Contents list available at CBIORE journal website

International Journal of Renewable Energy Development

Journal homepage: https://ijred.cbiore.id



Research Article

Source-grid-load-storage multi-terminal collaborative control technology for photovoltaic distribution network based on DQN

Weiru Niu, Suyuan Jiao, Zhihao Wang, Lingzhi Li, Tie Cao*

State Grid Gansu Electric Power Company Jiayuguan Power Supply Company, Jiayuguan 735100, China

Abstract. As the photovoltaic penetration rate continues to increase, the dynamic characteristics of the multi-dimensional coupling of source, grid, load and storage in the distribution network have become increasingly significant. Traditional control methods often face response lag and regulation instability in this multi-time-varying scenario. Therefore, the study proposes a source-grid-load-storage multi-terminal collaborative control strategy for photovoltaic distribution network. A centralized and distributed two-level collaborative architecture is established to enable deep reinforcement learning agents to learn the optimal power regulation strategy. The multi-time scale control of voltage regulator taps, photovoltaic, energy storage systems, and static reactive power compensation devices is coordinated to achieve voltage optimization and efficient energy allocation. A comparative test was conducted using the IEEE 33-node system as the simulation platform. The results showed that the proposed method could increase the peak photovoltaic output by about 20%-25% on the power generation side, and the average power generation utilization rate reached 93%. The load transfer rate increased to 26.2%, and the load balance dropped to 0.21. In terms of voltage stability, the system fluctuation was controlled within 0.15 p.u., and the steady-state error was less than 0.01 p.u. In typical operating scenarios, deep reinforcement learning control increased the energy storage utilization rate to 86.3%-88.9% and reduced the network loss rate to 3.1%-3.5%. The results show that this method is significantly better than the comparison strategy in terms of multi-source collaboration and dynamic stability, and provides an efficient and intelligent optimization solution for the safe and economical operation of high-penetration photovoltaic distribution networks.

Keywords: Photovoltaic distribution network; Deep reinforcement learning; Source-grid-load-storage collaborative control; Voltage control; Energy management

Check for updates

@ The author(s). Published by CBIORE. This is an open access article under the CC BY-SA license (https://creativecommons.org/licenses/by-sa/4.0/).

Received: 16th Dec 2025; Revised: 17th August 2026; Accepted: 19th Oct 2026; Available online: 1st Nov 2026

1. Introduction

As the penetration of renewable energy continues to increase, distribution networks are gradually evolving from traditional unidirectional power delivery systems into active networks characterized by strong coupling among source, grid, load, and storage. High-proportion photovoltaic integration increases the uncertainty of power flow and may lead to voltage deviation, power fluctuation, and insufficient utilization of flexible regulation resources. Therefore, coordinated voltage control based on distributed energy resources and hierarchical regulation has become an important research direction. Stanojev *et al.* (2025) proposed an online feedback droop scheduling method for distribution grids, which coordinated distributed energy resources to achieve frequency and local voltage regulation and improved the adaptability of voltage support under changing operating conditions. Irmak *et al.* (2024) developed an event-triggered distributed secondary control strategy for microgrids, reducing unnecessary communication while maintaining efficient and reliable coordinated regulation. Bosisio *et al.* (2025) introduced Q-learning into on-load tap changer operation and voltage control under high renewable energy penetration, enabling adaptive tap regulation according to system operating states. Manojkumar *et al.* (2024) coordinated photovoltaic generation and battery energy storage to compensate for net-load forecasting errors, thereby improving peak-shaving capability and reducing the influence of source-load uncertainty on system operation.

Hong *et al.* (2024) developed a multi-agent deep reinforcement learning-based bi-level Volt/Var optimization framework to coordinate grid-side static var compensators and customer-side photovoltaic inverters, thereby mitigating voltage deviations and reducing distribution network losses under high photovoltaic penetration. Wang *et al.* (2024) further coordinated dynamic network reconfiguration and Volt/Var control through a graph-aware deep reinforcement learning framework, demonstrating the potential of data-driven methods for handling high-dimensional discrete control decisions in active distribution networks. Mu *et al.* (2024) introduced graph multi-agent reinforcement learning into inverter-based active voltage control, using interactions among distributed agents to improve voltage regulation and network loss performance. In addition, Yu *et al.* (2024) proposed a risk-aware reinforcement learning framework for power fluctuation smoothing under high renewable energy penetration, highlighting the importance of considering uncertainty and operational risks in adaptive control. Peng *et al.* (2025) developed a deep reinforcement learning-based multi-timescale Volt/Var control method that coordinated devices with different response characteristics while considering distribution network reconfiguration, further

* Corresponding author
Email caotiee@outlook.com (T. Chao)

https://doi.org/10.61435/ijred.2026.62121
ISSN: 2252-4940/© 2026.The Author(s). Published by CBIORE

demonstrating the necessity of multi-timescale coordination in renewable-rich distribution systems.

In recent years, reinforcement learning and deep reinforcement learning have been increasingly applied to active distribution network control because of their capability to learn adaptive decision policies under nonlinear and time-varying operating conditions. D Wang *et al.* (2025) proposed a multi-agent deep reinforcement learning method for coordinating large-scale photovoltaic inverters and demonstrated its effectiveness in local distribution voltage regulation. Mukherjee *et al.* (2025) developed a vertical federated reinforcement learning framework for resilient control of networked microgrids and verified the coordinated control performance through a real-time test bed. Dang *et al.* (2025) proposed a meta-reinforcement-learning-based adaptive energy storage control method, which improved the adaptability and interpretability of storage scheduling under dynamic operating scenarios. Shalaby *et al.* (2025) further introduced safe deep reinforcement learning into resilient distribution grid control, incorporating operational security into the decision-making process and improving system control capability under severe disturbances. Guan *et al.* (2025) proposed a model-based safe reinforcement learning framework for active distribution network scheduling, embedding operational security constraints into the learning process to improve renewable energy utilization while ensuring feasible control actions. Locher *et al.* (2026) combined reinforcement learning with model predictive control for utility-level battery management in distribution grids, achieving fast real-time decision-making while explicitly considering battery state-of-charge and operational constraints. From the perspective of distributed decision-making, Vu *et al.* (2024) employed multi-agent deep reinforcement learning for distributed load restoration and demonstrated that centralized training and decentralized execution can effectively coordinate multiple agents while reducing constraint violations. These studies indicate that reinforcement learning-based control is gradually extending from single-device regulation toward multi-resource coordination, adaptive energy management, multi-timescale decision-making, and security-aware operation.

In summary, although existing studies have made certain progress in distribution network voltage regulation, reactive power optimization, and distributed energy scheduling, several limitations remain. First, most studies mainly focus on a single regulation object, such as the reactive power output of distributed energy resources, energy storage charging and discharging, or transformer tap adjustment, while insufficient attention has been paid to the coordination among OLTC, photovoltaic units, energy storage systems, SVC devices, and load-side response. Second, traditional model-driven optimization methods usually rely on accurate system parameters and continuous state estimation, which may lead to control delay and regulation instability under photovoltaic output fluctuations, load uncertainty, and communication delays. Third, although existing reinforcement learning methods can effectively address local voltage regulation and continuous power control problems, most studies still focus on single control objectives or single-agent optimization. Comprehensive coordination among multiple devices with different time scales and improvement of energy storage utilization efficiency under high photovoltaic penetration scenarios have not been sufficiently considered.

To address the above research gaps, this paper proposes a DQN-based source-grid-load-storage multi-terminal collaborative control method for photovoltaic distribution networks. The novelty of this study is mainly reflected in three aspects. First, a centralized-distributed two-level collaborative control architecture is constructed to coordinate global transformer tap regulation with local power regulation of photovoltaic units, energy storage systems, and SVC devices. Second, a multi-time-scale reinforcement learning control strategy is designed for high-penetration photovoltaic scenarios to improve voltage stability, photovoltaic accommodation capability, and energy storage utilization. Third, voltage deviation, network loss, equipment action smoothness, and energy scheduling efficiency are jointly considered in the reward function, enabling the control strategy to balance operational security and economic performance. Therefore, this study provides an intelligent optimization solution for the adaptive, robust, and economical operation of high-penetration photovoltaic distribution network.

