scientific reports

OPEN

Resilience driven EV coordination in multiple microgrids using distributed deep reinforcement learning

Yuxin Wu[1], Ting Cai[2]✉ & Xiaoli Li[1]

By integrating electric vehicles (EVs), the multi-microgrids (MMGs) can significantly enhance their resilient operation capabilities. However, existing works face challenges in formulating optimal routing and scheduling strategies for EVs, due to the spatial-temporal uncertainty of the distribution and transportation networks, as well as incomplete information. This paper addresses the coordination problem of EVs for the resilience enhancement of MMGs, using a distributed multi-agent deep reinforcement learning approach to minimize the load-shedding cost. Specifically, a coupled power-transportation network (CPTN) model is constructed to facilitate EV routing and scheduling for resilience enhancement, considering the uncertainties associated with distributed renewables, load profiles, and traffic flow. Then, the coordination problem of each EV is formulated as a partially observable Markov decision process, and an attention-based distributed multi-agent deep deterministic policy gradient method, namely AD-MADDPG, is proposed to learn the optimal strategies. The proposed method applies an architecture with multi-actor, single-learner to reduce training complexity, employing a convolutional neural network to capture spatial characteristics from the CPTN, and incorporating a long short-term memory to derive temporal sequence features across multiple time steps, thereby enhancing the exploration efficiency of the action space. Simulation results implemented on the modified IEEE 33-bus test feeder demonstrate that AD-MADDPG outperforms all other baselines in terms of load restoration, restoration fairness, and energy consumption when varying different numbers of EVs, maximum discharging proportion, and maximum moving distance.

Keywords Resilience-driven, Microgrids, EV coordination, Coupled power-transportation network, Distributed deep reinforcement learning

Background and motivation

The continuous warming of the global climate poses significant challenges to the stable operation of power systems, increasing the occurrence of high-risk, low-probability extreme events[1–3]. To alleviate the impacts stemming from extreme events, enhancing the resilience of power systems has become a recent research focus[4,5]. Multi-microgrids (MMGs) architecture, comprising distributed renewable energy sources (DRES), demand response strategies, and flexible load resources, has emerged as a highly promising solution to enhance the resilience of distribution systems[6]. In the event of regional faults or outages in the utility grid, MMGs can operate in an off-grid mode, curtailing a portion of the load to maintain voltage and frequency within acceptable ranges, coordinating DRES to restore load as much as possible[7]. However, relying solely on the reconfiguration of distribution networks and intermittent DRES to maintain loads is challenging in practical operation, as their load restoration capabilities are limited and unstable. Moreover, multiple isolated microgrids may not be able to support each other, due to the random damage of distribution network lines.

Electric vehicles (EVs), with their excellent mobility and flexible charging and discharging capabilities, have been utilized in a lot of recent research to enhance the resilience of distribution networks[8]. The core issues of using EVs in resilience enhancement of MMGs lie in optimizing the routing and charge-discharge scheduling of EVs, maximizing load restoration while reducing energy consumption, under the premise of satisfying operational constraints[9]. To solve the optimal routing and dispatching problems, many model-

[1]School of Computer Engineering, Hubei University of Arts and Science, Xiangyang 441053, China. [2]School of Computer Science, Hubei University of Technology, Wuhan 430068, China. ✉email: caiting@hbut.edu.cn

based offline one-shot optimization methods have been proposed, such as mixed integer linear programming (MILP)[10], stochastic programming[11], and robust optimization[12], etc. However, comprehensively considering the uncertainties of DRES and flexible loads imposes significant computational burdens, which may reduce the methods' responsiveness to dynamic load changes.

To improve the adaptability of methods to time-varying environments, *model-free* approaches have been developed, such as deep reinforcement learning (DRL)[13], offering a solution method to the decision-making problems in resilience enhancement of MMGs. DRL's capability to deliver fast responses and accommodate uncertainties and contingencies is particularly advantageous. By training agents to iteratively interact with the environment without prior knowledge and learning the optimal policies for the dynamic scheduling of EVs. However, the existing DRL-based approaches[14–16] usually treat the routing and scheduling of EVs separately, without considering their temporal and spatial coordination. Moreover, they simplify the routing of EVs to discrete path choices on a 2D vector map, ignoring the temporal and spatial characteristics of the movement process.

To solve these research gaps, we aim to design a coordinated and efficient routing and dispatch strategy for EVs to contribute to the resilience enhancement of MMGs, extract both temporal and spatial features, and improve the approaches' responsiveness under incomplete information.

Literature review

Model-based methods

In recent years, *model-based* mathematical approaches have been extensively developed to model the routing and scheduling of EVs to enhance the resilience of power networks. In works[17–19], a mixed-integer linear programming (MILP) model is proposed to achieve coordinated operation among EVs and MGs. In work[20], a mixed-integer quadratically constrained programming (MIQCP) model is proposed to optimize the rerouting and dispatching of EVs for resilience enhancement in coupled traffic-electric networks. In work[21], a two-stage stochastic optimization scheduling framework is proposed to coordinate the scheduling of EVs for the restoration of interconnected power-transportation systems under natural hazard risks. In works[22,23], a robust optimization method is proposed to schedule the EVs' mobility and its charging and discharging strategies to enhance the resilience of coupled networks by minimizing both investment and operational costs under uncertain traffic demands. In work[24], a particle swarm optimization (PSO) algorithm is proposed to optimize the load dispatch of the microgrid containing EVs. In work[25], a hybrid genetic algorithm and simulated annealing method (GA-SAA) is used to optimize the placement of EV charging station, reducing power losses and maintaining acceptable voltage levels. In work[26], four metaheuristic algorithms, including differential evolution (DE), PSO, whale optimization algorithm (WOA), and grey wolf optimizer (GWO), are applied to schedule the charging/discharging activities of EVs, reducing the daily costs. However, when the number of considered scenarios is large, addressing uncertainties through stochastic programming or robust optimization can impose a significant computational burden. Furthermore, problem-solving based on heuristic algorithms cannot guarantee the accuracy of the solutions obtained.

Value-based DRL methods

Considering these limitations, DRL emerges as a *data-driven* and *model-free* approach[27], offering a new paradigm for resilient operations involving MMGs and EVs. In one aspect of the existing literature, the *value-based* DRL methods, have been adopted to train the optimal routing and scheduling decisions of EVs. In references[28,29], a deep Q-network (DQN) based model is proposed to dispatch a set of EVs to supply energy to different consumers at different locations, enhancing power system resilience while considering the uncertainties in power supply and demand. In references[15,30], a double deep Q-networks (DDQN)-based method is proposed to solve the routing and scheduling coordination problem of mobile energy storage systems in the load restoration. In reference[31], an enhanced dueling DDQN algorithm with mixed penalty function is developed to optimize the energy management of MG incorporating EVs. Although DQN or DDQN utilize deep neural networks to effectively handle high-dimensional state space and mitigate the curse of dimensionality, they still face significant challenges in balancing exploration and exploitation in such large spaces. Furthermore, their capability to handle continuous action spaces is limited, rendering them inadequate for addressing stochastic decision problems.

Policy-based DRL methods

The other side, therefore, employs *policy-based* DRL methods that can directly optimize the probability of taking an action or the action value rather than estimating the Q-value function. In reference[32], a decentralized Actor-Critic (AC) method is proposed to solve the routing and scheduling of a large fleet of EVs, addressing the scalability issues in large-scale smart grid systems. In reference[33], a heterogeneous multi-agent hypergraph attention Actor-Critic (HMA-HGAAC) framework is proposed to solve the joint EVs routing and battery charging scheduling problem in a transportation network with multiple battery swapping stations. In references[34,35], a deep deterministic policy gradient (DDPG) method is proposed to enhance the resilience of EV charging stations in the presence of cyber uncertainties. In reference[36], a twin-delayed deep deterministic policy gradient (TD3) is developed to obtain the optimal driving strategy of EVs in a traffic scenario with multiple constraints. In reference[37], a hybrid parameter sharing proximal policy optimization (H-PSPPO) method is proposed to compute both discrete and continuous actions simultaneously that align with the nature of EV routing and scheduling problems in power-transportation network, aiming at enhancing power system resilience.

It can be concluded that, despite various types of DRL methods being applied to the routing and scheduling of EVs participating in the resilient operation of MMGs, the previous work exhibits the following potential limitations: i) existing studies typically treat EVs' routing as a discrete action problem, without considering the extraction of spatial-temporal features of EVs' movement within the transportation network; ii) the methods mentioned above may not deliver timely services for load restoration due to the increasing system scale and complexity in training within a multi-agent setting; iii) the inherent instability of the environment, coupled with the information asymmetry among multiple agents, exacerbates the difficulty of achieving stable learning dynamics and converging to optimal strategies.

Contribution

To address the above limitations and achieve multi-objective optimization, we propose an attention-based distributed MADRL algorithm with a multi-actors and single-learner framework to conduct the routing and scheduling of EVs. The main contributions of this paper can be summarized as follows:

- A coupled power-transportation network (CPTN) is constructed to investigate the coordination effect of EVs routing and scheduling problems for the resilient load restoration of MMGs. Each EV routing and scheduling coordination problem is then formulated as a partially observable Markov decision process, exploiting the EVs' flexibility in temporal and spatial.
- An attention-based distributed multi-agent deep deterministic policy gradient method, namely AD-MADDPG, is proposed to solve the routing and scheduling coordination decision-making problem of EVs. This method employs a multi-actor, single-learner interaction architecture and utilizes a convolutional neural network (CNN) to extract spatial state features from the CPTN.
- To enhance the efficiency of action exploration and exploitation, a long short-term memory (LSTM) network is employed to capture multi-step temporal characteristics. By incorporating a prioritized experience replay mechanism, the multi-step reward is calculated, giving a more precise approximation.
- Comparative simulation results with other benchmark algorithms demonstrate that the proposed AD-MADDPG achieves superior performance in terms of load restoration, energy consumption, and restoration fairness.

Paper organization

The remainder of this paper is organized as follows. The section II presents the spatial-temporal network model of CPTN. The section III describes the problem formulation about the optimal routing and scheduling of EVs. In the section IV, the proposed AD-MADDPG algorithm is introduced to improve the load restoration performance. Performance evaluations are carried out in Section V to demonstrate the effectiveness of AD-MADDPG. Finally, the section VI concludes the paper.

