reinforcement learning.

2. EQUIPMENT MODELING

Figure 1 shows the schematic diagram of the integrated energy system. Electricity is supplied by the internal combustion engine, photovoltaics, and the battery, with the grid compensating for any shortfall, while the cooling demand is met by the absorption and electric chillers.

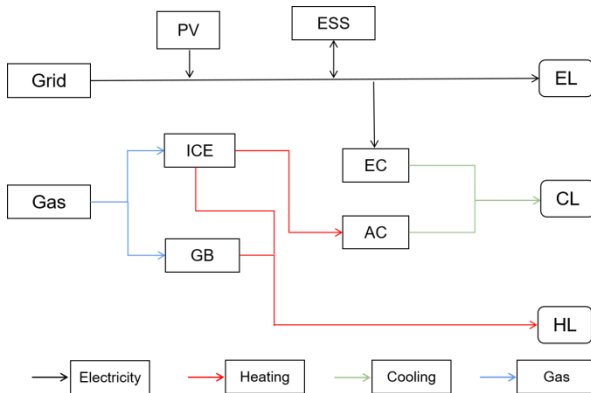


Figure 1. Integrated energy system structure diagram

2.1 Internal Combustion engine model

The internal combustion engine is widely used in combined heat and power (CHP) systems, and its mathematical model is as follows:

$$\eta_h = \beta \eta_e \quad (1)$$

$$P_I = M_{gas} \times q_{gas} \times \eta_e \quad (2)$$

$$H_I = M_{gas} \times q_{gas} \times \eta_h \quad (3)$$

$$P_{Lp} = \frac{P_I}{P_e} \quad (4)$$

$$P_I \leq P_{I_max} \quad (5)$$

η_e , η_h and β represent the electric efficiency, thermal efficiency, and thermoelectric ratio of the internal combustion engine, respectively; P_I and H_I respectively represent the actual generating power and thermal power of the internal combustion engine, kW ; M_{gas} represents the amount of natural gas, m^3 ; q_{gas} represents the low calorific value of natural gas, kWh/m^3 ; P_{Lp} represents the electric load rate of an internal combustion engine; P_e and P_{I_max} respectively represent the rated power and maximum generating capacity of the internal combustion engine.

2.2 The model of the electric chiller

$$L_{EC} = P_{EC} \times COP_{EC} \quad (6)$$

$$L_{EC} \leq L_{EC_max} \quad (7)$$

L_{EC} , P_{EC} , and L_{EC_max} represent the refrigerating capacity, power consumption, and maximum refrigerating capacity of the electric refrigerator, respectively, kW ; COP_{EC} represents the refrigeration coefficient of the electric refrigerator.

2.3 The model of absorption chiller

$$L_{ABS} = H_{ABS} \times COP_{ABS} \quad (8)$$

$$P_L = \frac{L_{ABS}}{L_e} \quad (9)$$

$$L_{ABS} \leq L_{ABS_max} \quad (10)$$

L_{ABS} , H_{ABS} , and L_e represent the cooling capacity, heat consumption, and rated power of absorption refrigeration unit, respectively, kW ; COP_{ABS} and P_L represent the refrigeration coefficient and cold load rate of the absorption chiller unit.

2.4 Electric Energy Storage and System Constraints

The electric energy storage unit maintains dynamic energy balance through the charging and discharging processes, with its state constrained by the maximum capacity and charging/discharging efficiency. The system operation must satisfy the balance of electricity, heat, and cooling, ensuring that energy supply and demand are matched at each time step. Renewable generation, grid power, storage devices, and multi-energy demands are

jointly coupled to form the overall energy interaction framework, which provides the basis for subsequent dispatch optimization.

3. DRL METHODS

3.1 Fundamentals of DRL

3.1.1 Overview of the RL Framework

Reinforcement learning (RL) is typically modeled as a Markov Decision Process (MDP), MDP is generally defined by a four-tuple, where: (S, A, R, π) . These four elements represent a set of states, actions, a reward function, and a policy that maps states to actions to maximize cumulative rewards.

The objective of the agent is to learn a policy $\pi(a|s)$ so as to maximize the expected cumulative return, that is:

$$\max_{\pi} E_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) \right] \quad (11)$$

γ is the Discount factor.