2.Methodology

2.1 Construction of integrated voltage regulation model of source-grid-load-storage

With the expansion of distribution network scale and increasing photovoltaic penetration, voltage regulation has gradually evolved from local compensation toward coordinated control involving multiple distributed resources (Nematshahi *et al.*, 2023; Chaturvedi *et al.*, 2024). Therefore, the study proposes a global and local collaborative voltage control method, which achieves global voltage optimization by adjusting the On-load Tap Changer (OLTC) tap adjustment, and combines photovoltaics, energy storage and SVC to achieve precise local regulation. OLTC adjusts the secondary side voltage by adjusting the tap position to maintain the node voltage within the allowable range. In actual operation, OLTC introduces delay and dead zone mechanisms to avoid equipment losses caused by frequent actions. The voltage deviation and its discrete time

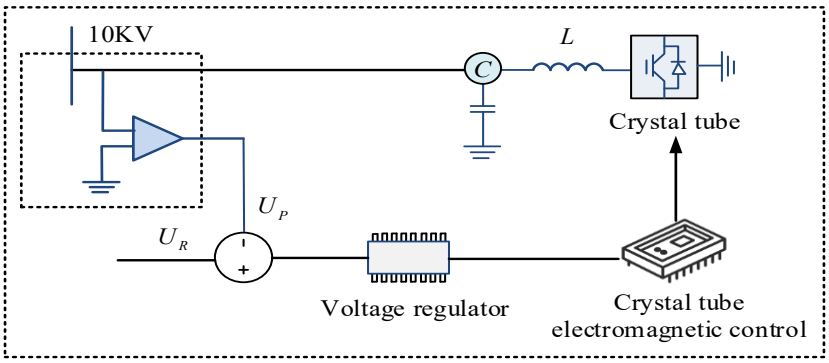


Fig.1 SVC grid connected structure

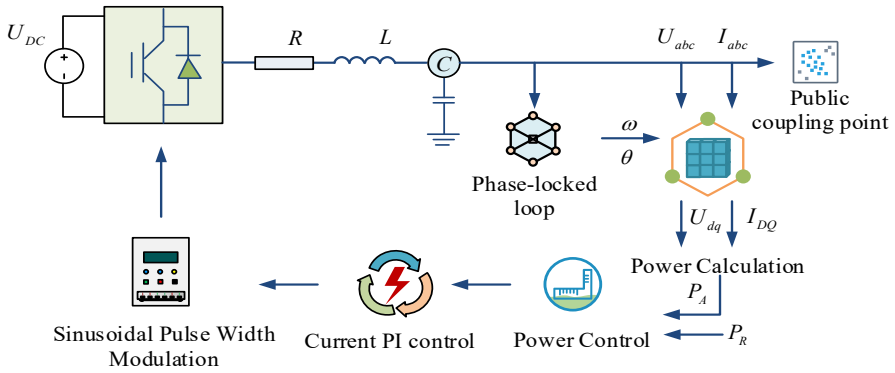


Fig.2 PQ control model

accumulation are calculated, as shown in equation (1) (Bosisio *et al.*, 2025).

$$\begin{cases} e = V - V_{ref} \\ t(\tau + 1) = t(\tau) + \Delta S_\tau \end{cases} \quad (1)$$

In equation (1), e represents the voltage error. V is the actual voltage measured twice by the transformer. V_{ref} is the nominal voltage expected to be maintained. t represents the operating time. τ is the counter. ΔS_τ is the step size. The voltage regulating action judgment function is shown in equation (2).

$$f(e,t) = \begin{cases} 1 & e > \phi/2, t \geq T_d \\ 0 & otherwise \\ -1 & e < -\phi/2, t \geq T_d \end{cases} \quad (2)$$

In equation (2), ϕ represents the voltage dead zone. T_d is the allowed duration. The tap gear update process is shown in equation (3).

$$N(t+1) = N(t) - L \cdot f(e(t), t) \quad (3)$$

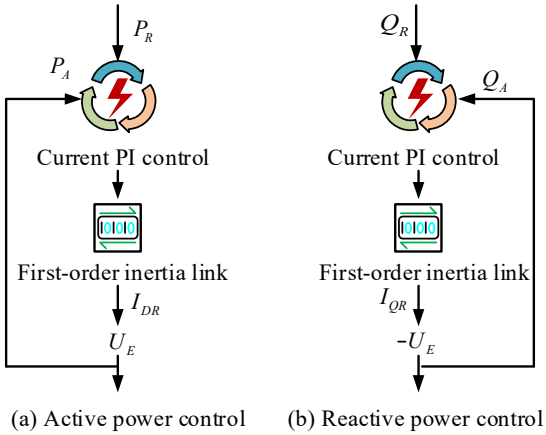
In equation (3), $N(t)$ represents the current tap gear. L is the gear step size for each adjustment, usually 1. SVC based on thyristor phase-controlled reactors is used to achieve rapid dynamic adjustment of the node voltage of the distribution network. Its grid-connected structure is shown in Figure 1.

In Figure 1, SVC adopts a parallel structure of capacitor and thyristor-controlled reactor to form an adjustable reactive power compensation branch. The system monitors the grid-connection point voltage U_P in real time. After comparing it with the reference voltage U_R, the voltage regulator generates control instructions to drive the thyristor to adjust the switching degree of the reactance branch, dynamically changing the equivalent susceptance, and rapidly adjusting the reactive power output. Capacitor C provides capacitive reactive power to increase the voltage, and thyristor-controlled reactor L absorbs inductive reactive power to suppress overvoltage, jointly maintaining the voltage at the grid-connection point stable within the set range. The equivalent transfer function of SVC is shown in equation (4).

$$\begin{cases} \Delta E_{SVC} = \frac{1}{sT+1} \cdot \Delta c_{SVC} \\ \Delta Q_{SVC} = \Delta E_{SVC} \cdot U_{SVC}^2 \end{cases} \quad (4)$$

In equation (4), ΔE_{SVC} represents the change in SVC's current equivalent susceptance relative to the previous moment. Δc_{SVC} represents the change amount of the control command. s is the Laplacian operator. T is the control time constant. ΔQ_{SVC} represents the change in SVC's current output reactive power relative to the previous moment. U_{SVC} is the voltage at the grid-connection point. Both energy storage systems and photovoltaic units are connected to the distribution network through power electronic converters, enabling coordinated active and reactive power regulation (Chaturvedi *et al.*, 2024). The PQ control model adopted in this study is shown in Figure 2.

In Figure 2, the inverter converts the DC side voltage U_{DC} of the energy storage device to the AC side after sinusoidal pulse width modulation and uses a series filter branch to form anIC filter composed of resistor R, inductor I and C to suppress high-order harmonics and smooth current output. The inverter is connected in parallel to the public coupling point of the



(a) Active power control (b) Reactive power control

Fig.3 Current PI controller module

distribution network, and the grid-connected voltage U_{abc} and grid-connected current I_{abc} are measured in real time. The grid voltage phase θ and angular frequency ω are extracted through the phase-locked loop, and U_{abc} and I_{abc} are transformed into a synchronous rotating coordinate system to obtain the voltage component U_{dq} and current component I_{DQ}. Then, the instantaneous active and reactive power are calculated. The power controller generates a current command based on a given power reference value P_R, outputs a control signal through a current Proportional-Integral (PI) regulator, and then generates a sinusoidal pulse width modulation trigger signal through coordinate inverse transformation, ultimately achieving precise decoupling control of grid-connected power. The current PI controller module is shown in Figure 3.

Figure 3(a) and Figure 3(b) show the active power and reactive power control modules of the energy storage system respectively. In Figure 3(a), the active power control module takes the deviation between P_R and the actual output power P_A as input. After PI adjustment and dynamic smoothing of the first-order inertia link, the reference current I_{DR} is out, which is used to control the active component of the inverter and accurately track the active power. In Figure 3(b), the reactive power control module takes the deviation between the reactive power Q_R and the actual output power Q_A as input. Based on PI adjustment and inertial link processing, a reactive current command I_{QR} is generated, which is used to adjust the inverter's reactive output in real time. The current commands output by the two modules work together on the inverter voltage control variable U_E to complete active-reactive power decoupling control and ensure stable operation of the system under power fluctuation conditions.

2.2 Multi-time scale voltage collaborative control strategy for distribution network based on DQN

After completing the modeling and control mechanisms of multiple regulation units in the distribution network, a multi-time-scale coordinated control strategy based on DQN is developed to further achieve dynamic coordination among power sources, the grid, loads, and energy storage terminals, as well as global voltage optimization. A two-level coordinated control architecture combining centralized and distributed control is adopted. The upper-level centralized controller generates coordinated control decisions according to the global operating state, and is responsible for voltage regulator tap adjustment and multi-device coordination (Amer *et al.*, 2023; Mohamed and Tsuji, 2023). The lower-level distributed controllers perform power regulation of photovoltaic systems, energy storage systems, and SVC devices based on regional operating states. Through information exchange between the upper- and lower-level controllers, global voltage optimization and local fast response are achieved simultaneously. Unlike reinforcement learning algorithms designed for continuous action spaces, DQN is selected in this study mainly to address the discrete coordination decision problem among multiple regulation devices. In practical distribution network operation, voltage regulator tap switching, device operation mode selection, and power regulation level determination exhibit clear discrete characteristics, whereas the continuous power outputs of photovoltaic inverters, energy storage converters, and SVC devices are regulated by the underlying PQ control and PI control loops. Therefore, applying DQN for upper-level strategy optimization avoids the expansion of the action space caused by directly discretizing multiple continuous control variables. Meanwhile, its stability in discrete action decision-making and strong policy search capability enable effective multi-source coordinated control under complex and time-varying operating conditions.