Spatial-temporal network modeling of coupled power-transportation network

The constructed spatial-temporal network model for the coupled power-transportation network (CPTN) is illustrated in Fig. 1. Considering disconnection from the utility grid, the radial network of MMGs operating in off-grid mode can be represented as a tree-topological graph $\boldsymbol{G}_{mg} = (\boldsymbol{B}, \boldsymbol{L})$, where $\boldsymbol{B} = \{1, 2, ..., B\}$ denotes the set of buses, and $\boldsymbol{L} = \{1, 2, ..., L\}$ indicates the set of branches. Similarly, the transportation network is modeled as a directed connected graph $\boldsymbol{G}_{tn} = (\boldsymbol{V}_{tn}, \boldsymbol{E}_{tn})$, where $\boldsymbol{V}_{tn} = \{1, 2, ..., V\}$ represents the sequentially numbered vertices and $\boldsymbol{E}_{tn} = \{1, 2, ..., E\}$ denotes the edges of the connected graph, indicating road intersections (including origin and destination points) and directed road segments, respectively. The CPTN

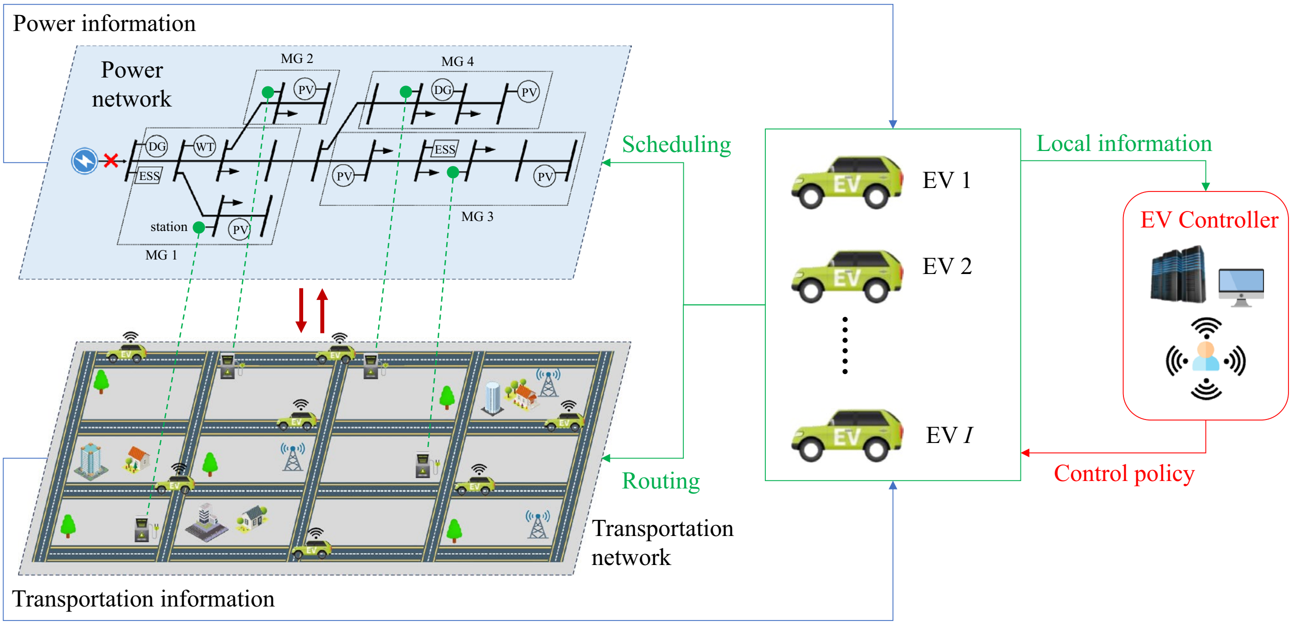


Fig. 1. The optimal routing and scheduling of multiple EVs in a coupled power-transportation network model.

model also incorporates charging-discharging stations (CDSs) that couple MMGs and transportation networks together, defined as $\boldsymbol{H} = \{1, 2, ..., H\}$. Additionally, the set of microgrids is represented as $\boldsymbol{M} = \{1, 2, ..., M\}$.

During extreme events such as earthquakes, typhoons, prolonged icy weather, etc., which result in failures or power outages in the utility grid area, MMGs can operate in islanded mode, utilizing local DRES, diesel generation (DG) and energy storage system (ESS) to maintain power supply to critical loads. In this paper, we focus on the optimal routing and scheduling of EVs $\boldsymbol{I} = \{1, 2, ..., I\}$ in the CPTN to enhance the resilient load restoration of isolated MMGs. Each MG monitors and collects information on bus and branch failures in the network, as well as the local generation and load, to form the supply-demand balance deviation requirements for the islanded operation of MGs. Additionally, each EV moves between different CDSs of MGs, considering the real-time traffic conditions of the transportation network, to facilitate load restoration after extreme events and ensure the stable operation of MMGs.

MMGs network modeling

After extreme events occur (e.g., line faults at B1-B2, B4-B5, B5-B6, B11-B12 in Fig. 6), the distribution network is segmented into multiple autonomous MGs. Each MG operates independently by leveraging local resources (DRES, DG, ESS) and internal grid reconfiguration capabilities[38]. MMGs typically shed a portion of the load as an emergency response to prevent sustained frequency and voltage decline when operating in islanded mode[38]. Subsequently, based on the locally available renewable energy generation capacity and dispatchable resources on the load side, MMGs gradually restore critical loads during the blackout period. We assume no internal bus-level damage within MGs, focusing on post-segmentation coordination. This simplification aligns with standard test systems (e.g., IEEE 33-bus) for resilience studies[3,6]. In this paper, the load restoration problem aims to maximize load recovery while minimizing economic costs, subject to safety constraints.

Operation costs of MG

The operation costs of each MG consist of three components: (1) DG generation cost, (2) load shedding cost, and (3) ESS battery degradation cost, as expressed in Eq. (1).

$$C_m^{mg} = \sum_{t=1}^{T} \left[\alpha_m P_{m,t}^{dg} + c_m^{ls} P_{m,t}^{ls} + c_m^{ess} P_{m,t}^{essd} \right] \tag{1}$$

where α_m and $P_{m,t}^{dg}$ represent the unit generation cost and active power output of DG, respectively. c_m^{ls} and $P_{m,t}^{ls}$ refer to the load shedding cost and the quantity of load shedding for MG *m*. c_m^{ess} and $P_{m,t}^{essd}$ represent the battery degradation cost coefficient of ESS and the amount of energy discharged. The objective of each MG is to minimize its operational cost while satisfying operational constraints. Additionally, to prioritize the load restoration of islanded MG, we set $c_m^{ls} \gg \alpha_m$ and $c_m^{ls} \gg c_m^{ess}$, indicating that the penalty cost for load shedding is higher than the DG generation cost and ESS battery degradation cost.

Operation constraints

System-level constraints: The secure operation of MGs requires compliance with corresponding system-level power flow constraints as well as internal component-level constraints within the MGs. In this paper, the linearized DistFlow model[39] is adopted to describe the power flow and voltage of MGs.

$$p_{nj,t} = \sum_{k \in \boldsymbol{B}_j} p_{jk,t} + P_{j,t}^{load} - P_{j,t}^{dg} - P_{j,t}^{pv} - P_{j,t}^{wt} - P_{j,t}^{essd} \tag{2}$$

$$q_{nj,t} = \sum_{k \in \boldsymbol{B}_j} q_{jk,t} + Q_{j,t}^{load} - Q_{j,t}^{dg} \tag{3}$$

$$v_{j,t} = v_{n,t} - \frac{(r_{nj} p_{nj,t} + w_{nj} q_{nj,t})}{v_1} \tag{4}$$

$$v_{n,\min} \leqslant v_{n,t} \leqslant v_{n,\max} \tag{5}$$

where Eqs. (2) and (3) represent the active power and reactive power from bus *n* to *j* at time step *t*, respectively. $\boldsymbol{B}_j$ is the set of buses that take bus *j* as the parent node. Equations (4) and (5) represent the voltage of bus and related constraints. Here, r_{nj} and w_{nj} represent the resistance and reactance of the branch (bus *n* to *j*), where the branch $(n, j) \in \boldsymbol{L}$.

Component-level constraints: The operational constraints of each component (including DG, ESS, PV, WT) within the MGs are represented as follows:

$$P_{m,\min}^{dg} \leqslant P_{m,t}^{dg} \leqslant P_{m,\max}^{dg}, \forall m, t \tag{6}$$

$$Q_{m,\min}^{dg} \leq Q_{m,t}^{dg} \leq Q_{m,\max}^{dg} \tag{7}$$

$$P_{m,\min}^{essd} \leq P_{m,t}^{essd} \leq P_{m,\max}^{essd}, \forall m, t \tag{8}$$

$$E^{ess}_{m,t+1} = E^{ess}_{m,b,t} + \eta^{\text{ch}}_m P^{essc}_{m,t} \cdot \Delta t - \frac{\Delta t \cdot P^{essd}_{m,t}}{\eta^{\text{dis}}_m}, \forall t \tag{9}$$

$$E^{ess}_{m,\min} \leqslant E^{ess}_{m,t} \leqslant E^{ess}_{m,\max} \tag{10}$$

$$0 \leqslant P^{wt}_{m,t} \leqslant P^{wt}_{m,\max}, \forall m, t \tag{11}$$

$$0 \leqslant P^{pv}_{m,t} \leqslant P^{pv}_{m,\max}, \forall m, t \tag{12}$$

where Eqs. (6) and (7) represent the active power and reactive power constraints of DG, while (8) and (9) represent the power constraints of ESS.

Transportation network modeling

The transportation network model is illustrated in Fig. 2. Each node $e, g \in \boldsymbol{V}_{tn}$ represents a road intersection, which could also serve as the original or destination point of a journey. An edge $(e, g) \in \boldsymbol{E}_{tn}$ denotes an available road connecting nodes e and g. For each road, the length $D(e, g) \in \mathbb{R}^{\geq 0}$ determines the distance traveled along the road, while $t_{rd}(e, g) \in \{1, 2, \ldots, T\}$ represents the travel time on the road without congestion. The maximum capacity of the road, $V^{\max}_{(e,g)} \in \mathbb{R}^{\geq 0}$, indicates the maximum number of vehicles (per unit time) the road can accommodate without causing congestion from external traffic.

EVs traveling through the transportation network can both charge and discharge at the CDSs of MGs. When there is an excess of renewable energy generation in isolated MGs, EVs can serve as charging loads to absorb the surplus energy. Conversely, when there is insufficient generation in isolated MGs leading to load shedding, EVs can discharge, thereby enhancing the operational resilience of MGs. $N^{\max}_{h \in \boldsymbol{H}} \in \mathbb{N}$ denotes the maximum number of EVs that can charge or discharge simultaneously at CDS h. In this paper, EVs freely enter/exit the network. Our model optimizes participating EVs (fixed during scheduling horizons), consistent with real-world V2G contracts.