3.1.2 Principle of the DQN Algorithm

Its core idea is to use a parameterized Q-network $Q(s, a; \theta)$ to approximate the action-value function, which is defined as:

$$\max_{\pi} E_{\pi} \left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t) | s_0 = s, a_0 = a \right] \quad (12)$$

The parameters of the Q-network are updated by minimizing the following loss function:

$$\mathcal{L}(\theta) = E_{(s,a,r,s') \sim D} [(y - Q(s, a; \theta))^2] \quad (13)$$

where the target Q-value is calculated as:

$$y = r + \gamma \max_{a'} Q(s', a'; \theta^-) \quad (14)$$

In Equations (20)–(22), s_0, s , and s' represent the initial, current, and next states, respectively, while a_0, a , and a' denote the initial, current, and next actions. r is the immediate reward at the current time step, and θ and θ^- refer to the parameters of the Q-network and the target network, respectively.

DQN is suitable for discrete action space problems and has been widely applied in energy systems for tasks such as load allocation and energy storage control.

3.1.3 Principle of the TD3 Algorithm

TD3 (Twin Delayed DDPG) is a reinforcement learning algorithm designed for continuous action

spaces, based on the policy gradient and Actor-Critic framework. Its primary goal is to directly learn a deterministic policy $\mu(s; \theta^{\mu})$ that maximizes the Q-value:

$$\max_{\theta^{\mu}} E_{s \sim D} [Q(s, \mu(s; \theta^{\mu}))] \quad (15)$$

The actor network is optimized using the following gradient:

$$\nabla_{\theta^{\mu}} J \approx E_{s \sim D} \left[\nabla_a Q_1(s, a; \theta^Q) /_{a=\mu(s)} \nabla_{\theta^{\mu}} \mu(s) \right] \quad (16)$$

TD3 is particularly well suited for integrated energy systems (IES), where it can handle continuous control variables such as internal combustion engine output and battery charging/discharging power.

3.2 MDP Modeling for IES Dispatch

3.2.1 Design of the State Space

In this study, the environment corresponds to an integrated energy system. The observable information from the environment includes the renewable energy output at each time step, cooling load, electricity load, state of charge (SOC) of the battery, and electricity price. Thus, the state space is defined as follows:

$$s = [P_{RG}(t), E_c(t), E_{el}(t), SOC_t, Pr(t)] \quad (17)$$

Where $P_{RG}(t)$ denotes the output of renewable energy at time t , kW ; $Pr(t)$ denotes the electricity purchase price from the grid at time t .

3.2.2 Design of the Action Space

In integrated energy systems (IES), the agent's actions typically correspond to the power outputs of various devices. Due to the presence of energy balance equations, the action space of the system can ultimately be reduced to two dimensions: the electrical output of the internal combustion engine and the charging/discharging power of the energy storage system.

$$a = [P_I(t), P_{BA}(t)] \quad (18)$$

$P_{BA}(t)$ is the charging and discharging power of the battery at the time, kW .

3.2.3 Design of the Reward Function

The operating cost of the system primarily consists of the expenses for electricity purchase and natural gas consumption. The reward function is formulated with the total operating cost as its principal component, while incorporating penalty terms for power and cooling energy imbalances, as well as for overcharging and

overdischarging behaviors of the energy storage unit. In this way, the reinforcement learning agent achieves a trade-off between minimizing operating costs and maintaining system stability throughout the learning process.

3.3 Algorithm Implementation and Network Architecture

Since the DQN algorithm requires a discrete action space, the output power of the internal combustion engine and the charging/discharging power of the battery were discretized based on their rated capacities. To improve resolution in high-power and frequent charge/discharge regions, a non-uniform discretization strategy was adopted, resulting in 108 possible action combinations. In contrast, the TD3 algorithm inherently supports continuous control and therefore does not require discretization; however, its action range was set consistent with that of the DQN to ensure a fair comparison.

Table 1 summarizes the key hyperparameters and neural network architectures of the DQN and TD3 algorithms.