To account for the different response characteristics of various regulation devices, a coordinated execution mechanism with fast and slow time scales is adopted. The lower-level PV inverters, energy storage converters, and SVC controllers operate with a 1 s control interval to perform rapid power regulation according to local voltage and power states, while the upper-level centralized controller operates with a 15 min coordination interval to update the OLTC tap position and device coordination strategy based on the aggregated regional states. The lower-level controllers update local states every 1 s and upload the regional state information to the upper-level controller at the end of each 15 min coordination interval. After generating new coordination commands, the upper-level controller synchronously distributes them to all regions at the beginning of the next coordination interval. The OLTC performs tap adjustment only at the upper-level coordination instants and when the voltage deviation satisfies the action condition, thereby avoiding frequent mechanical operations and achieving

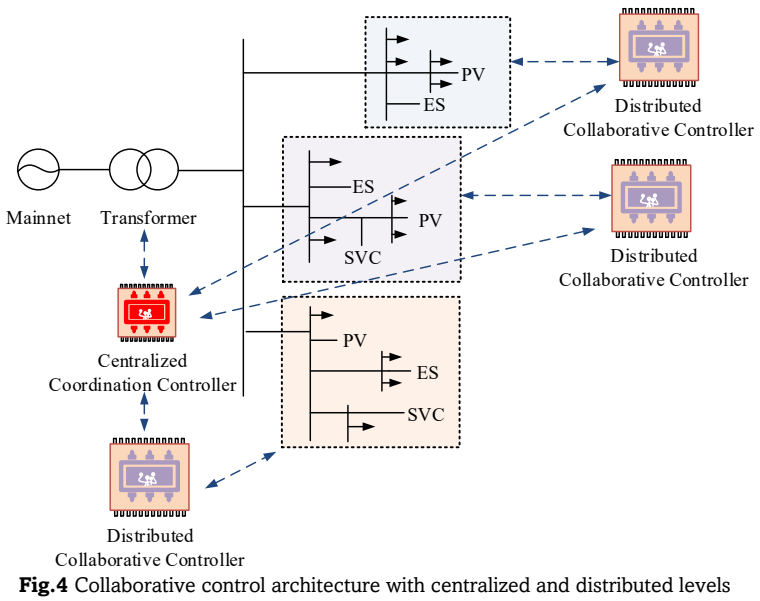


Fig.4 Collaborative control architecture with centralized and distributed levels

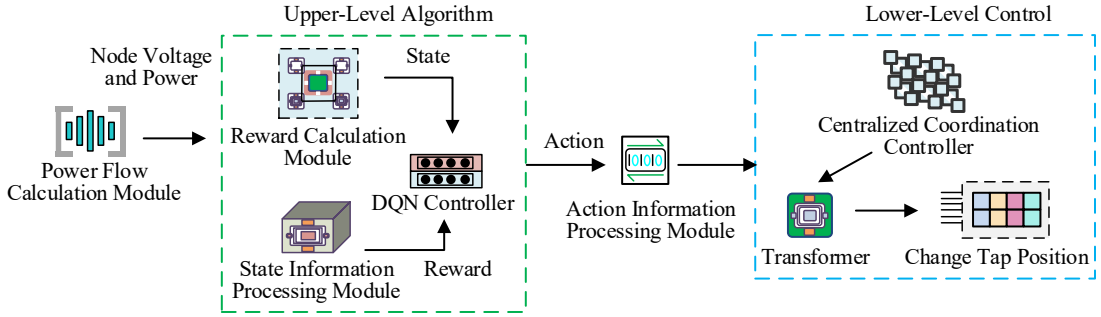


Fig.5 Multi-time scale collaborative control strategy based on DQN

effective time-scale coordination between slow global regulation and fast local control.

The overall architecture is shown in Figure 4. In Figure 4, the upper-layer centralized coordinator collects global information and issues voltage and power instructions uniformly. The distributed controllers in each region directly manage distributed Photovoltaic (PV), Energy Storage system (ES) and SVC equipment, and perform rapid adjustments based on real-time measurements. The two exchange data through communication links to achieve global optimization while ensuring rapid local response, and finally complete the collaborative control of multi-regional voltage and reactive power. The distributed intelligent agent collects the node voltage and regulates equipment power in the area, and calculates the node voltage per unit value accordingly, as shown in equation (5).

$$V_{p,i} = V_i(t) \cdot V_{r,i}^{-1} \quad (5)$$

In equation (5), $V_{p,i}$ is the voltage per unit. $V_{r,i}^{-1}$ is the nominal voltage. $V_i(t)$ represents the real-time voltage value of node i at time t. To alleviate the bus voltage over-limit problem, a multi-time scale collaborative control strategy based on DQN is proposed. When a voltage abnormality is detected, the centralized coordination controller generates optimal power adjustment instructions based on the training results and sends them to the regional distributed controllers, which coordinate the photovoltaic, energy storage and SVC outputs to activate the voltage regulator taps in parallel to achieve dynamic stable control of the voltage. The multi-time scale collaborative control strategy based on DQN is shown in Figure 5.

In Figure 5, the power flow calculation module first obtains node voltage and power information in real time, and then sends the relevant state quantities to the state processing module and calculates the corresponding rewards, forming the input of the reinforcement learning controller. The DQN controller outputs the optimal action command based on the status. The action is parsed by the action processing module and then handed over to the centralized collaborative controller for execution. The position of the transformer tap is further adjusted to effectively control the voltage of the distribution network. To comprehensively describe the dynamic operating state of the high-penetration photovoltaic distribution network, the node voltages, photovoltaic output power, load demands, energy storage operating state, and regulation device states are jointly constructed as the state space of the DQN controller. Compared with using only node voltages as state variables, the multi-source operating information can reflect the dynamic coupling relationships among the source-grid-load-storage components and provide more comprehensive decision-making information for the controller. The state space is expressed as equation (6).

$$S_t = [V_1(t), V_2(t), \cdots, V_N(t), P_{PV}(t), P_L(t), Q_L(t), SOC(t), T_{OLTC}(t), B_{SVC}(t)] \quad (6)$$

In equation (6), $V_i(t)$ represents the voltage magnitude of the i -th node at time t; $P_{PV}(t)$ represents the photovoltaic output power; $P_L(t)$ and $Q_L(t)$ represent the active and reactive load demands, respectively; $SOC(t)$ represents the state of charge of the energy storage system; $T_{OLTC}(t)$ represents the current tap position of the on-load tap changer; and $B_{SVC}(t)$ represents the equivalent susceptance of the static var compensator. By integrating the operating states of different regulation units, the DQN can perform comprehensive decision-making based on real-time system conditions, thereby improving the adaptability of the multi-device coordinated control strategy. To achieve coordinated control among multiple regulation resources, the action space of DQN is defined as a set of multiple discrete coordination actions, including on-load tap changer regulation, photovoltaic power adjustment levels, energy storage operation modes, and SVC reactive power compensation levels. Since the internal control loops of individual devices are responsible for continuous power tracking, DQN only outputs high-level coordination actions. The action space is formulated as equation (7).

$$A_t = \{a_{OLTC}, a_{PV}, a_{ES}, a_{SVC}\} \quad (7)$$

In equation (7), A_t represents the action space of the DQN controller at time t; a_{OLTC} represents the tap adjustment action of the on-load tap changer; a_{PV} represents the photovoltaic output adjustment level; a_{ES} represents the energy storage charging/discharging mode selection; and a_{SVC} represents the static var compensator adjustment level. Each action variable is represented using discrete levels. For example, photovoltaic output can be configured with different power adjustment ratios, energy storage actions can be represented as charging, maintaining, or discharging modes, and SVC actions can be represented by different reactive power compensation levels. Through this action space design, DQN can perform multi-device coordination decisions within a finite action set while avoiding the rapid expansion of the action space caused by directly optimizing continuous control variables. To balance the operational safety, economic performance, and equipment lifetime of the distribution network, a comprehensive reward function is constructed in this study. The proposed reward function incorporates voltage deviation, network losses, equipment operation costs, and energy storage operating conditions into the optimization objective. When the system operating state changes, the DQN updates the control strategy according to the comprehensive reward generated by the selected action. The reward function is expressed as equation (8) (Ye *et al.*, 2023a).

$$R_{DQN,t} = -\lambda(\Delta\bar{v}_i)^2 \quad (8)$$

In equation (8), R_t represents the immediate reward at time t; $E_V(t)$ represents the voltage deviation index, which is used to evaluate voltage stability; $P_{loss}(t)$ represents the active power loss of the distribution network; $C_{tap}(t)$ represents the tap-switching cost of the on-load tap changer, which is used to reduce frequent switching and extend equipment lifetime; $C_{ES}(t)$ represents the energy storage operation cost, which is used to constrain the charging and discharging process and improve energy utilization efficiency; and w_1, w_2, w_3, w_4 represent the weighting coefficients of different optimization objectives, respectively. By comprehensively considering multiple operational indicators, the reward function enables DQN to maintain voltage stability while reducing network losses, limiting unnecessary equipment actions, and improving energy storage utilization efficiency. The network loss rate and line loading conditions were calculated using the full AC power flow model embedded in the simulation environment. After each control action, bus voltages, branch power flows, and line loading states were updated, and the corresponding voltage and thermal operating constraints were checked. The voltage deviation is calculated in equation (9) (Hou *et al.*, 2025).

$$\Delta\bar{v}_i = \begin{cases} \bar{v}_i - 0.95, & \bar{v}_i > 0.95 \\ 1.05 - \bar{v}_i, & \bar{v}_i < 1.05 \end{cases} \quad (9)$$

In equation (9), $\bar{v}_i$ represents the average unit voltage of the global distribution network bus voltage. After designing the state space, action space and reward function, the DQN-based photovoltaic distribution source-grid-load-storage multi-terminal collaborative control process is shown in Figure 6.

The framework shown in Figure 6 first collects bus voltage information and selects a control strategy based on whether the node voltage exceeds the limit. When a limit violation is detected, the distributed controller issues a reactive power adjustment instruction to restore the voltage. When the voltage is normal, the centralized controller comprehensively determines and coordinates the joint adjustment of photovoltaic, energy storage and reactive power equipment. After meeting the operating safety conditions, the system continuously monitors the voltage status to coordinate and stably control the power supply, network, load, and storage terminals.

3.Results and Analysis

3.1 Comparison of power generation side performance of source-grid-load-storage collaborative control

To verify the voltage control effect and cooperative regulation performance of the proposed method, the research employs reinforcement learning comparative experiments and uses the IEEE 33-node system as the simulation platform for evaluation. The training parameters of the DQN controller were uniformly configured. During training, the learning rate α was set to 0.001, the discount factor γ to 0.95, the replay buffer capacity to 10,000, the batch size to 64, and the target network update interval to 100 steps. The DQN employed a fully connected neural network for state feature mapping and action-value estimation, with 128 and 64 neurons in the two hidden layers, respectively. The ReLU activation function was adopted, and the network parameters were updated using the Adam optimizer.