Real-time traffic flow

Due to the spatial-temporal dynamics of traffic flow in the transportation network, the travel time of road (e, g) at time t is significantly influenced by the real-time traffic volume. This paper models the impact of real-time traffic volume on road travel time as follows[15]:

$$T^{dri}_{(e,g),t} = t^{rd}_{(e,g)} + t^{rd}_{(e,g)} \cdot \delta^{rd} \left(\frac{F^{rt}_{(e,g),t}}{F^{\max}_{(e,g)}} \right)^{\rho^{rd}}, \forall t \tag{13}$$

$$F^{rt}_{(e,g),t} = d^{rd}_{(e,g),t} + \sum_{i \in \boldsymbol{I}} N^{rd}_{i,(e,g),t}, \forall t \tag{14}$$

where $T^{dri}_{(e,g),t}$ represents the real-time travel time of the road (e, g), $F^{rt}_{(e,g),t}$ denotes the real-time traffic flow, δ^{rd} and ρ^{rd} represent congestion factors. $d^{rd}_{(e,g),t}$ indicates the basic traffic volume[40], i.e., the other types of vehicles with specific daily travel patterns in the transportation network. $\sum_{i \in \boldsymbol{I}} N^{rd}_{i,(e,g),t}$ represents the number of EVs participating in the resilient operation of islanded MGs on the road.

Dispatching costs of EVs

As EVs move within the transportation network, they consume energy, and their travel time is also affected by real-time road congestion. Therefore, we consider the energy consumption cost, time cost, and battery degradation cost of EVs moving between different CDSs for charging and discharging, calculated as follows:

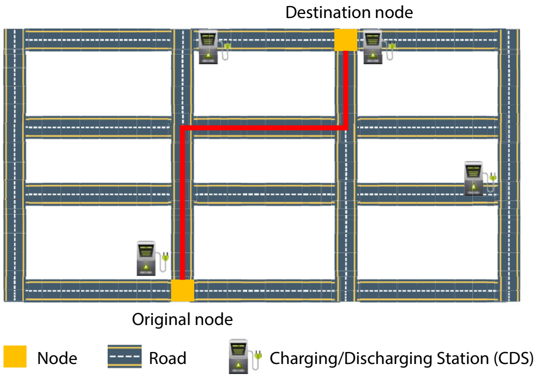


Fig. 2. The transportation network integrating charging/discharging stations.

$$C_i^{ev} = \sum_{t \in \boldsymbol{T}} \left(d_{bat}^{ev} \cdot \left(P_{i,t}^{evd} + P_{i,t}^{evc} \right) + \kappa \cdot D_{i,t}^{ev} + \sigma \cdot T_{i,t}^{dri} \right) \tag{15}$$

where d_{bat}^{ev} represents the battery depreciation cost per unit of charging/discharging for EVs, $P_{i,t}^{evd}/P_{i,t}^{evc}$ indicates the discharging/charging power of EV *i* at time *t*. κ and σ represent the energy consumption cost per unit distance and the cost coefficient per unit time for EVs, respectively. $D_{i,t}^{ev}$ and $T_{i,t}^{dri}$ represent the moving distance and traveling time.

Moreover, the constraints of routing and scheduling behaviors for EVs are represented as follows:

$$0 \leqslant P_{i,t}^{evc} \leqslant \mu_{i,t}^{c} \cdot P_{i,\max}^{evc}, \forall i \in \boldsymbol{I} \tag{16}$$

$$0 \leqslant P_{i,t}^{evd} \leqslant \mu_{i,t}^{d} \cdot P_{i,\max}^{evd}, \forall i \in \boldsymbol{I} \tag{17}$$

$$\mu_{i,t}^{c} + \mu_{i,t}^{d} \leqslant 1 \tag{18}$$

$$E_{i,t}^{ev} = E_{i,t-1}^{ev} + \eta_i^{evc} P_{i,t}^{evc} \Delta t - \frac{P_{i,t}^{evd} \Delta t}{\eta_i^{evd}} \tag{19}$$

$$E_{i,\min}^{ev} \leqslant E_{i,t}^{ev} \leqslant E_{i,\max}^{ev} \tag{20}$$

where Eqs. (16) and (17) represent the charging and discharging constraints of EVs. The binary variables $\mu_{i,t}^{c}$ and $\mu_{i,t}^{d}$ represent the charging and discharging decisions of EV. Equation (18) guarantees that the charging and discharging patterns of EV *i* cannot be triggered simultaneously. Equations (19) and (20) indicate the energy storage dynamic and the minimum and maximum energy storage levels of EVs.

Problem formulation

The key issue of this paper is to optimize the routing and charging-discharging scheduling of EVs in the CPTN. The goal is to achieve maximum load restoration in MMGs with the minimum energy consumption cost of the EVs.

POMDP modeling

To address the resilient load restoration problem of MMGs, the optimal routing and scheduling of EVs is formulated as a partially observable Markov decision process (POMDP[33]), denoted as $\langle \boldsymbol{S}, \boldsymbol{O}, \boldsymbol{A}, \boldsymbol{R}, \Upsilon \rangle$. Then a distributed MADRL-based method is proposed to solve the optimization problem.

State space

The state space of the CPTN system is defined as $\boldsymbol{S}_t = \{\boldsymbol{S}_1, \boldsymbol{S}_2, \boldsymbol{S}_3\}$, comprising three parts:

(1) The first channel of state $\boldsymbol{S}_1$ is shown in Fig. 3a, including the positions $(x(h), y(h))_{h \in \boldsymbol{H}}$ of the CDSs and the quantity of shedding load $P_{h,t}^{ls}$, i.e., $\boldsymbol{S}_1 = \left\{ \left(x(h), y(h), P_{h,t}^{ls} \right) \right\}_{h \in \boldsymbol{H}, t \in \boldsymbol{T}}$. Here, we utilize the index of CDSs to represent their corresponding MGs. Additionally, when $P_{h,t}^{ls}$ is positive, it indicates that the MG corresponding to CDS *h* experiences load shedding; when $P_{h,t}^{ls}$is negative, it signifies that the MG corresponding to CDS *h* has surplus electricity.

(2) The second channel of state $\boldsymbol{S}_2$ is shown in Fig. 3b, including the positions $(x(e), y(e))_{e \in \boldsymbol{V}}$ of nodes, and the traffic volume $F_{(e,g),t}^{rt}$ of road (*e*, *g*), i.e., $\boldsymbol{S}_2 = \left\{ \left(x(e), y(e), F_{(e,g),t}^{rt} \right) \right\}_{e \in \boldsymbol{V}, t \in \boldsymbol{T}}$.

(3) The third channel of state $\boldsymbol{S}_3$ is shown in Fig. 3c, which includes the positions $(x(i), y(i))_{i \in \boldsymbol{I}}$ and remaining energy $E_{i,t}^{ev}$ of all EVs, i.e., $\boldsymbol{S}_3 = \left\{ \left(x(i), y(i), E_{i,t}^{ev} \right) \right\}_{t \in \boldsymbol{T}}$.

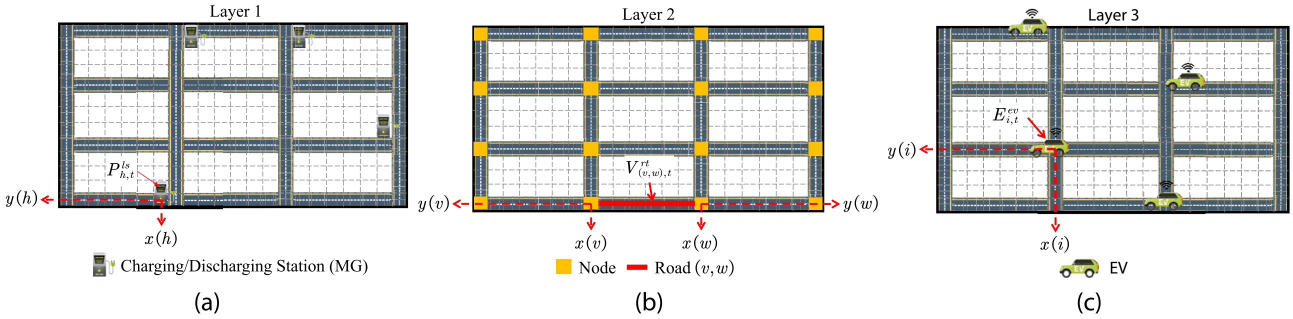


Fig. 3. The input state of the coupled power-transportation network.

Action space

The action space is defined as $\boldsymbol{A} \triangleq \left\{a_{i,t} = \left(\theta_{i,t}, l_{i,t}, P_{i,t}^{sch}\right)_{i \in \boldsymbol{I}}\right\}$, where $\theta_{i,t}$ and $l_{i,t}$ represent the moving direction and distance of EV *i* at time step *t*, respectively, $P_{i,t}^{sch} \in [-1,1]$ denotes the charging/discharging scheduling of EVs. At each time slot *t*, each EV decides its direction $\theta_{i,t} \in [0, 2\pi)$ and distance $l_{i,t} \in [0, l_{\max}]$ of movement based on the load reduction of each MG and the real-time traffic conditions, aiming to achieve the maximum load restoration with minimum movement cost. Additionally, EVs must comply with driving regulations and travel along the roads, incurring penalties if the direction and distance of movement exceed the boundaries (coordinates) of the roads. The charging/discharging schedule $P_{i,t}^{sch} \in [-1,1]$ represents the percentage of charging (positive) or discharging (negative) power of EVs relative to its battery capacity $[-P_{i,\max}^{evd}, P_{i,\max}^{evc}]$ when it arrives at a specific CDS. It is important to note that, unlike existing work, in this paper, the direction and distance of the EV's movement are treated as continuous variables.