Table 1. Hyper Parameters used for DQN and TD3

Parameters	Value/DNQ	Value/TD3
Discount Factor	0.99	0.99
Replay buffer capacity	10000	50000
Batch size	64	128
Actor learning rate	-	(1e-4, 1e-5)
Critic learning rate	-	(1e-4, 1e-5)
Q-network learning rate	(1e-4, 1e-5)	-
total_timesteps	168*1000	168*1000

4. COMPARATIVE STUDY AND SIMULATION RESULTS

4.1 Experimental Setup and Data Sources

This study uses a simulated office building in Shanghai as the user-side scenario. Meteorological and pricing inputs are based on 2019 solar data and local time-of-use electricity and gas rates. Key equipment parameters of the IES are listed in Table 2

Table 2. Device Parameters

Device	Parameters	Value
Natural gas	$q_{gas}/kWh \cdot m^{-3}$	10

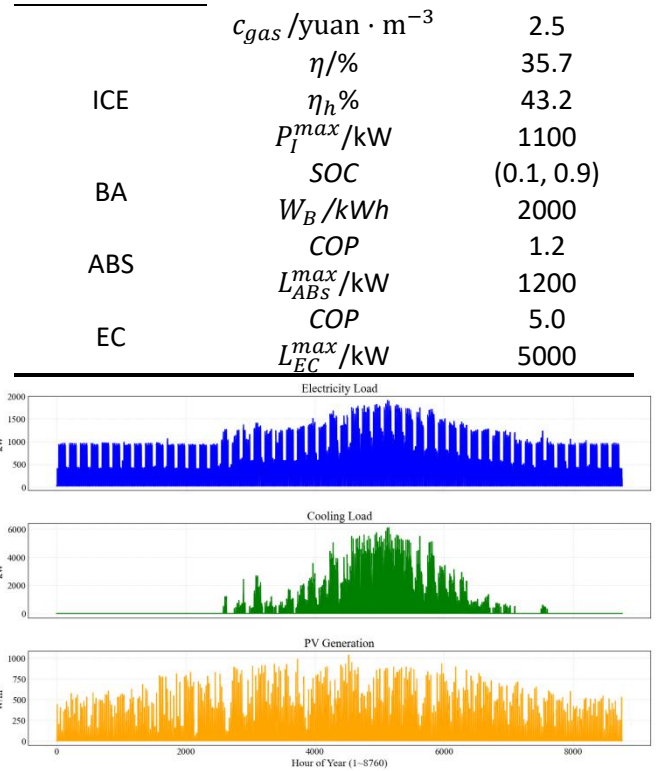


Figure 2. Temporal Distribution of Electricity Load, Cooling Load, and PV Output

The annual profiles of electricity load, cooling load, and photovoltaic (PV) irradiance over 8760 hours are shown in Figure 2. Under summer operating conditions, data from June 2 to July 31 (a total of 60 days) are selected as training data, while data from August 1 to August 21 (21 days) are used for testing. The time-of-use electricity pricing adopted in this study corresponds to the Shanghai region, as illustrated in Figure 3.

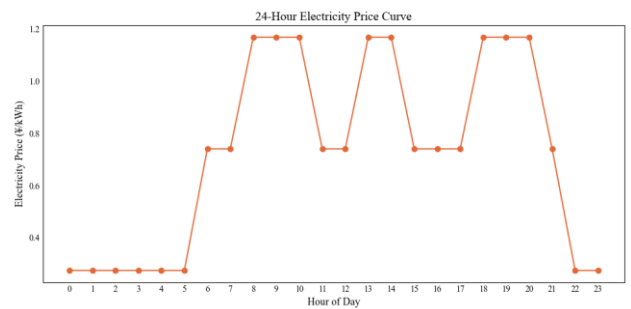


Figure 3. Time-of-Use Electricity Price Chart

4.2 Convergence Performance Comparison

Figure 4 illustrates the training reward curves of DQN and TD3 in the economic dispatch task of the integrated energy system. Compared to DQN, TD3 converges faster, achieves higher final rewards, and exhibits smaller fluctuations, demonstrating stronger policy learning capability and training stability. It is more suitable for

energy system dispatch optimization with continuous action spaces.

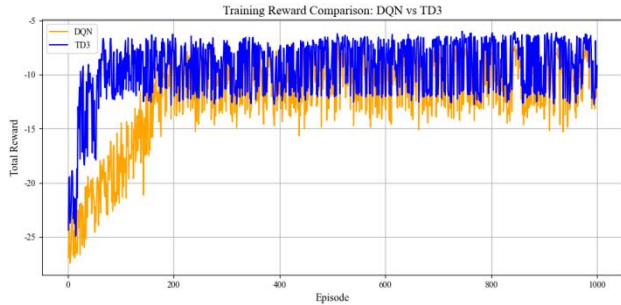


Figure 4. Training Convergence Comparison between DQN and TD3

4.3 Comparative Analysis of DQN and TD3

Table 3 presents a comparison of the system operating costs under the DQN and TD3 dispatch strategies across three representative test days. The results indicate that TD3 consistently achieves lower operating costs than DQN on all test days, demonstrating its superiority in optimizing the economic performance of the integrated energy system.