The simulation experiments were conducted on a platform equipped with an Intel Core i7 processor, 16 GB of RAM, and an NVIDIA RTX 3060 GPU. The total offline training time of the DQN for 500 episodes was approximately 42.6 min, with a peak memory consumption of approximately 2.14 GB during training. After training, the average policy inference time per control step was 8.7 ms, with a maximum inference time of 13.4 ms, which was substantially shorter than the 1 s execution interval of the

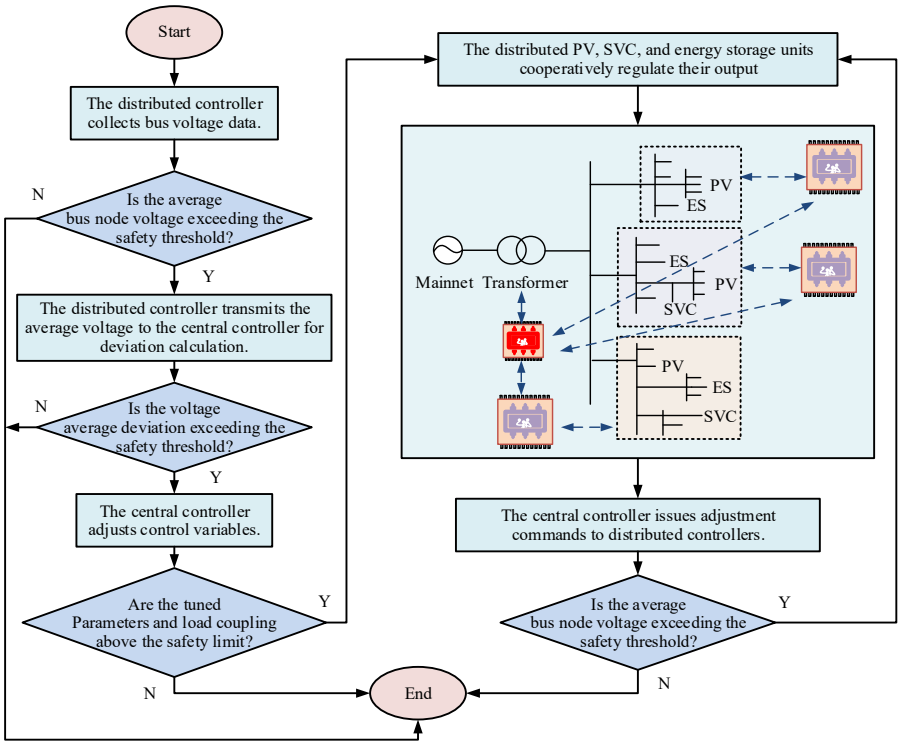


Fig.6 Source-grid-load-storage multi-terminal collaborative control based on DQN

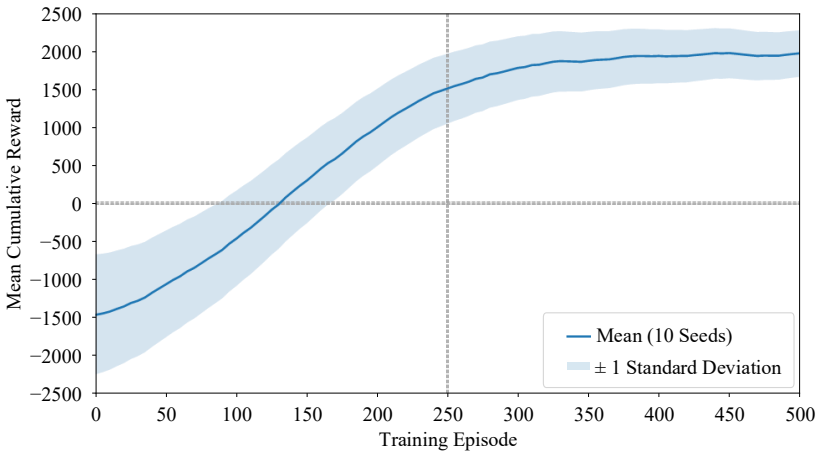


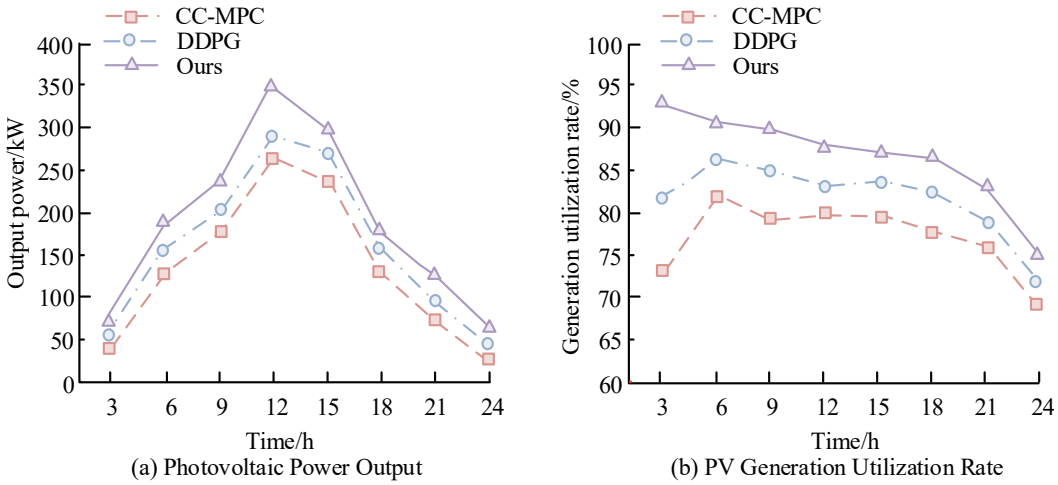
Fig 7 Mean cumulative reward and standard deviation of DQN training under 10 random seeds

lower-level controller. Therefore, under the current simulation hardware configuration, the proposed method can satisfy the computational timeliness requirements of online coordinated control.

The load and photovoltaic data used in this study were obtained from typical practical operating conditions. To simulate the prediction uncertainty of photovoltaic output and load demand in actual operation, zero-mean Gaussian random errors were superimposed on the baseline photovoltaic and load profiles. The standard deviations of the photovoltaic and load prediction errors were set to 5% and 3%, respectively, and the random disturbances at different time steps were assumed to be independent. To prevent the random errors from producing extreme values beyond realistic operating ranges, the disturbance amplitudes were truncated, with the load disturbance limited to within ±10% of the baseline value. The photovoltaic-side errors were used to represent short-term irradiance fluctuations and prediction deviations, while the load-side errors were used to simulate random variations in user power demand. The generated random errors were added to the corresponding baseline power profiles at each time step to construct operating scenarios with prediction uncertainty.

First, the convergence of the DQN training process and its stability under different random initializations were evaluated. Ten different random seeds were used for independent training, and the mean and standard deviation of the cumulative reward at each training episode were calculated. The results are shown in Figure 7. As shown in Figure 7, the mean cumulative reward of DQN gradually increases with the number of training episodes and tends to stabilize after approximately 400 episodes, indicating that the policy learning process has essentially converged. Meanwhile, the standard deviation band gradually narrows as training proceeds, indicating a decrease in the dispersion of the results across different random seeds. Overall, the proposed method exhibits a consistent convergence trend under different random initializations, demonstrating good training stability and reproducibility. It operates in a power distribution environment with high penetration of photovoltaic access, and compared with the Deep Deterministic Policy Gradient (DDPG) method and the stochastic real-time control method based on Chance-Constrained Model Predictive Control (CC-MPC). The advantages of the strategy in terms of voltage stability, accommodation capability and equipment action friendliness are verified. The results of changes in photovoltaic output and power generation utilization under the three control methods are shown in Figure 8.

Figure 8(a) and Figure 8(b) show the changes in photovoltaic output and power generation utilization using different methods, respectively. In Figure 8(a), at 3 h, the output powers of CC-MPC, DDPG and the proposed method were 40kW, 60kW and 70kW, respectively. The curve peaked at around 12 hours at noon, with CC-MPC at about 260kW, DDPG at about 280kW, and the method proposed in the study reaching 360kW, which was the highest point throughout the day. After that, the output gradually decreased. At 18 hours, the three