Reward function

The objective of this paper is to minimize the load reduction costs of isolated MGs in an energy-efficient way. Then, combining Eqs. (1) and (15), the reward function can be defined as follows:

$$r_{i,t} = \frac{\omega_t \cdot \left(-c_m^{ls} P_{m,t}^{ls}\right)}{C_{i,t}^{ev}\left(P_{i,t}^{evd}\right)} + \frac{\omega_t \cdot P_{i,t}^{evc}}{C_{i,t}^{ev}\left(P_{i,t}^{evc}\right)} \tag{21}$$

where ω_t denotes the fairness of load restoration in the MG, $c_m^{ls} P_{m,t}^{ls}$ represents the load shedding cost of MG, $C_{i,t}^{ev}\left(P_{i,t}^{evd}\right)$ and $C_{i,t}^{ev}\left(P_{i,t}^{evc}\right)$ are the discharging and charging cost of EV *i*, respectively. Rewards incentivize load restoration ($-c_m^{ls} P_{m,t}^{ls}$) and efficient EV usage. Traffic congestion costs are inherently captured by $T_{i,t}^{dri}$ in C_i^{ev} (Eq. 15).

Load restoration fairness

However, driven by the rational pursuit of profit, an EV is inclined to discharge at nearby MGs and then recharge at the closest ones, potentially neglecting MGs located in more remote areas. This tendency may result in insufficient or even no load restoration for certain MGs. Therefore, we introduce a fairness index ω_t for load restoration across all isolated MGs, which can be represented by the Jain's fairness index[41], defined as:

$$\omega_t(\pi) = \frac{\left(\sum_{m=1}^{M} P_t^{load}(\pi, m)\right)^2}{M \sum_{m=1}^{M} P_t^{load}(\pi, m)^2} \tag{22}$$

where $P_t^{load}(\pi, m)$ represent the load restoration amount of MG *m* under EV scheduling strategy π. It can be seen that as the value of $\omega_t(\pi) \in [\frac{1}{M}, 1]$ increases, the load restoration process among different MGs becomes more equitable.

According to the predetermined state transition Υ and reward function $r_{i,t}$, the optimization problem in this paper can be formulated as:

$$V_i^*(s_t) = \max_{a_{i,t}} \left[r_{i,t}(s_t, a_{i,t}) + \gamma \int_{s_{t+1} \in \boldsymbol{S}} \Upsilon(s_t, a_{i,t}, s_{t+1}) V_i^*(s_{t+1}) \right] \tag{23}$$

where γ denotes the discount factor. At time slot *t*, each EV *i* makes decisions to learn the optimal scheduling policy π_i^*, aiming to maximize its cumulative total reward, which can be defined as follows:

$$\pi_i^* = \arg\max_{a_{i,t}} \left[r_{i,t}(s_t, a_{i,t}) + \gamma \int_{s_{t+1} \in \boldsymbol{S}} \Upsilon(s_t, a_{i,t}, s_{t+1}) V_i^*(s_{t+1}) \right] \tag{24}$$

To solve the continuous optimal routing and scheduling problem of EV *i*, we proposed a distributed multi-agent DRL (MADRL) method to learn the optimal policy π_i^* for each EV. However, the optimal routing decision of an EV involves the spatial features of the CPTN. Additionally, due to the limited maximum distance an EV can move within a single time slot *t*, reaching a CDS may require multiple steps. Therefore, exploiting both the spatial and temporal features of EV movements is critical to solving our problem. In this paper, we use a convolutional neural network (CNN) to capture the spatial characteristics and integrate a long short-term memory (LSTM) network to extract temporal features, thereby improving the long-term performance of our model.

Distributed deep deterministic policy gradient framework

In this section, we propose an attention-based distributed multi-agent deep deterministic policy gradient (AD-MADDPG) framework to address the continuous routing and scheduling problem for each EV.

Attention-based distributed actor-critic with spatial state modeling

Each EV aims to navigate through the transportation network to reach the corresponding CDS associated with each isolated MG, minimizing load shedding losses for the MG during disconnection periods. Simultaneously, to reduce its energy consumption, EV *i* needs to optimize both its charging/discharging decisions and its movement paths within the transportation network. Therefore, acquiring effective spatial state information about the CPTN is crucial for the EV's decision-making. To facilitate the extraction of spatial features from the CPTN, we divide the transportation network into a grid graph and utilize a CNN to extract the spatial information, as illustrated in Fig. 4.

Furthermore, for each EV *i*, four deep neural networks (DNNs) are implemented as Actor network θ_i^a, Target Actor network θ_i^{ta}, Critic network θ_i^c and Target Critic network θ_i^{tc}. At each time slot *t*, EV *i* obtains an observation $o_{i,t}$ and then the Actor θ_i^a decides on an action $a_{i,t}$ to execute according to the policy $\pi_i\left(a_{i,t}\,|o_{i,t}\right)$ combined with an attention mechanism. Here the attention mechanism learns a weight distribution W_i over the input observation and applies it to the original features, enabling the learning task to focus on the most important features, thereby improving efficiency. The attention mechanism is calculated as follows:

$$o_{i,t}' = \mathrm{Softmax}\left(W_i \cdot o_{i,t}\right) \cdot o_{i,t} \tag{25}$$

The attention mechanism can adaptively adjust the attention weights W_i and feature weighting based on changes in the environment and tasks, thereby optimizing policy selection and improving reinforcement learning performance.

After each EV takes actions according to its private observation, the corresponding reward $r_{i,t}$ is calculated for each EV, and the state s_t transitions to s_{t+1}. This transition includes the amount of load shedding $P_{m,t}^{ls}$, the traffic volume $F_{(e,g),t}^{rt}$ of road (*e*, *g*), the EV location $\left(x\left(i\right), y\left(i\right)\right)_{i\in\boldsymbol{I}}$, and the remaining energy $E_{i,t}^{ev}$ of each EV. At the end of each time slot *t*, the state transition tuple $\left(o_{i,t}, a_{i,t}, r_{i,t}, o_{i,t+1}\right)$ generated from the interaction between EV and the environment is stored into the experience replay memory. After a fixed number of interactions, a mini-batch of state transitions is sampled from the experience replay memory to train the DNNs. Then, the critic network θ_i^c is updated by minimizing the following loss function:

$$L\left(\theta_i^c\right) = \mathbb{E}\left[\left(y_{i,t}^{\theta_i^{tc}} - Q\left(s_t, a_{1,t}, \ldots, a_{I,t}\,|\theta_i^c\right)\right)^2\right] \tag{26}$$

where the target *Q*-value $y_{i,t}^{\theta_i^{tc}}$ is calculated by:

$$y_{i,t}^{\theta_i^{tc}} = r_{i,t} + \gamma Q_i'\left(s_{t+1}, a_{1,t+1}, \ldots, a_{I,t+1}\,\middle|\theta_i^{tc}\right) \tag{27}$$

Finally, the actor network θ_i^a is updated using gradients as:

$$\nabla_{\theta_i^a} J \approx \nabla_a Q_i(s, a_{1,t}, \ldots, a_{I,t}|\theta_i^c)_{a_{i,t}=\pi_i(o_{i,t})} \cdot \nabla_{\theta_i^a}\theta_i^a(o_i) \tag{28}$$

Given the scale of the CPTN network, EVs require significant time and numerous transitions to fully explore their spatial and temporal features. To reduce the training complexity and enable the agent to learn optimal routing

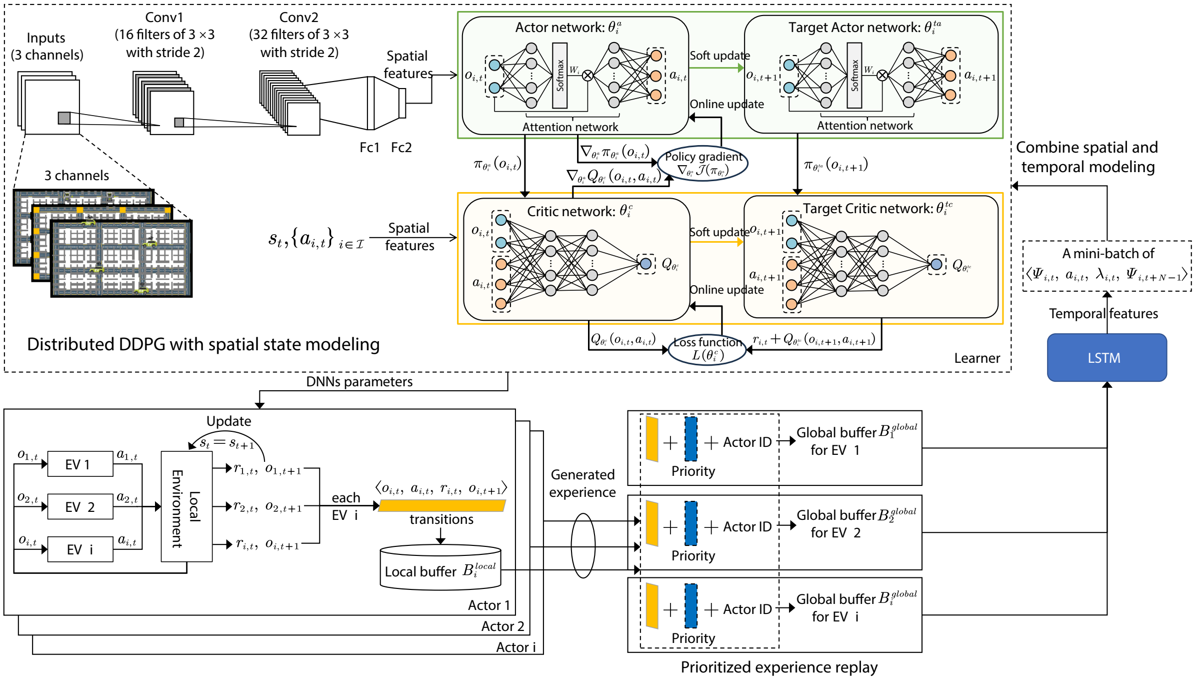

Inputs
(3 channels)
Conv1
(16 filters of 3 ×3
with stride 2)
Conv2
(32 filters of 3 ×3
with stride 2)
Spatial
features
Fc1
Fc2
3 channels
$s_t, \{a_{i,t}\}_{i\in\mathcal{I}}$
Spatial
features
Distributed DDPG with spatial state modeling
Actor network: θ_i^a
Softmax
W_i
$o_{i,t}$
$a_{i,t}$
Attention network
Soft update
Online update
Target Actor network: θ_i^{ta}
$o_{i,t+1}$
$a_{i,t+1}$
Attention network
$\pi_{\theta_i^a}(o_{i,t})$
$\nabla_{\theta_i^a}\pi_{\theta_i^a}(o_{i,t})$
Policy gradient
$\nabla_{\theta_i^a}\mathcal{J}(\pi_{\theta_i^a})$
$\nabla_a Q_{\theta_i^c}(o_{i,t}, a_{i,t})$
$\pi_{\theta_i^{ta}}(o_{i,t+1})$
Critic network: θ_i^c
$Q_{\theta_i^c}$
Target Critic network: θ_i^{tc}
$Q_{\theta_i^{tc}}$
$Q_{\theta_i^c}(o_{i,t}, a_{i,t})$
Loss function
$L(\theta_i^c)$
$r_{i,t} + Q_{\theta_i^{tc}}(o_{i,t+1}, a_{i,t+1})$
Learner
Combine spatial and
temporal modeling
A mini-batch of
$\langle \Psi_{i,t},\ a_{i,t},\ \lambda_{i,t},\ \Psi_{i,t+N-1} \rangle$
Temporal features
LSTM
DNNs parameters
Update
$s_t = s_{t+1}$
$o_{1,t}$
EV 1
$a_{1,t}$
$o_{2,t}$
EV 2
$a_{2,t}$
$o_{i,t}$
EV i
$a_{i,t}$
Local
Environment
$r_{1,t},\ o_{1,t+1}$
$r_{2,t},\ o_{2,t+1}$
$r_{i,t},\ o_{i,t+1}$
each
EV i
$\langle o_{i,t},\ a_{i,t},\ r_{i,t},\ o_{i,t+1} \rangle$
transitions
Local buffer B_i^{local}
Actor 1
Actor 2
Actor i
Generated
experience
+ Actor ID
Priority
Global buffer B_1^{global}
for EV 1
Global buffer B_2^{global}
for EV 2
Global buffer B_i^{global}
for EV i
Prioritized experience replay