Table 3. Operating Cost Comparison between DQN and TD3

Test day	Operating Cost	
	DQN	TD3
1	20434.14493	20209.96977
2	17233.7602	17161.8293
3	16377.3224	15831.22826

Specifically, on Day 1, TD3 reduces the operating cost by approximately ¥224.17 compared to DQN. On Days 2 and 3, the cost differences are ¥71.93 and ¥546.09, respectively. Notably, the advantage of TD3 is most significant on Day 3, with a cost reduction of approximately 3.33%. This indicates that TD3 maintains greater robustness and cost-efficiency even under conditions with high load fluctuations or significant variations in photovoltaic output.

Figures 5 and 6 present the scheduling results of electrical and cooling subsystems on a typical test day. In terms of power load dispatch, the internal combustion engine (ICE) output under the TD3 strategy is significantly higher than that of DQN, particularly at 10:00, 13:00, and 17:00, reflecting its advantage in cost reduction. TD3 also demonstrates superior performance in controlling the charge and discharge behavior of the energy storage system. Benefiting from its continuous action space, TD3 enables fine-grained energy regulation, thereby avoiding

large power fluctuations. As observed from the dispatch results, TD3 achieves stable and uniform charging during most daytime periods without exhibiting any apparent irrational charge/discharge behaviors. In contrast, DQN is constrained by its discrete action space, leading to discontinuous and jumpy control decisions. This includes charging behavior during economically unfavorable periods—such as 11:00 when electricity prices are high—which compromises the economic efficiency and rationality of system operation. The two dispatch diagrams also reveal significant differences in the output ranges of various devices within the system. When the output capacities of system components vary widely, reinforcement learning algorithms with continuous action spaces offer clear advantages over those based on discrete actions. They enable more precise power regulation and coordinated control, thereby enhancing the adaptability and overall performance of the dispatch strategy.

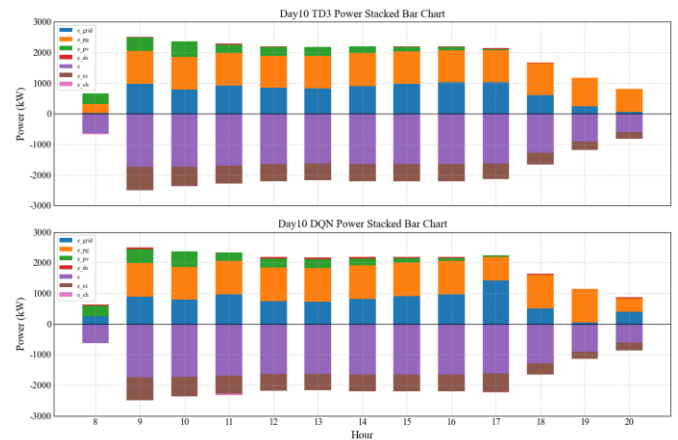


Figure 5. Comparison of Power Dispatch between TD3 and DQN

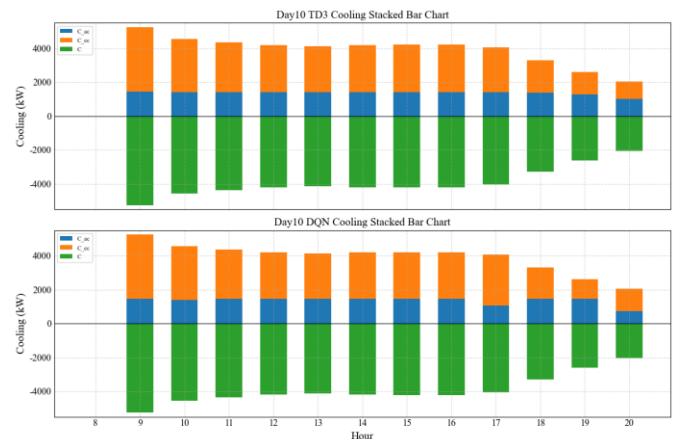


Figure 6. Comparison of Cooling Dispatch between TD3 and DQN

In terms of cooling load dispatch, the absorption chiller output under the TD3 strategy is noticeably higher

than that of DQN, particularly at 10:00 and 17:00. Compared to DQN, TD3 more effectively utilizes the advantages of absorption-based cooling in the combined cooling and heating system, thereby reducing the electric cooling load and improving overall system energy efficiency. Its strategy enables effective coordination between thermal and cooling resources, highlighting the benefits of multi-energy complementary dispatch optimization.