(a) Photovoltaic Power Output (b) PV Generation Utilization Rate

Fig.8 Changes in photovoltaic output and power generation utilization rate

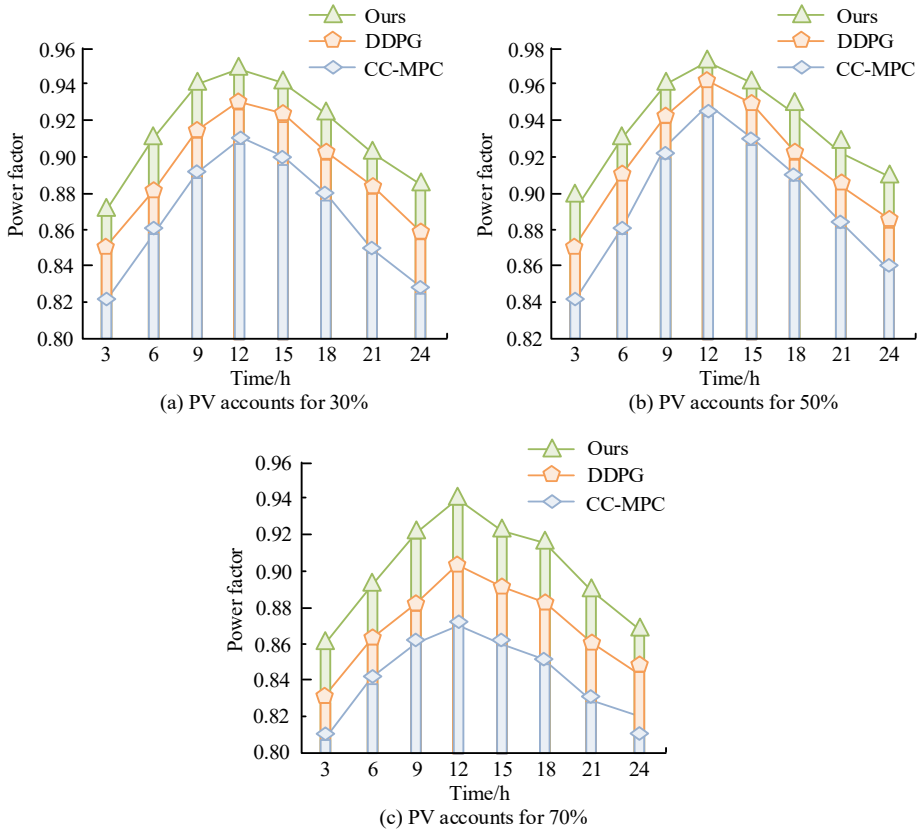


Fig.9 Process of power factor curve variation

outputs dropped to 140kW, 160kW, and 180kW, respectively. At the end of 24 hours, it was close to 45kW, 50kW, and 75kW. Overall, the control curve of the proposed method was always higher than that of DDPG and CC-MPC throughout the entire period, with the peak value increased by about 20%-25%, and the rising and falling stages were smoother, indicating that it could make full use of light and suppress fluctuations. In Figure 8(b), the three curves were 74%, 82%, and 93% respectively at 3 h, and then increased as the irradiation increased. By 9h, the utilization rate of the proposed method was 90%. After that, the utilization rate dropped slightly, from 87% at 18h and to 75% at 24h. Generally speaking, the power generation utilization rate shows an "increasing-stable-decreasing" trend over time, and this method controls to maintain the highest level at all times throughout the day. This shows that the proposed method optimizes the output distribution through reinforcement learning, so that photovoltaic power generation can achieve higher energy conversion efficiency and operational stability under different irradiation conditions. The changing process of the power factor curve of different methods under different PV capacity ratios is shown in Figure 9.

Figure 9(a), Figure 9(b) and Figure 9(c) show the power factor changes of the three control strategies during a typical day and night cycle when the PV capacity ratio is 30%, 50% and 70%, respectively. In Figure 9(a), the PV ratio was 30%. The power factors of each method were 0.82, 0.85, and 0.87 respectively at 3 hours in the morning. As the sunshine increases, they gradually increase, reaching peak values of 0.91, 0.93, and 0.95 at 12 hours, and then slowly decrease to 0.83, 0.86, and 0.89 at 24 hours. The overall trend shows that the power factor improvement is limited under low ratio conditions, and the proposed control method maintains the highest value at all times. In Figure 9(b), the overall level of the three curves was significantly improved. The power factors were 0.84, 0.87, and 0.90 respectively at 3 hours, reaching peak values of 0.94, 0.96, and 0.97 at 12 hours, and then gradually falling back to 0.86,

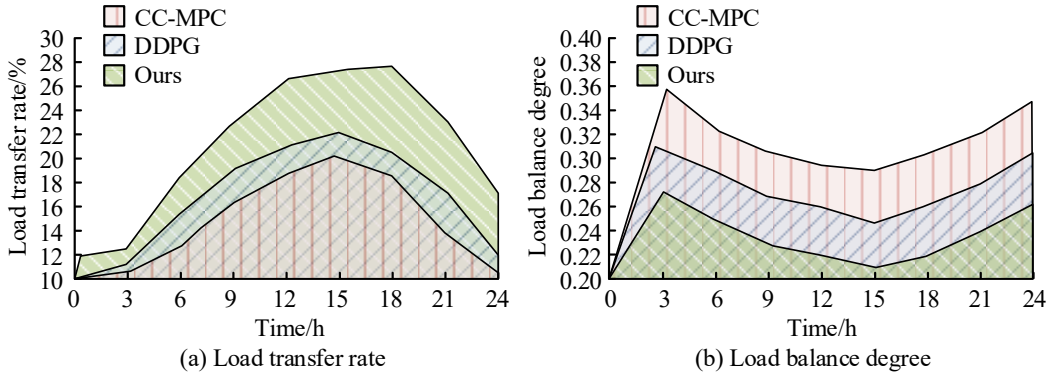


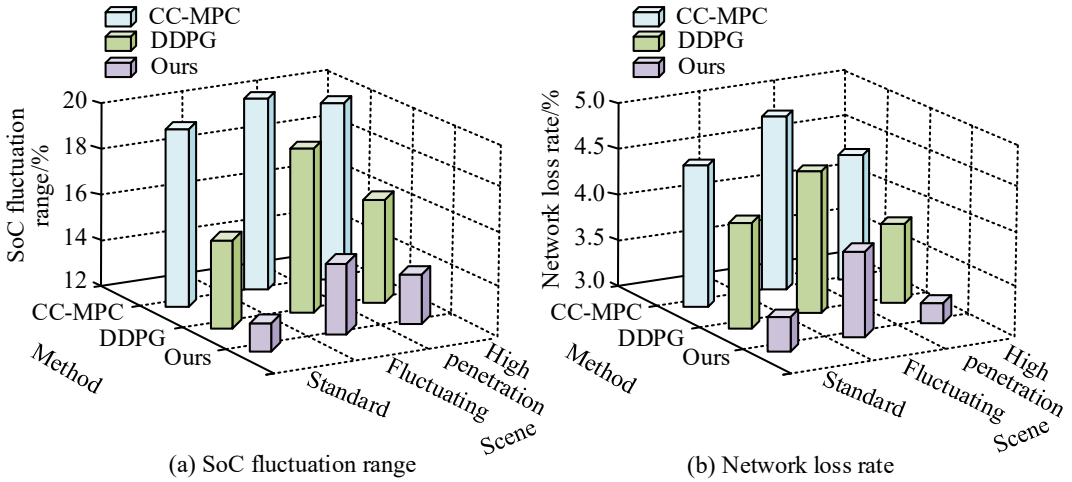
Fig.10 The process of changes in load transfer rate and balance degree

0.89, and 0.91 at 24 hours. At this time, the power factor remained at a high level, indicating that the coordination of active and reactive power was best when the ratio was 50%. In Figure 9(c), the proportion of photovoltaics increased to 70%, and the curves of the three methods overall shifted slightly downward compared to the first two groups. They were 0.80, 0.83, and 0.86 respectively at 3 hours, and the peak values at 12 hours were 0.87, 0.90, and 0.94, respectively, and fell back to 0.81, 0.85, and 0.87 at night. Under high penetration rates, the voltage regulation pressure increased and the system power factor decreased slightly, but the curve of the proposed method still remained ahead, and the peak value was about 6% higher than that of CC-MPC. In summary, it can be seen that the proposed control method achieves the highest power factor and smoothest curve in all scenarios, indicating that it has the best dynamic coordination capabilities in multi-source collaboration and reactive power optimization. To further quantify the effect of peak shaving and valley filling, the load transfer rate and balance degree are calculated, and the results are shown in Figure 10.

Figure 10(a) and Figure 10(b) respectively show the change process of load transfer rate and load balance degree over time under different control strategies. In Figure 10(a), the load transfer rates of CC-MPC, DDPG and the proposed method were 10.8%, 11.2% and 12.4% respectively at 3h. With the enhancement of time-of-use electricity price incentives and energy storage participation, they increased to 16.0%, 19.1% and 22.8% at 9h, and reached the peak value at 15h, which were 19.6%, 21.8% and 26.2% respectively. After that, the transfer rate gradually decreased, reaching 10.9%, 12.4% and 17.9% respectively at 24h. Overall, the curves of the proposed method were higher than those of DDPG and CC-MPC at all times, especially during the peak period of 12-18h. The results show that this control method can more accurately respond to time-of-use electricity price signals, coordinate energy storage and load, and achieve a higher proportion of peak shaving and valley filling. In Figure 10(b), at 3h, the load balance degree of CC-MPC, DDPG and the proposed method were 0.36, 0.31, and 0.27, respectively. With the continuous optimization of the control strategy, they dropped to 0.29, 0.25, and 0.21 respectively at 15h, which were the lowest points throughout the day. They then recovered slightly and were 0.35, 0.30, and 0.26 respectively at 24h. Compared with CC-MPC, the balance degree of the proposed method decreased by 0.11 on average, indicating that this method can effectively balance the load of each node and improve the coordination and stability of the power grid operation.