Fig. 4. Distributed DDPG with spatial information and temporal sequence modeling.

and scheduling strategies efficiently, we adopt a multi-actor, single-learner training framework. Specifically, the learner is implemented using the DDPG method, which consists of four DNNs (see Fig. 4). Each independent actor (EV) copies the policy network parameters from the learner and updates its policy network periodically. Each actor maintains a local experience replay memory B_i^{local} to store the generated transitions sequentially within its local observation, while the learner utilizes a global memory B_i^{global} for each EV i. The local memory B_i^{local} sends all data to the global memory B_i^{global} once B_i^{local} is full. Since the multi-actors and single-learner operate in parallel, this framework allows for distributed execution of exploration and learning tasks, improving efficiency.

LSTM-enhanced temporal sequence features extraction

Within a single time slot *t*, the immediate step reward $r_{i,t}$ only reflects benefits or losses at that moment and cannot capture the cumulative rewards of the EV over multiple time slots, which is essential for long-distance travel (since reaching a CDS may require multiple steps due to the maximum distance constraint within a single time slot *t*). Therefore, relying solely on the value of previous observations and actions based on $Q_i\left(o_{t+1}, a_{i,t+1}\right)$ may not be accurate during the initial stages of training. Additionally, some EVs may develop a tendency to rely on specific CDSs that are suboptimal or may frequently return to recharge without contributing to load restoration rewards. Hence, it is crucial to extract multi-step temporal sequence features to address this issue. To better capture the long-term impact of policy choices and help the agent evaluate the future value of its actions, this paper calculates the cumulative reward over multiple steps as follows:

$$\lambda_{i,t} = r_{i,t} + \gamma r_{i,t+1} + \cdots + \gamma^{N-1} r_{i,t+N-1}, \forall i \in \boldsymbol{I} \tag{29}$$

Therefore, the new transition will use $o_{i,t}$ and $o_{i,t+N-1}$ as the initial and final observations, generating a new transition $\left(o_{i,t}, a_{i,t}, \lambda_{i,t}, o_{i,t+N-1}\right)$.

To account for the multi-step reward for each EV *i*, with a transition starting from *t* to $t+N-1$, we employ an LSTM network to capture additional temporal sequence information. Specifically, we fed two sequences of observations $\left\{o_{i,t-\varepsilon+1}, o_{i,t-\varepsilon+2}, \cdots, o_{i,t}\right\}$ and $\left\{o_{i,t+N-\varepsilon}, o_{i,t+N-\varepsilon+1}, \cdots, o_{i,t+N-1}\right\}$ into the LSTM network to generate new observations $\varPsi_{i,t}$ and $\varPsi_{i,t+N-1}$ respectively, as shown in Fig. 5. This allows us to obtain a new transition $\left(\varPsi_{i,t}, a_{i,t}, \lambda_{i,t}, \varPsi_{i,t+N-1}\right)$ based on the LSTM network modeling. Here, ε is the LSTM sequence length. In this way, as an EV learns from mini-batches, it retrospectively considers multiple time slots, gaining insight into the cumulative impact of a sequence of actions and decisions. Additionally, the EV can simultaneously learn optimal charging or load restoration locations through this iterative process.

Action clipping for safe exploration

When exploring the action space using DDPG, the physical system constraints of the EVs and MMGs must be satisfied. Otherwise, the resulting actions may threaten the safe operation of the MMGs and lead to severe consequences. Therefore, we clip the agents' actions to ensure that the action exploration of EV agents complies with the constraints. In the load restoration problem of an isolated microgrid system, Eqs. (16)–(20) account for the constraints of EV charging/discharging scheduling. To satisfy these constraints, the Sigmoid activation function is designed in the output layer of the Actor network, guaranteeing that the output actions $\left[\theta_{out,i,t}, l_{out,i,t}, P_{out,i,t}^{sch}\right]$ are normalized values between 0 and 1. These normalized values are then mapped back to the absolute values within the operational range, expressed as follows:

$$\theta_{i,t} = 2\pi \cdot \theta_{out,i,t} \tag{30}$$

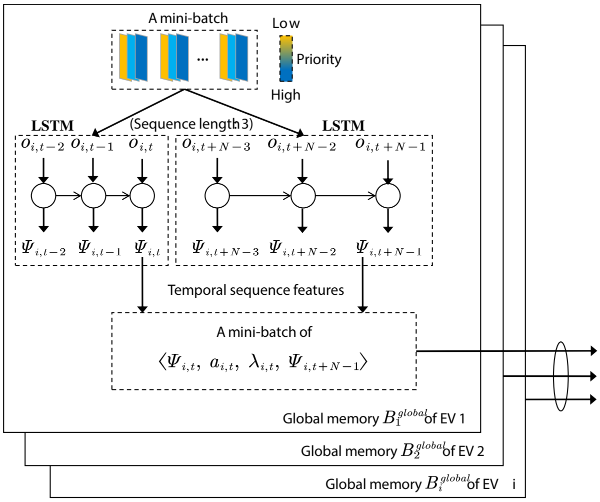

A mini-batch
Low
Priority
High
LSTM
(Sequence length 3)
LSTM
$o_{i,t-2}$ $o_{i,t-1}$ $o_{i,t}$
$o_{i,t+N-3}$ $o_{i,t+N-2}$ $o_{i,t+N-1}$
$\varPsi_{i,t-2}$ $\varPsi_{i,t-1}$ $\varPsi_{i,t}$
$\varPsi_{i,t+N-3}$ $\varPsi_{i,t+N-2}$ $\varPsi_{i,t+N-1}$
Temporal sequence features
A mini-batch of
$\langle\varPsi_{i,t}, a_{i,t}, \lambda_{i,t}, \varPsi_{i,t+N-1}\rangle$
Global memory B_1^{global} of EV 1
Global memory B_2^{global} of EV 2
Global memory B_i^{global} of EV i

Fig. 5. Prioritized experience replay integrating LSTM-enhanced multi-step temporal sequence features extraction.

$$l_{i,t} = l_{\max} \cdot l_{out,i,t} \tag{31}$$

$$P_{i,t}^{sch} = 2P_{out,i,t}^{sch} - 1 \tag{32}$$

$$E_{i,t}^{ev} = \begin{cases} \max\left(E_{i,t}^{ev} + \Delta t \cdot P_{i,t}^{sch} \cdot P_{i,\max}^{evd} / \eta_i^{evd}, E_{i,\min}^{ev}\right), & \text{if } P_{i,t}^{sch} \leqslant 0 \\ \min\left(E_{i,t}^{ev} + \Delta t \cdot \eta_i^{evc} P_{i,t}^{sch} \cdot P_{i,\max}^{evc}, E_{i,\max}^{ev}\right), & \text{Otherwise} \end{cases} \tag{33}$$

where Eqs. (30) and (31) denote the action constraints for the EV's movement direction and distance, respectively. Equation (32) represents the charging/discharging power constraint for the EV, while Eq. (33) ensures that the EV's energy level at time step *t* always remains within its lower and upper bound.

Algorithm description

In this subsection, the AD-MADDPG algorithm for load restoration in isolated MMGs is illustrated, which combines a CNN for spatial information extraction and integrates an LSTM to capture multi-step temporal sequence features. The single-learner, utilizing the Deep Deterministic Policy Gradient (DDPG), replays mini-batch of transitions to update the policy network. The multi-actors operate independently within the environment, evaluating a local policy π derived from the Learner, and recording the observed transition data into their local replay memory. Periodically and asynchronously, these actors transmit their local buffer data to the Learner for each EV *i*. The training process of AD-MADDPG is described in Algorithms 1 and 2.

```
Initialize the local environment state $s_t = s_0$;
Initialize the local policy network by duplicating the network parameters of learner;
Set the local replay memory $B_i^{local}$, the number of EV $I$ and time slot count $T$;
for each episode do
    for $t$ = 1, 2, ..., $T$ do
        for $i$ = 1, 2, ..., $I$ do
            Get the current observation $o_{i,t}$ from local environment;
            Select an action $a_{i,t} = \pi_{\theta_i^{a'}}(o_{i,t}) + \mathcal{N}_t$ using the current policy and exploration noise;
            Clipping the action $a_{i,t}$ according to Eqs. (30)-(33);
        Perform action $a_t = (a_{1,t}, a_{2,t}, \cdots, a_{i,t})$ in the local environment, calculate the reward $r_t = (r_{1,t}, r_{2,t}, \cdots, r_{i,t})$ and
          obtain the next state $s_{t+1}$;
        for $i$ = 1, 2, ..., $I$ do
            Get the reward $r_{i,t}$ and obtain the next observation $o_{i,t+1}$ according to state $s_{t+1}$;
            Compute the transition's priority $\delta_{i,t}^{\beta} = |y_{i,t}^{\beta} - Q_i(s_t, a_{1,t}, ..., a_{I,t})|$;
            Store transition $(o_{i,t}, a_{i,t}, r_{i,t}, o_{i,t+1}, \delta_{i,t}^{\beta}, e)$ into the local memory $B_i^{local}$, where $e$ is the ID of actor;
            if $B_i^{local}$ is full then
                Transfer all memory data of local $B_i^{local}$ to the global $B_i^{global}$;
            Update local policy network by copying the learner's latest network parameters;
        Set current state $s_t = s_{t+1}$;
```