Through comparative analysis, it is evident that TD3 outperforms DQN in multi-energy coordinated dispatch. By increasing the utilization of the internal combustion engine and absorption chiller, and effectively regulating the charge/discharge behavior of the energy storage system, TD3 is able to meet both electrical and cooling load demands while reducing dependence on the power grid. This leads to enhanced economic performance and better coordination within the integrated energy system.

5. CONCLUSIONS

Based on the modeling and optimal dispatch of an integrated energy system (IES) incorporating renewable energy sources, this study conducts a comparative analysis of two deep reinforcement learning algorithms: DQN and TD3. The main conclusions are summarized as follows:

(1) The dispatch problem of the integrated energy system is formulated as a Markov Decision Process (MDP), with a state space, action space, and reward function specifically designed to reflect the system's operational objectives and constraints.

(2) Experimental results demonstrate that both DQN and TD3 are capable of learning feasible dispatch strategies through interaction with the environment. However, TD3 exhibits superior performance in terms of convergence speed, training stability, and long-term dispatch effectiveness, achieving higher average rewards than DQN.

(3) The TD3 algorithm enables a more balanced energy allocation strategy and more stable battery discharge behavior, thereby reducing dependence on grid electricity and enhancing the overall economic efficiency of system operation.

ACKNOWLEDGEMENT

This research was supported by National Key R&D Program of China (No.2023YFC3807100).

REFERENCE

- [1] Wang, Y., Zhou, H., Ma, K., & others. (2025). Dynamic pricing and low-carbon economy dispatch of multi-park integrated energy system based on mixed game. *Electric Power Systems Research*, 248, 111908.
- [2] Solanki, B. V., Bhattacharya, K., & Cañizares, C. A. (2017). A sustainable energy management system for isolated microgrids. *IEEE Transactions on Sustainable Energy*, 8(4), 1507–1517.
- [3] Zhang, Y., Ai, Q., & Hao, R. (2019). Economic dispatch of integrated energy system at building level based on chance constrained programming. *Power System Technology*, 43(1), 108–116.
- [4] Arroyo, J., Spiessens, F., & Helsen, L. (2020). Identification of multi-zone grey-box building models for use in model predictive control. *Journal of Building Performance Simulation*, 13(4), 472–486.
- [5] Lee, S., Prabawa, P., & Choi, D. H. (2025). Joint peak power and carbon emission shaving in active distribution systems using carbon emission flow-based deep reinforcement learning. *Applied Energy*, 379, 103274.
- [6] Liu, Z. Y., Zhang, Y. X., & Ning, Y. (2025). Emergency mobile energy storage optimal allocation in microgrid-integrated distribution networks considering economic and resilience benefits. *Energy*, 322, 135633
- [7] Li, Y., Ma, W., Li, Y., & others. (2025). Enhancing cyber-resilience in integrated energy system scheduling with demand response using deep reinforcement learning. *Applied Energy*, 379, 103274.
- [8] Cardo-Miota, J., Beltran, H., Pérez, E., & others. (2025). Deep reinforcement learning-based strategy for maximizing returns from renewable energy and energy storage systems in multi-electricity markets. *Applied Energy*, 388, 125561.
- [9] Ruan, Y., Liang, Z., Qian, F., & others. (2023). Operation strategy optimization of combined cooling, heating, and power systems with energy storage and renewable energy based on deep reinforcement learning. *Journal of Building Engineering*, 65, 105682.
- [10] Lu, R. Z., Hong, S. H., & Zhang, X. F. (2018). A dynamic pricing demand response algorithm for smart grid: Reinforcement learning approach. *Applied Energy*, 220, 220–230.
- [11] Ruan, Y. J., Hou, Z. Q., Qian, F. Y., & others. (2022). Optimization of the operation of distributed energy system based on deep reinforcement learning. *Science Technology and Engineering*, 22(17), 7021–7030.