3.2 System comprehensive effect analysis of source-grid-load-storage collaborative control

To further verify the stability and comprehensive performance of the proposed multi-time scale collaborative control strategy based on DQN under different operating environments, simulation analysis was carried out from three aspects: system voltage quality, energy storage dynamic response and energy loss. The voltage amplitude of each bus node after adjustment by different control strategies is shown in Figure 11. Figures 11(a), 11(b), and 11(c) show the temporal variations in node voltage magnitude under different control methods. To reduce the influence of random initialization and the training process of reinforcement learning on the results, each method was independently evaluated using five different random seeds, and the statistical results are reported as mean ± standard deviation. CC-MPC exhibits the largest voltage fluctuation, with an average fluctuation amplitude of 0.169 ± 0.011 p.u. and a peak voltage of 1.112 ± 0.009 p.u., repeatedly exceeding the safety upper limit of 1.05 p.u., which indicates pronounced over-regulation and under-regulation under disturbance conditions. The voltage fluctuation under DDPG is reduced, with an average fluctuation amplitude of 0.151 ± 0.008 p.u. and a peak voltage of 1.086 ± 0.007 p.u.; however, the stabilization process remains relatively slow. In comparison, the proposed method produces a smoother voltage response, with an average fluctuation amplitude of 0.143 ± 0.006 p.u. and a steady-state voltage deviation of 0.008 ± 0.002 p.u. Moreover, the results obtained under different random seeds show relatively low dispersion. These results indicate that the proposed DQN-based coordinated control strategy can maintain stable voltage regulation performance under different training initializations, demonstrating better dynamic smoothness and result consistency. To comprehensively evaluate the overall performance of the proposed DQN-based source-grid-load-storage collaborative control strategy under different operating conditions, the study set up three typical scenarios: The standard scenario simulates stable operation on sunny days to verify the baseline performance, the fluctuating scenario introduces light intermittent and ±10% load disturbance to test the system robustness, and the high-penetration scenario sets the photovoltaic capacity to account for 70% to examine the operating efficiency under a high proportion of renewable energy. The three together constitute a typical operating environment and provide sufficient basis for comparing control methods. The network loss rate results in different battery State of Charge (SoC) fluctuation ranges are shown in Figure 12.



(a) SoC fluctuation range (b) Network loss rate

Fig. 12 Comparison results of SoC fluctuation range and network loss rate

Table 1
Comparison of equipment action smoothness under different control strategies

Indicator	CC-MPC	DDPG	Proposed method
Number of OLTC operations within 24 h / times	22	17	11
Average interval between adjacent OLTC operations / min	65.8	86.4	132.7
Number of reverse tap-switching operations / times	6	4	1
Maximum variation rate of PV power command / %	12.8	10.4	7.1
Maximum variation rate of ES power command / %	15.6	12.3	8.5
Maximum variation rate of SVC reactive power command / %	13.9	11.1	7.6

Table 2
Energy storage utilization rates and system operating costs of different methods

Scene Type	Methods	Energy storage utilization rate (%)	System operating cost (¥/kWh)
Standard scenario	CC-MPC	71.4	0.734
	DDPG	78.2	0.691
	Ours	86.3	0.653
Fluctuating scenario	CC-MPC	68.9	0.751
	DDPG	74.5	0.705
	Ours	83.1	0.669
High penetration scenario	CC-MPC	75.8	0.728
	DDPG	80.5	0.686
	Ours	88.9	0.648

Figures 12(a) and 12(b) present the SoC fluctuation range and network loss rate, respectively, for different control strategies under three operating scenarios. All results were obtained from five independent runs with different random seeds and are reported as mean ± standard deviation. As shown in Figure. 12(a), the SoC fluctuation rates of CC-MPC under the standard, fluctuating, and high-penetration scenarios were 18.4 ± 1.2%, 19.8 ± 1.4%, and 19.2 ± 1.3%, respectively. For DDPG, the corresponding values decreased to 14.7 ± 0.9%, 17.9 ± 1.1%, and 15.3 ± 1.0%. The proposed method achieved the lowest fluctuation rates, at only 11.9 ± 0.6%, 13.8 ± 0.8%, and 12.1 ± 0.7%, respectively. As the operating scenarios became more complex, the SoC fluctuations of CC-MPC and DDPG increased, whereas the proposed method maintained relatively stable performance, with variations among the three scenarios remaining within 2% and relatively small standard deviations. This indicates better smoothness and stability during the charging and discharging process. As shown in Figure 12(b), the network loss rates of CC-MPC under the three scenarios were 4.32 ± 0.18%, 4.58 ± 0.21%, and 4.12 ± 0.17%, respectively. The corresponding values for DDPG were reduced to 3.76 ± 0.14%, 4.09 ± 0.16%, and 3.69 ± 0.13%, while the proposed method further reduced them to 3.18 ± 0.10%, 3.54 ± 0.12%, and 3.10 ± 0.09%, respectively. The relatively small dispersion across different random seeds indicates that the proposed method can consistently reduce energy storage power fluctuations and network losses under different operating scenarios. To further evaluate the smoothness of equipment actions under different control strategies, the OLTC tap operations and the power regulation command variations of PV, ES, and SVC over a 24 h operating cycle were statistically analyzed, as shown in Table 1. As shown in Table 1, the numbers of OLTC operations within 24 h under CC-MPC, DDPG, and the proposed method were 22, 17, and 11, respectively. Compared with CC-MPC and DDPG, the proposed method reduced the number of OLTC operations by 50.00% and 35.29%, respectively. The average interval between adjacent OLTC operations reached 132.7 min, while only one reverse tap-switching operation occurred, indicating that the proposed method can effectively suppress frequent tap adjustments. Meanwhile, the maximum variation rates of the PV, ES, and SVC power commands under the proposed method were 7.1%, 8.5%, and 7.6%, respectively, all lower than those of the two comparative methods. These results indicate that the proposed multi-time-scale coordinated control strategy can reduce frequent operations of mechanical regulation devices and abrupt changes in power commands while maintaining dynamic regulation capability, thereby improving the smoothness of the control process. To evaluate the overall performance under source-grid-load-storage coupling, the energy storage utilization and system operating costs of the three methods in different scenarios are compared. The results are shown in Table 2.

In Table 2, the energy storage utilization rate and system operation cost of the three control methods in different scenarios showed an obvious reverse trend, and the proposed method performed optimally in each scenario. In the standard scenario, the energy storage utilization rate of the proposed method reached 86.3%, which was about 8.1% higher than that of DDPG and about 14.9% higher than that of CC-MPC. The operating cost was the lowest, only 0.653 ¥/kWh, which was about 0.081 ¥/kWh lower than that of CC-MPC and about 0.038 ¥/kWh lower than that of DDPG. Under the fluctuation scenario, the overall performance of the three methods has declined, but the proposed method still maintains the highest utilization rate of 83.1%, with the smallest increase in cost and stable control effect. The performance of CC-MPC dropped most significantly in this scenario, with the energy storage utilization rate falling to 68.9% and the cost rising to 0.751 ¥/kWh. In high-penetration scenarios, as the proportion of photovoltaic output increases, the optimization advantages of the proposed method are further highlighted. Its energy storage utilization rate reached 88.9% and the cost dropped to 0.648 ¥/kWh. Overall, the proposed method can significantly improve the utilization efficiency of energy storage resources and reduce system economic costs under different operating conditions, showing excellent robustness and global coordination capabilities.

Finally, to verify the adaptability of the proposed method under high-temporal-resolution data and different network topologies, MADDPG and MAPPO were introduced as comparison methods. The PV and load data were represented by 24 h operating sequences with a temporal resolution of 1 min, resulting in 1,440 consecutive state points per test cycle to characterize short-term PV output variations and stochastic

Table 3
Performance comparison of different reinforcement learning methods under different network topologies

Test scenario	Method	PV utilization rate / %	Voltage violation rate / %	Network loss rate / %	Energy storage utilization rate / %	Inference time / ms
IEEE 33-bus standard topology	MADDPG	90.8 ± 1.3	1.84 ± 0.21	3.47 ± 0.13	84.6 ± 1.2	11.6 ± 0.8
	MAPPO	91.7 ± 1.1	1.52 ± 0.18	3.35 ± 0.11	85.8 ± 1.0	12.9 ± 0.9
	Ours	93.1 ± 0.8	0.96 ± 0.12	3.16 ± 0.09	87.2 ± 0.8	8.9 ± 0.6
IEEE 33-bus reconfigured topology	MADDPG	87.9 ± 1.6	3.26 ± 0.34	3.91 ± 0.16	81.7 ± 1.5	11.8 ± 0.9
	MAPPO	89.1 ± 1.4	2.71 ± 0.29	3.76 ± 0.14	83.2 ± 1.3	13.1 ± 1.0
	Ours	91.5 ± 1.0	1.63 ± 0.20	3.42 ± 0.11	85.6 ± 1.0	9.2 ± 0.7
IEEE 119-bus system	MADDPG	86.8 ± 1.8	3.74 ± 0.39	4.08 ± 0.18	80.9 ± 1.6	18.7 ± 1.2
	MAPPO	88.3 ± 1.5	3.08 ± 0.31	3.89 ± 0.16	82.7 ± 1.4	20.4 ± 1.4
	Ours	90.6 ± 1.2	1.91 ± 0.23	3.53 ± 0.13	84.9 ± 1.1	14.8 ± 1.0

load fluctuations. In addition to the standard IEEE 33-bus topology, an IEEE 33-bus reconfigured topology and the IEEE 119-bus system were further considered. All methods used the same equipment parameters and operating constraints, and five independent training and testing runs with different random seeds were conducted. The results are shown in Table 3.