Algorithm 1. Actors

```
Initialize parameters $\theta_i^a$, $\theta_i^c$ randomly;
Initialize parameters $\theta_i^{ta} \leftarrow \theta_i^a$, $\theta_i^{tc} \leftarrow \theta_i^c$;
Initialize LSTM sequence length $\varepsilon$, the size of multi-step $N$;
Initialize the global replay buffer $B_i^{global}$;
Set the number of MG $I$ and time slot count $T$;
for each episode do
    for $t$ = 1, 2, ..., $T$ do
        for $i$ = 1, 2, ..., $I$ do
            if $|B_i^{global}| \geqslant F$ then
                Sample a mini-batch of size $F$ from global memory $B_i^{global}$;
                for transition $(o_{i,\beta}, a_{i,\beta}, r_{i,\beta}, o_{i,\beta+1})$ in $F$ do
                    Calculate $N$-step reward $\lambda_{i,\beta}$ according to Eq. (29);
                    Get the two sequence of observation $(o_{i,\beta-\varepsilon+1}, o_{i,\beta-\varepsilon+2}, \cdots, o_{i,\beta})$ and
                    $(o_{i,\beta+N-\varepsilon}, o_{i,\beta+N-\varepsilon+1}, \cdots, o_{i,\beta+N-1})$ from $B_i^{global}$;
                    Obtain the new observation $\Psi_{i,\beta}$ and $\Psi_{i,\beta+N-1}$ by LSTM;
                    Replace $(o_{i,\beta}, a_{i,\beta}, r_{i,\beta}, o_{i,\beta+1})$ by $(\Psi_{i,\beta}, a_{i,\beta}, \lambda_{i,\beta}, \Psi_{i,\beta+N-1})$ ;
                Using a CNN to extract the spatial information of EV $i$;
                Calculate the target $Q$-value of sampled transitions as:
                $y_{i,\beta} = \lambda_{i,\beta} + \gamma Q_i'(s_{\beta+N-1}, a_{1,\beta+N-1}, \cdots, a_{I,\beta+N-1} | \theta_i^{tc})$ ;
                Update the Critic network $\theta_i^c$ according to Eq. (26);
                Update the Actor network $\theta_i^a$ according to Eq. (28);
                Soft update the Target Critic $\theta_i^{tc}$ and Target Actor $\theta_i^{ta}$ network;
                Update the priorities $\delta_i^{\beta} = \left| y_{i,t}^{\beta} - Q_i(s_t, a_{1,t}, ..., a_{I,t}) \right|$ of the $F$ transitions;
            if $B_i^{global}$ is full then
                Replace the earliest stored transition;
```

Algorithm 2. Learner

Performance evaluation
Simulation settings

To validate the performance of the proposed AD-MADDPG algorithm for resilient load restoration in EV-coordinated MMGs, we model the transportation network as a 2D square grid with dimensions of 32 × 32 units. For the power network, we construct an off-grid operation scenario consisting of 5 MGs, based on a modified IEEE 33-bus test system, with the topological structure shown in Fig. 6. The CPTN includes 5 CDSs, which support EVs for charging and discharging. We assume that at 10:00 a.m., multiple faults occur, causing the MMGs to disconnect from the utility grid. The expected duration of the utility grid outage is 6 hours, with an EV scheduling interval of 0.5 hours (i.e., $T = 12$). Additionally, to validate the performance of EVs in participating in the load restoration of MMGs, we assume that the following lines are disconnected due to faults: B1-B2, B4-B5, B5-B6, and B11-B12. As a result, the MMGs are decomposed into 5 autonomously operating MGs, each

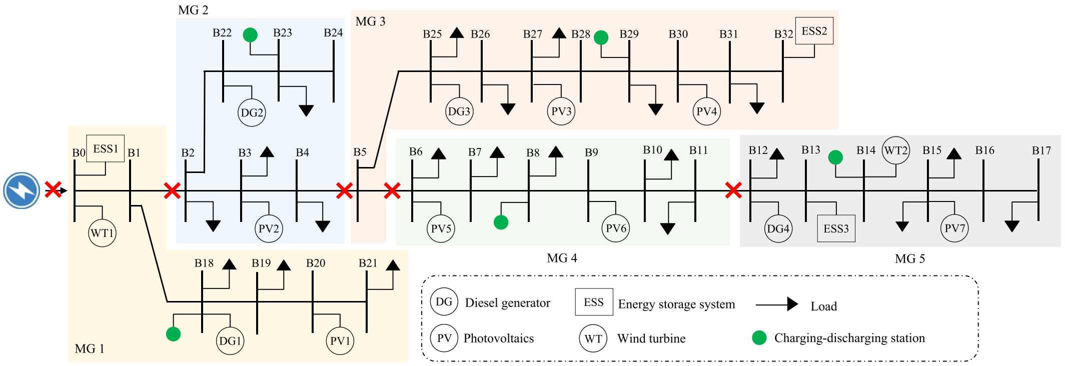


Fig. 6. The operation scenario consists of 5 independent MGs based on the modified IEEE 33-bus test system.

Parameter	Value	Parameter	Value
DG 1/DG 2 (MW)	0.4/0.3	PV 1/PV2 (MW)	0.4/0.3
DG 3/DG 4 (MW)	0.5/0.4	PV 3/PV 4 (MW)	0.4/0.3
WT 1/WT 2 (MW)	0.25/0.32	PV 5/PV 6 (MW)	0.25/0.35
ESS 1 Power/Capacity (MW/MWh)	0.5/1.0	PV 7 (MW)	0.5
ESS 2 Power/Capacity (MW/MWh)	1.5/3.0	DG unit generation cost α_m	0.65
ESS 3 Power/Capacity (MW/MWh)	0.5/1.0	ESS degradation cost c_m^{ess} (CNY/kWh)	0.2
ESS discharging efficiency η_m^{dis}	0.95	EV capacity $E_{\max}^{ev}$ (kWh)	100
EV energy consumption cost κ (kWh/mile)	3.8	EV degradation cost d_{bat}^{ev} (CNY/kWh)	0.1
EV unit time cost σ (CNY/h)	2	Load shedding cost c_m^{ls} (CNY/kWh)	10

Table 1. Technical parameters of each MG and EV.

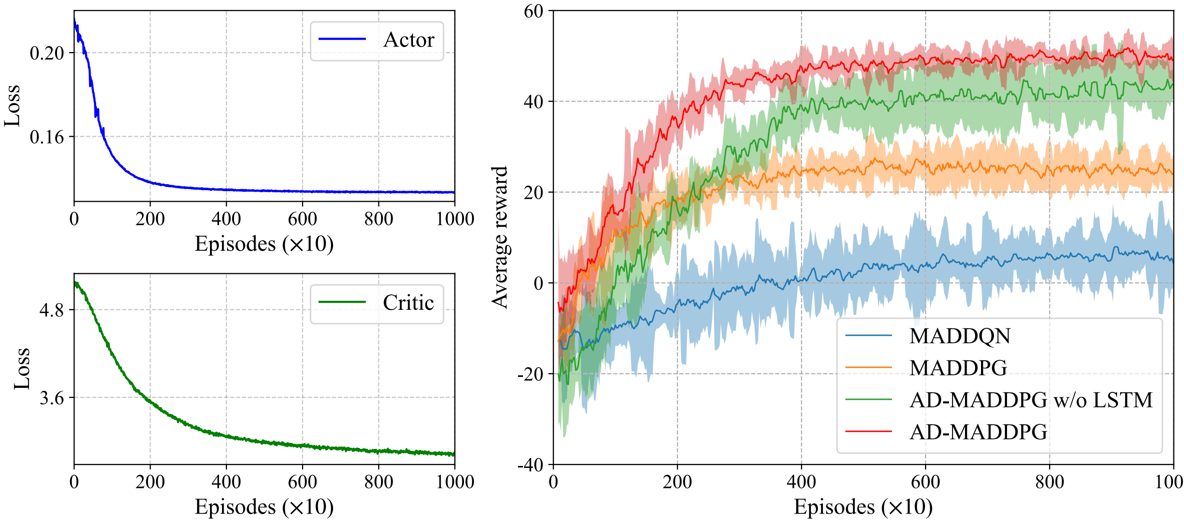


Fig. 7. Episodic average reward of EV agent over 10000 episodes for different methods.

utilizing internal DRES (e.g., PV and WT), DG, ESS and EVs to supply power to the loads within their respective islands. The technical parameters of each MG and EV are shown in Table 1.

For the DNN structure design, we use a 4-layer fully connected neural network with 2 hidden layers for the Actor, Critic, and Target networks. The neurons of hidden layers are 64, and the activation function is ReLU. Additionally, we employ a CNN with 3 hidden layers to extract spatial features. In the i-th layer of CNN, the number of filters is $16 \times 2^{(i-1)}$, each of size 3×3 with stride 2. To prevent gradient explosion, batch normalization is applied in the CNN, and layer normalization is adopted in the LSTM. For the LSTM, the gain is set to 1.0, and the shift is set to 0.0. The sequence length ε is chosen from the set $\{2, 3, 4\}$.

To validate the effectiveness of the proposed AD-MADDPG algorithm, we compared it with several baseline algorithms, which are listed as follows:

- MADDQN[15]: It investigates the optimization of routing and scheduling of mobile energy storage systems (MESS) for load restoration. To implement the MADDQN-based simulation, we discretized the routing and scheduling of EV to verify its convergence.
- MADDPG[35]: It is regarded as the state-of-the-art method for multi-agent deep reinforcement learning, demonstrating superior performance in cooperative and competitive multi-agent environments compared to other DRL methods. In this paper, we employ MADDPG's centralized training with decentralized execution to learn optimal routing and scheduling policies.
- AD-MADDPG w/o LSTM: In this version, the policy network relies only on the current observation $o_{i,t}$ at time slot t, without utilizing the proposed LSTM-enhanced Multi-step temporal sequence features extraction.