As shown in Table 3, under the 1 min high-temporal-resolution data, the proposed method achieved a PV utilization rate of 93.1 ± 0.8%, a voltage violation rate of 0.96 ± 0.12%, a network loss rate of 3.16 ± 0.09%, and an energy storage utilization rate of 87.2 ± 0.8% under the standard IEEE 33-bus topology, outperforming MADDPG and MAPPO overall. After a line outage and network reconfiguration in the IEEE 33-bus system, the PV utilization rate of the proposed method decreased to 91.5%, while the voltage violation rate increased to 1.63%, which remained lower than those of MADDPG and MAPPO, indicating good adaptability to topology changes. In the IEEE 119-bus system, the PV utilization rate, voltage violation rate, and network loss rate of the proposed method were 90.6 ± 1.2%, 1.91 ± 0.23%, and 3.53 ± 0.13%, respectively. The average inference time was 14.8 ms, which remained substantially lower than the 1 s control interval, demonstrating good control performance and computational feasibility as the network scale increased.

4.Discussion

The results show that the proposed DQN-based source-grid-load-storage coordinated control method outperforms CC-MPC and DDPG in photovoltaic accommodation, voltage stability, load transfer, energy storage utilization, and network loss reduction. The photovoltaic power generation utilization rate reaches 93%, the maximum load transfer rate reaches 26.2%, the energy storage utilization rate reaches 88.9% under the high-penetration scenario, and the network loss rate is reduced to 3.10%. These improvements indicate that the proposed controller can coordinate multiple regulation resources more effectively when photovoltaic output and load demand vary simultaneously. Recent studies have explored coordinated control from different perspectives. Wang *et al.* (2025) coordinated large-scale photovoltaic inverters using multi-agent deep reinforcement learning, while Ye *et al.* (2023b) focused on coordinated energy trading and flexibility services. Doostizadeh *et al.* (2024) further considered decentralized economic and resilient operation with demand response. Similar trends can also be observed in recent studies on distribution-system and virtual-power-plant coordination, network reconfiguration, and decentralized Volt/Var control (Sun and Lu, 2024; Gao *et al.*, 2024; Yan and Xu, 2025). These studies mainly demonstrate the value of coordinated decision-making among heterogeneous flexible resources. In comparison, the present study simultaneously integrates OLTC, photovoltaic generation, energy storage, SVC, and load-side response into a unified multi-time-scale framework. Its main advantage therefore lies in multi-terminal source-grid-load-storage coordination rather than the optimization of a single regulation resource or operating objective. The hierarchical coordination between slower global regulation and faster local response also helps reduce unnecessary control actions while maintaining acceptable voltage quality.

The proposed method also maintains energy storage utilization rates of 83.1% and 88.9% under fluctuating and high-penetration scenarios, with corresponding network loss rates of 3.54% and 3.10%. This suggests that the learned policy is not limited to a single operating condition, but can preserve relatively stable regulation performance when photovoltaic output and load demand are disturbed. Existing studies have shown that distributed flexible resources and transactive energy mechanisms can improve adaptive energy management under uncertain operating conditions (Xu *et al.*, 2024; Mohsenzadeh-Yazdi *et al.*, 2024). Meanwhile, increasing attention has been given to the safety and robustness of data-driven grid control. Eichelbeck *et al.* (2026), Hou *et al.* (2025), and Shalaby *et al.* (2025) emphasized explicit operational constraints in reinforcement learning, while Yan *et al.* (2024), Yang *et al.* (2024), and Zhang *et al.* (2024) further considered safe Volt/Var control, observation perturbations, and low-carbon demand management. These studies indicate that safety constraints, robustness to uncertain observations, and feasible action generation are becoming essential for practical reinforcement learning-based grid control. Compared with these methods, the proposed DQN framework currently relies mainly on the reward function and predefined operating limits to discourage infeasible or uneconomic actions. This design is simple and computationally efficient, but its safety mechanism is still weaker than methods with explicit constraint projection, safety filters, or formally defined feasible action sets.

From the perspective of flexibility and scalability, Rousseau *et al.* (2026), Juma *et al.* (2024), and Shahbazian *et al.* (2024) highlighted the value of flexible loads and coordinated microgrid operation under uncertainty. More recent studies have also investigated topology adaptation and large-scale applicability through graph-based reinforcement learning and networked microgrid scheduling (Ge *et al.*, 2025; Kaloti and Chowdhury, 2025). These developments suggest that future intelligent distribution control should not only pursue higher performance under a fixed test system, but also maintain adaptability under changing topology, communication uncertainty, operating disturbances, and larger network scales. In the present study, the proposed DQN strategy demonstrates

good performance on the selected simulation systems and operating scenarios, but its generalization to unseen network structures and field-level disturbances still requires further validation. In addition, the present framework assumes that sufficient historical or simulated operating samples are available for offline training, which may increase the preparation cost when the operating environment changes substantially. Future work will therefore incorporate safety-constrained reinforcement learning, topology-aware state representation, and larger-scale distribution networks with time-varying operating conditions to further improve the robustness, transferability, and engineering applicability of the proposed framework.

5. Conclusion

To address voltage fluctuation, power imbalance and insufficient energy storage utilization in high-penetration photovoltaic distribution networks, this study proposed a DQN-based multi-time-scale collaborative optimization method for source-grid-load-storage coordinated control. The proposed method integrates voltage regulator tap adjustment, photovoltaic output regulation, energy storage charging and discharging, and reactive power compensation into a unified reinforcement learning framework, enabling coordinated decision-making among multiple regulation resources under dynamic operating conditions. The simulation results based on the IEEE 33-bus system demonstrate that the proposed method can effectively enhance the operational performance of photovoltaic distribution networks. The photovoltaic power generation utilization rate is increased by up to 25%, the load transfer rate reaches 26.2%, and the voltage fluctuation is controlled within 0.15 p.u. Under different operating scenarios, the energy storage utilization rate is improved by approximately 15%-20%, the SoC fluctuation range is reduced to about 12%, the network loss rate decreases to 3.1%, and the system operating cost is reduced by about 11%. These results confirm that the proposed DQN-based collaborative control framework provides better voltage regulation, smoother energy dispatch and higher economic efficiency than CC-MPC and DDPG. Overall, the main conclusion of this study is that multi-terminal coordination among source, grid, load and storage can effectively improve the stability, flexibility and economy of distribution networks with high renewable energy penetration. The proposed method provides an adaptive control solution that reduces the dependence on fixed model parameters and improves the ability of the system to respond to photovoltaic fluctuation and load variation. Nevertheless, this study is still mainly based on simulation verification, and the training process requires sufficient operating samples. Future work will further extend the method to larger-scale distribution networks, multi-agent distributed reinforcement learning, real-time edge computing and field operation scenarios, so as to improve its generalization ability, real-time performance and engineering applicability.

Author Contributions: W. N.: Conceptualization, methodology, formal analysis, writing—original draft, S.J.: supervision, resources, project administration, Z.W.: writing—review and editing, project administration, validation, L.L. and T.C.: writing—review and editing, project administration, validation. All authors have read and agreed to the published version of the manuscript.

Funding: The authors received no financial support for the research, authorship, and/or publication of this article.

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

References

Amer, A. A., Shaban, K. B., & Massoud, A. M. (2023). DRL-HEMS: Deep reinforcement learning agent for demand response in home energy management systems considering customers and operators perspectives. *IEEE Transactions on Smart Grid*, 14(1), 239–250. https://doi.org/10.1109/TSG.2022.3198401

Bosisio, A., Soldan, F., Pisani, M., Bionda, E., Belloni, F., & Morotti, A. (2025). A Q-learning algorithm for optimizing on-load tap changer operation and voltage control in distribution networks with high integration of renewable energy sources. *Journal of Modern Power Systems and Clean Energy*, 13(6), 2063–2073. https://doi.org/10.35833/MPCE.2024.000528

Chaturvedi, S., Bui, V. H., Su, W., & Wang, M. (2024). Reinforcement learning-based integrated control to improve the efficiency of DC microgrids. *IEEE Transactions on Smart Grid*, 15(1), 149–159. https://doi.org/10.1109/TSG.2023.3286801

Dang, Y., Xu, J., Yang, F., Jiang, C., & Li, D. (2025). Meta reinforcement learning based adaptive and interpretable energy storage control meets dynamic scenarios. *IEEE Transactions on Sustainable Energy,* 16(4), 2560–2572. https://doi.org/10.1109/TSTE.2025.3555002

Doostizadeh, M., Jalili, H., & Babaee, A. (2024). A decentralized framework for economic-resilient operation of networked microgrids, considering demand response program. *IET Renewable Power Generation,* 18(8), 1428–1454. https://doi.org/10.1049/rpg2.12984

Eichelbeck, M., Markgraf, H., & Althoff, M. (2026). CommonPower: A framework for safe data-driven smart grid control. *IEEE Transactions on Smart Grid,* 17(1), 71–82. https://doi.org/10.1109/TSG.2025.3616402

Gao, H., Jiang, S., Li, Z., Wang, R., Liu, Y., & Liu, J. (2024). A two-stage multi-agent deep reinforcement learning method for urban distribution network reconfiguration considering switch contribution. *IEEE Transactions on Power Systems*, 39(6), 7064–7076. https://doi.org/10.1109/TPWRS.2024.3371093

Ge, L., Li, J., Hou, L., & Lai, J. (2025). Autonomous voltage regulation for smart distribution network with high-proportion PVs: A graph meta-reinforcement learning approach. *IEEE Transactions on Sustainable Energy,* 16(4), 2768–2781. https://doi.org/10.1109/TSTE.2025.3566469

Guan, Y., Ma, W., Che, L., & Shahidehpour, M. (2025). Model-based safe reinforcement learning for active distribution network scheduling. *IEEE Transactions on Smart Grid*, 16(3), 2375–2388. https://doi.org/10.1109/TSG.2025.3547843

Hong, L., Wu, M., Wang, Y., Yan, M., Chen, Z., & Yan, Z. (2024). MADRL-based DSO-customer coordinated bi-level Volt/VAR optimization method for power distribution networks. *IEEE Transactions on Sustainable Energy,* 15(3), 1834–1846. https://doi.org/10.1109/TSTE.2024.3380605

Hou, S., Fu, A., Salazar Duque, E. M., Palensky, P., Chen, Q., & Vergara, P. P. (2025). DistFlow safe reinforcement learning algorithm for voltage magnitude regulation in distribution networks. *Journal of Modern Power Systems and Clean Energy,* 13(1), 300–311. https://doi.org/10.35833/MPCE.2024.000253

Irmak, E., Kabalcı, E., & Calpbinici, A. (2024). Event-triggered distributed secondary control for enhancing efficiency, reliability and communication in island mode DC microgrids. *IET Renewable Power Generation,* 18(1), 78–94. https://doi.org/10.1049/rpg2.12897

Juma, S. A., Ayeng'o, S. P., & Kimambo, C. Z. M. (2024). A review of control strategies for optimized microgrid operations. *IET Renewable Power Generation,* 18(14), 2785–2818. https://doi.org/10.1049/rpg2.13056

Kaloti, S., & Chowdhury, B. (2025). Dual agent framework for scheduling networked microgrids using DRL to improve resilience. *IEEE Transactions on Sustainable Energy*, 16(4), 2989–3002. https://doi.org/10.1109/TSTE.2025.3576153

Lee, S., Seon, J., Sun, Y. G., Kim, S. H., Kyeong, C., Kim, D. I., & Kim, J. Y. (2024). Novel architecture of energy management systems based on deep reinforcement learning in microgrid. *IEEE*

Transactions on Smart Grid, 15(2), 1646–1658. https://doi.org/10.1109/TSG.2023.3317096

Locher, L., Stai, E., & Hug, G. (2026). A safe combined reinforcement learning and model predictive control scheme for utility-level battery control in distribution grids. *IEEE Transactions on Smart Grid,* 17(1), 144–160. https://doi.org/10.1109/TSG.2025.3617966

Manojkumar, R., Kumar, C., Ganguly, S., & Catalão, J. P. S. (2024). Net load forecast error compensation for peak shaving in a grid-connected PV storage system. *IEEE Transactions on Power Systems,* 39(2), 4372–4381. https://doi.org/10.1109/TPWRS.2023.3317309

Meng, Q., Hussain, S., Luo, F., Wang, Z., & Jin, X. (2025). An online reinforcement learning-based energy management strategy for microgrids with centralized control. *IEEE Transactions on Industry Applications,* 61(1), 1501–1510. https://doi.org/10.1109/TIA.2024.3430264

Mohsenzadeh-Yazdi, H., Kebriaei, H., & Aminifar, F. (2024). Multi-agent reinforcement learning in a new transactive energy mechanism. *IET Generation, Transmission & Distribution*, 18(18), 2943–2955. https://doi.org/10.1049/gtd2.13244

Mohamed, M., & Tsuji, T. (2023). Battery scheduling control of a microgrid trading with utility grid using deep reinforcement learning. *IEEJ Transactions on Electrical and Electronic Engineering,* 18(5), 665–677. https://doi.org/10.1002/tee.23768

Mu, C., Liu, Z., Yan, J., Jia, H., & Zhang, X. (2024). Graph multi-agent reinforcement learning for inverter-based active voltage control. *IEEE Transactions on Smart Grid,* 15(2), 1399–1409. https://doi.org/10.1109/TSG.2023.3298807

Mukherjee, S., Hossain, R. R., Mohiuddin, S. M., Liu, Y., Du, W., Adetola, V., Jinsiwale, R. A., Huang, Q., Yin, T., & Singhal, A. (2025). Resilient control of networked microgrids using vertical federated reinforcement learning: Designs and real-time test-bed validations. *IEEE Transactions on Smart Grid*, 16(2), 1897–1910. https://doi.org/10.1109/TSG.2024.3466768

Nematshahi, S., Shi, D., Wang, F., Yan, B., & Nair, A. (2023). Deep reinforcement learning based voltage control revisited. *IET Generation, Transmission & Distribution,* 17(21), 4826–4835. https://doi.org/10.1049/gtd2.13001

Peng, H., Liao, K., Yang, J., Pang, B., & He, Z. (2025). Deep reinforcement learning based multi-timescale Volt/Var control in distribution networks considering network reconfiguration. *IEEE Transactions on Sustainable Energy,* 16(4), 2948–2958. https://doi.org/10.1109/TSTE.2025.3574806

Rousseau, J., Cai, H., Heer, P., Orehounig, K., & Hug, G. (2026). Uncertainty-aware flexibility of HVAC systems in buildings: From quantification to provision. *IEEE Transactions on Smart Grid*, 17(1), 269–282. https://doi.org/10.1109/TSG.2025.3611455

Shahbazian, S., Kharrati, S., & Rastgou, A. (2024). Developing a cooperative approach under normal and contingency conditions for generation expansion planning of microgrids. *IET Renewable Power Generation,* 18(13), 2065–2079. https://doi.org/10.1049/rpg2.13034

Shalaby, A. A., Abdeltawab, H., & Mohamed, Y. A. R. I. (2025). Safe deep reinforcement learning for resilient self-proactive distribution grids against wildfires. *IEEE Transactions on Smart Grid*, 16(6), 5065–5082. https://doi.org/10.1109/TSG.2025.3602849

Stanojev, O., Guo, Y., & Hug, G. (2025). Online feedback droop scheduling in distribution grids for frequency and local voltage control. *IEEE Transactions on Power Systems*, 40(5), 4105–4118. https://doi.org/10.1109/TPWRS.2025.3528752

Sun, Z., & Lu, T. (2024). Collaborative operation optimization of distribution system and virtual power plants using multi-agent deep reinforcement learning with parameter-sharing mechanism. *IET Generation, Transmission & Distribution*, 18(1), 39–49. https://doi.org/10.1049/gtd2.13037

Vu, L., Vu, T., Vu, T. L., & Srivastava, A. (2024). Multi-agent deep reinforcement learning for distributed load restoration. *IEEE Transactions on Smart Grid,* 15(2), 1749–1760. https://doi.org/10.1109/TSG.2023.3310893

Wang, R., Bi, X., & Bu, S. (2024). Real-time coordination of dynamic network reconfiguration and Volt-VAR control in active distribution network: A graph-aware deep reinforcement learning approach. *IEEE Transactions on Smart Grid*, 15(3), 3288–3302. https://doi.org/10.1109/TSG.2023.3324474

Wang, Y., Hu, W., Cao, D., Zhao, P., Abulanwar, S., Chen, Z., & Blaabjerg, F. (2025). Local distribution voltage control using large-scale coordinated PV inverters: A novel multi-agent deep reinforcement learning-based approach. *IEEE Transactions on Smart Grid,* 16(3), 2683–2686. https://doi.org/10.1109/TSG.2025.3533958

Xu, T., Chen, T., Gao, C., Song, M., Wang, Y., & Yuan, H. (2024). Distributed flexible resource regulation strategy for residential communities based on deep reinforcement learning. IET Generation, Transmission & Distribution, 18(21), 3378–3391. https://doi.org/10.1049/gtd2.13284

Yan, R., Xing, Q., & Xu, Y. (2024). Multi-agent safe graph reinforcement learning for PV inverters-based real-time decentralized Volt/Var control in zoned distribution networks. *IEEE Transactions on Smart Grid,* 15(1), 299–311. https://doi.org/10.1109/TSG.2023.3277087

Yan, R., & Xu, Y. (2025). Multi-objective and multi-agent deep reinforcement learning for real-time decentralized Volt/VAR control of distribution networks considering PV inverter lifetime. *IEEE Transactions on Power Systems*, 40(2), 1558–1569. https://doi.org/10.1109/TPWRS.2024.3452154

Yang, X., Liu, H., & Wu, W. (2024). Attention-enhanced multi-agent reinforcement learning against observation perturbations for distributed Volt-VAR control. *IEEE Transactions on Smart Grid,* 15(6), 5761–5772. https://doi.org/10.1109/TSG.2024.3423700

Ye, Y., Papadaskalopoulos, D. A., Yuan, Q., Tang, Y., & Strbac, G. (2023b). Multi-agent deep reinforcement learning for coordinated energy trading and flexibility services provision in local electricity markets. *IEEE Transactions on Smart Grid*, 14(2), 1541–1554. https://doi.org/10.1109/TSG.2022.3149266

Ye, Y., Wang, H., Chen, P., Tang, Y., & Strbac, G. (2023a). Safe deep reinforcement learning for microgrid energy management in distribution networks with leveraged spatial-temporal perception. *IEEE Transactions on Smart Grid*, 14(5), 3759–3775. https://doi.org/10.1109/TSG.2023.3243170

Yu, P., Zhang, H., & Song, Y. (2024). Adaptive tie-line power smoothing with renewable generation based on risk-aware reinforcement learning. IEEE Transactions on Power Systems, 39(6), 6819–6832. https://doi.org/10.1109/TPWRS.2024.3379513

Zhang, J., Sang, L., Xu, Y., & Sun, H. (2024). Networked multiagent-based safe reinforcement learning for low-carbon demand management in distribution networks. *IEEE Transactions on Sustainable Energy,* 15(3), 1528–1545. https://doi.org/10.1109/TSTE.2024.3355123

© 2026. The Author(s). This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-ShareAlike 4.0 (CC BY-SA) International License (http://creativecommons.org/licenses/by-sa/4.0/)