Evaluation of AD-MADDPG

Convergence analysis

The training curves for the EVs are compared and presented in Fig. 7. The left side of Fig. 7 shows the training loss for both the Actor and Critic networks, while the right side displays the average reward along with the standard deviation (shaded region) over 10000 episodes. From the figure, it can be observed that the loss values

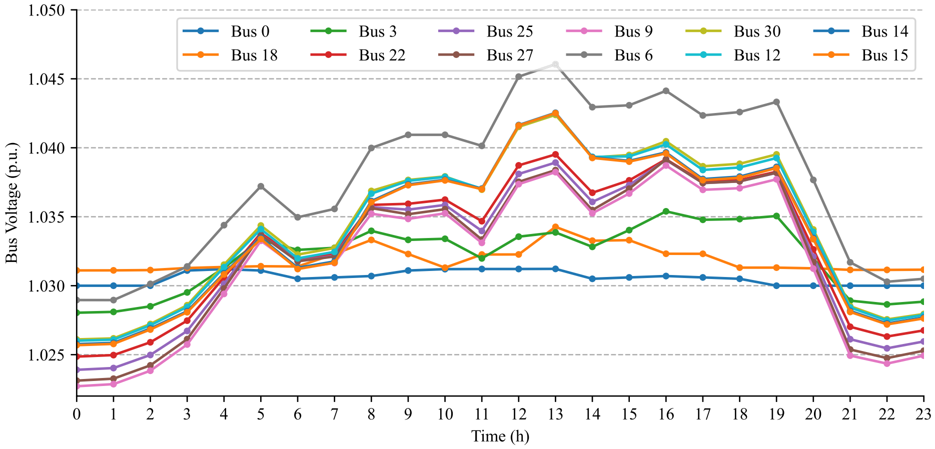


Fig. 8. The hourly voltage fluctuation curves for different buses in MMGs..

Method	MADDQN	MADDPG	AD-MADDPG w/o LSTM	AD-MADDPG
Num. of DNNs	*I*	2*I*	2*I*+3	2*I*+4
Num. of episodes[a]	6200	5500	4500	4200
Tot. training time (h)	1.48	0.96	1.02	1.03
Average training time per episode (s)	0.86	0.63	0.82	0.88
Tot. average reward	4.23	25.53	42.65	51.83

Table 2. Computational performance of the evaluated methods. [a]Each episode includes 12 time slots (i.e., $T = 12$) in our simulation setting

for both the Actor and Critic networks initially decrease rapidly due to the imprecise approximation, which cause high loss values at the beginning. With the use of multi-actor, single-learner, along with LSTM-enhanced Multi-step temporal sequence features extraction, the losses gradually converge to a smaller numerical range around 4000 episodes, effectively demonstrating the convergence of the proposed AD-MADDPG algorithm. Furthermore, the right side of Fig. 7 shows that our AD-MADDPG algorithm outperforms the three baselines in terms of both the converged average rewards and convergence speed. Specifically, the converged average reward for AD-MADDPG is 21.5% higher than AD-MADDPG w/o LSTM. These results indicate that our method is learning a more effective policy, maximizing load restoration in isolated MMGs while reducing the energy consumption of EVs.

Voltage analysis of MGs

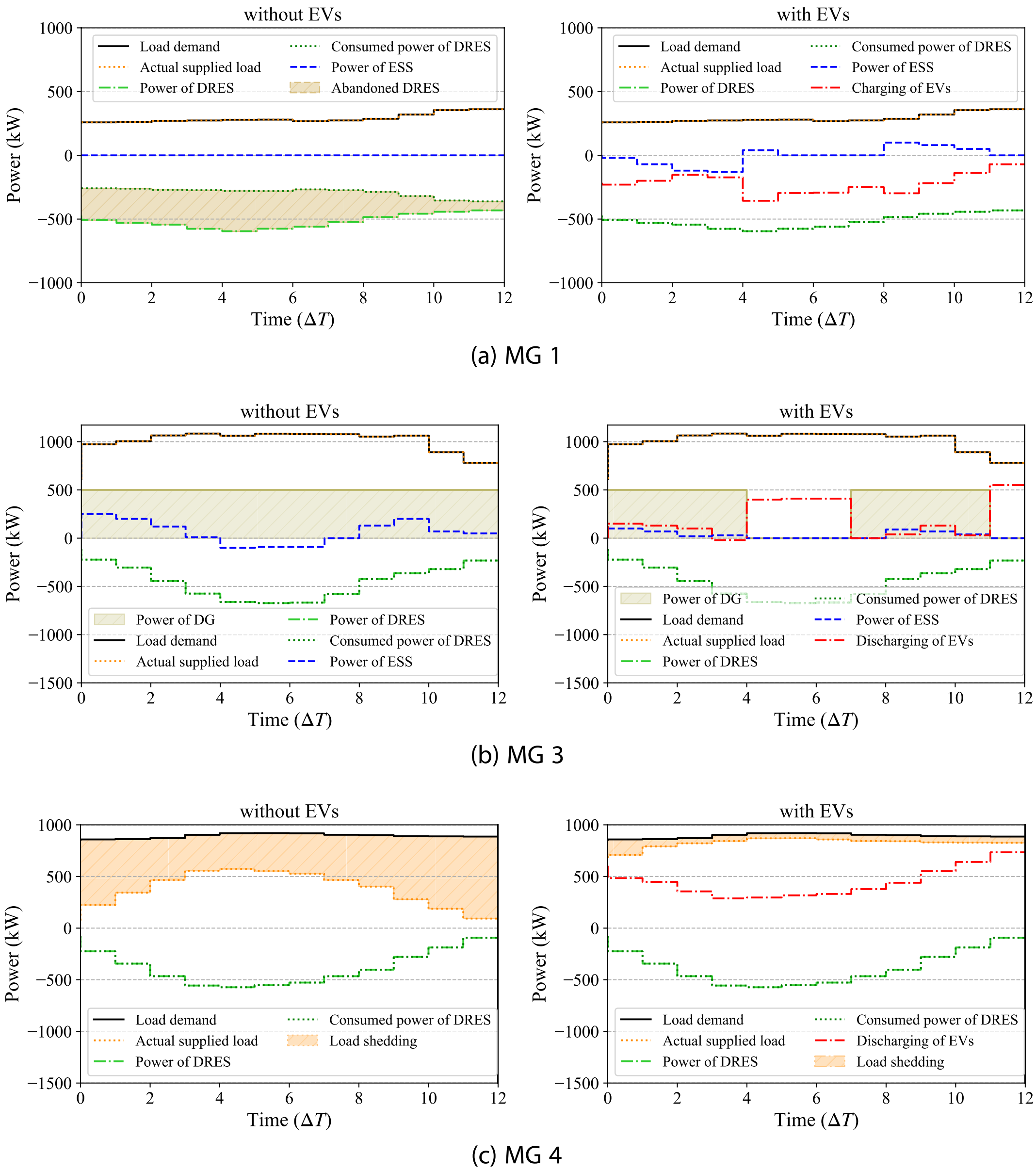
The hourly voltage fluctuation curves for different buses in MMGs are shown in Fig. 8. All bus voltages remain within the range of 0.95-1.05 p.u., confirming that each MG independently sustains stable voltage profiles. In particular, even Bus 6 with the largest fluctuations (±0.015 p.u.) strictly adheres to the safe range. In addition, voltage fluctuations intensify during 08:00–19:00 due to concurrent spikes in MG load consumption and renewable generation (PV/WT). The increased PV/WT output elevates local voltages, while load surges create temporary drops. Our coordination framework dynamically balances these opposing effects through EV scheduling, ensuring overall stability.

Computational performance

The computational performance of the evaluated methods is shown in Table 2. It is evident that MADDQN has the longest total training time, primarily due to the near-linear increase in the number of DNNs as the number of EVs grows and the large-scale discrete action space. In contrast, MADDPG has the shortest training time, thanks to its deterministic policy updates and simplified action selection process. Additionally, the AD-MADDPG algorithm converges in the fewest episodes, benefiting from the multi-actor, single-learner architecture of MADRL and the LSTM-enhanced Multi-step temporal sequence modeling. These results indicate that our AD-MADDPG algorithm achieves higher action exploration efficiency during training, leveraging multi-step expected discounted reward calculation. This leads to higher average rewards for EVs in load restoration and demonstrates superior computational performance within a reasonable training time.

Evaluation of load restoration

To verify the performance of EV scheduling policies for load restoration in islanded MMGs, we conducted a comparative analysis between scenarios with and without EV participation. The comparison results are illustrated in Fig. 9.



(a) MG 1

(b) MG 3

(c) MG 4

Fig. 9. The independent operation of each MG under different EVs strategies.

From Fig. 9a, it can be observed that MG 1 experiences issues with DRES (wind and solar) power curtailment due to an abundance of renewable energy when EVs are not participating in the strategy. However, when EVs are incorporated into MG 1's operation scheduling, the surplus renewable energy is effectively utilized. After being fully charged in MG 1, these EVs can leverage their mobile energy storage capability to discharge in other MGs where power generation is insufficient. Fig. 9b demonstrates that MG 3 achieves a basic supply-demand balance without EVs. However, this balance comes at the cost of extensive diesel generator use and frequent ESS charge-discharge cycles, leading to high operational costs and significant carbon emissions. With the participation of EVs, the reliance on diesel generation and ESS scheduling in MG 3 is significantly reduced, resulting in substantial cost savings. Fig. 9c shows that MG 4 experiences substantial load shedding when EVs are not involved, in order to maintain power balance due to insufficient output from DRES. When EVs participate, they facilitate a better load recovery, utilizing their batteries to discharge energy back into the MGs. Table 3 provides a comparison of the total costs under the two different strategies. Compared to Strategy 1, Strategy 2 achieves a reduction in both DRES curtailment costs and load shedding costs, directly saving an economic loss of 42659.33 (CNY), which accounts for approximately 52.53% of the total cost under Strategy 1.

Accordingly, under the strategy where EVs do not participate in the operation scheduling of MMGs, due to the lack of effective energy transmission channels, the surplus DRES in MG 1 cannot support the load of MG 4, resulting in curtailment of wind and solar power in MG 1 and load shedding in MG 4. Conversely, when EVs participate in the scheduling of MMGs, the mobile energy storage feature of EVs can be fully utilized to

Cost type	Strategies	
	Without EVs (Strategy 1)	With EVs (Strategy 2)
DRES curtailment cost (CNY)	2959.64	0
Load shedding cost (CNY)	68367.39	9700.32
Diesel generation (DG) cost (CNY)	9612.34	7223.86
ESS scheduling cost (CNY)	264.54	187.26
EVs scheduling cost (CNY)	0	21433.14
Total cost (CNY)	81203.91	38544.58

Table 3. The cost comparison of different scheduling strategies.

establish temporary electricity transmission channels, thereby transferring surplus renewable energy from MG 1 to support critical loads in MG 4, enhancing the overall operational resilience of the entire MMGs.

Evaluation of EVs scheduling

To evaluate the scheduling performance of EVs, we compare the proposed AD-MADDPG with three other baselines (as stated in section V.A.) in terms of the following three metrics.

- Load restoration ratio σ^l: determined as the ratio of the total amount of load restored to the initial amount of shed load over T time slots.
- Restoration fairness ω_t: defined by Eq. (22) to illustrate how evenly the load associated with MGs is restored by all EVs over T time slots.
- Energy consumption ratio σ^e: computed as the ratio of the total energy consumed (for movement) by all EVs to the initial energy reserve over T time slots.

We conduct three sets of simulations by varying the number of EVs *I*, maximum discharging proportion $P_{\max}^{sch}$, and the maximum moving distance l_{max} in a time slot. The comparison results in terms of load restoration ration σ^l, restoration fairness ω_t, and energy consumption ratio σ^e are illustrated in Fig. 10. As shown in Fig. 10a, we fixed $P_{\max}^{sch} = 70\%$, $l_{max} = 1.0$, while the number of EVs I changes form 20 to 60. In this case, we observe that our AD-MADDPG consistently outperforms the other three baselines in terms of load restoration. For example, when $I = 30$, AD-MADDPG achieves a load restoration ratio of 0.924, compared to 0.814 achieved by the best baseline AD-MADDPG w/o LSTM, making a 13.5% improvement. On average, our AD-MADDPG improves the load restoration ratio by 9.5%, 22.6%, and 41.5% over AD-MADDPG w/o LSTM, MADDPG and MADDQN, respectively. In addition, we can see that with the increase in the number of EVs, the load restoration, restoration fairness and energy consumption are increasing due to the increased total energy consumption. However, benefiting from the LSTM-based N-step temporal sequence modeling and multi-actor, single-learner architecture designing, our AD-MADDPG method facilitates effective collaboration among multiple EVs. This leads to better scheduling strategies during the learning process, thereby improving the load restoration ratio and restoration fairness, while keeping the increase in energy consumption relatively slow. Specifically, for the energy consumption ratio, our AD-MADDPG achieved an average reduction of 8.9%, 17.8%, and 21.4% over AD-MADDPG w/o LSTM, MADDPG, and MADDQN, respectively.

In Fig. 10b, we fixed $I = 30$, $l_{\max} = 1.0$, while the maximum discharging ratio $P_{\max}^{sch}$ changes from 0.4 to 0.8 with a step size of 0.1. As can be seen, AD-MADDPG outperforms all baselines in terms of load restoration, restoration fairness, and energy consumption. This is because as the maximum discharging ratio increases, EVs can strategically adjust their discharging rates based on their remaining energy status during each time slot. This enables them to supply more electrical energy to MGs experiencing energy deficits, thereby increasing the load restoration ratio. Furthermore, by discharging more energy in each time slot, EVs can reduce the frequency of their movements required for charging and discharging, consequently lowering energy consumption costs and enhancing overall operational efficiency. These results demonstrate that our spatial-temporal cooperative method effectively trains multiple agents to cooperate in a distributed manner, thereby enhancing the load restoration ratio while reducing energy consumption and maintaining a high level of restoration fairness.

Figure 10c evaluates the impact of maximum moving distance $l_{\max}$ of EVs in a time slot on the three metrics. We fixed $I = 30$, $P_{\max}^{sch} = 80\%$, while the maximum moving distance $l_{\max}$ changes from 0.6 to 1.4 with a step size of 0.2. As can be observed from Fig. 10c, compared to the three benchmark algorithms, the proposed algorithm demonstrates a significant performance improvement. Specifically, it achieves an average increase in load restoration ratio of 9.6%, 19.8%, and 25.5%, respectively. Additionally, it improves restoration fairness by 7.9%, 16.5%, and 40.9%, and reduces the energy consumption ratio by 8.2%, 15.7%, and 23.6% on average, respectively. We can see that as the maximum moving distance increases, the load restoration ratio of MMGs and restoration fairness gradually improve and reach the bottleneck, while the energy consumption ratio exhibits only slight fluctuations. This occurs because increasing the maximum moving distance of EVs within a single time slot allows the EVs to cover greater distances over the entire episode with a fixed number of slots. As a result, EVs can reach more distant MGs during the resilient load restoration process, providing more flexible routing and charge-discharge scheduling. This enhances the overall system's collaborative performance.

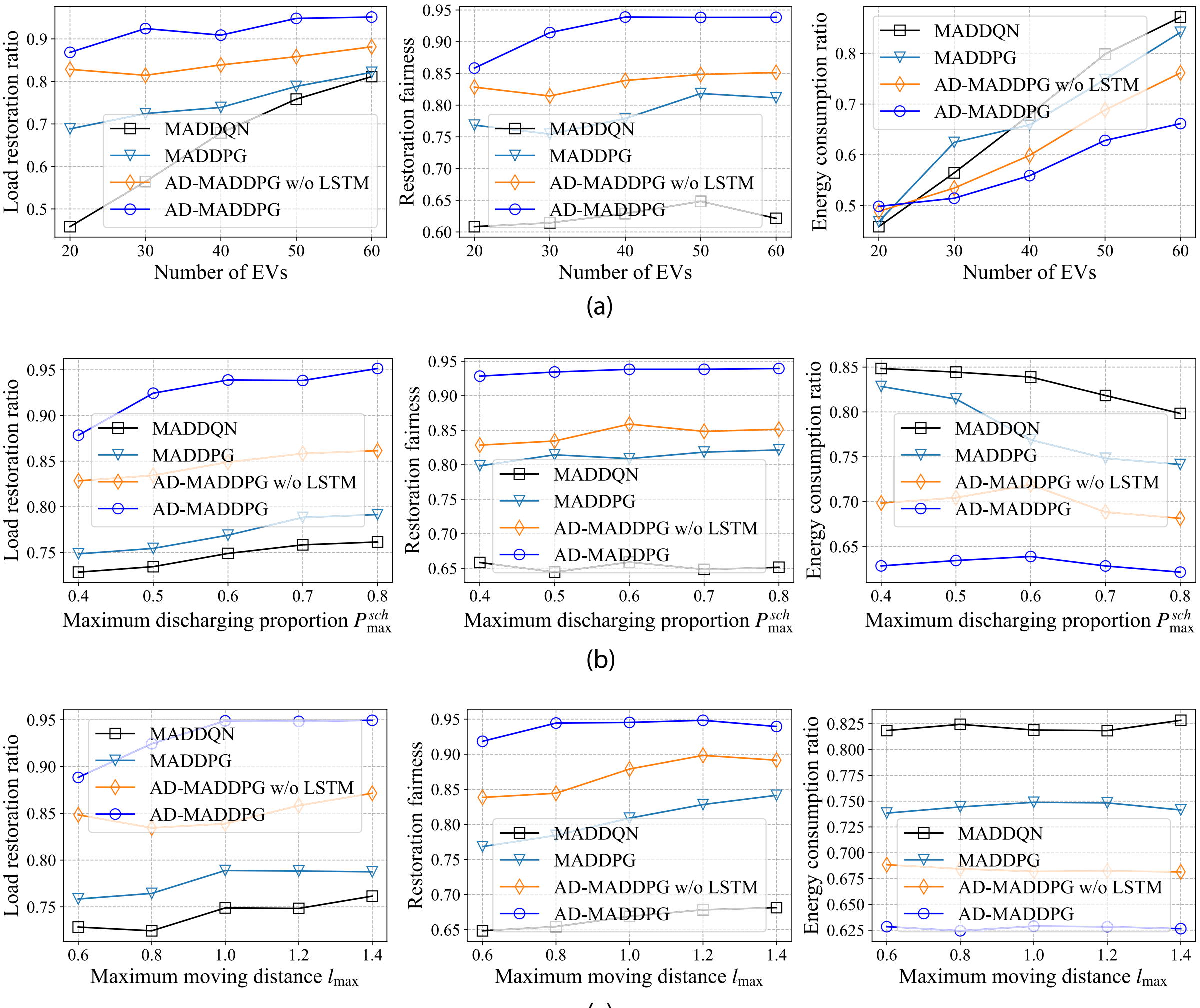


Fig. 10. The comparison of AD-MADDPG with the other three baselines in terms of load restoration, restoration fairness, and energy consumption.

Conclusion

This paper focuses on improving the resilient load restoration of MMGs by integrating EVs. The impact of EVs on load restoration of MMGs is analyzed using a constructed CPTN system, which couples the MMGs' power network with the transportation network. To reduce load-shedding costs during off-grid operation and save energy consumption of EVs, a distributed DDPG algorithm with a multi-actor, single-learner training structure is implemented to explore the performance of routing and charging-discharging scheduling. The main findings of this paper can be concluded as follows:

- To investigate resilient load restoration in MMGs, we constructed a coupled power-transportation network. This model represents the off-grid operation of MMGs and the optimal routing and scheduling of EVs, taking into account the uncertainty factors of both MGs and the transportation network.
- To maximize the load restoration and restoration fairness among MMGs while minimizing the energy consumption of EVs, a distributed multi-agent DDPG approach is proposed, called AD-MADDPG. This approach incorporates CNN for spatial feature extraction and LSTM for temporal sequence modeling. Furthermore, we make an improvement for AD-MADDPG by integrating a multi-actor, single-learner framework to improve the learning speed and quality.
- Simulation results demonstrate that: (1) the involvement of EVs in the resilient operation of MMGs significantly reduces the load shedding costs; (2) Compared to benchmark algorithms, the proposed AD-MADDPG utilizes a framework of multi-actor, single-learner, along with LSTM-enhanced Multi-step temporal features extraction, effectively accelerates DNN training and facilitates the learning of multi-agent cooperation strategies, thereby improving load restoration and restoration fairness while reducing energy consumption.

In future work, we will provide a more detailed modeling of the uncertainty factors within transportation networks and analyze their impact on EV scheduling. Additionally, curriculum learning and parameter sharing will be integrated into the multi-actor single-learner framework to further enhance the learning efficiency of the distributed multi-agent training process, thereby establishing a robust foundation for the practical application of the AD-MADDPG algorithm in real-world environments.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 14 March 2025; Accepted: 17 July 2025
Published online: 25 July 2025

References

1. Xiong, H. et al. Resilience enhancement for distribution system with multiple non-anticipative uncertainties based on multi-stage dynamic programming. *IEEE Trans. Smart Grid* **15**, 5706–5720 (2024).
2. Amini, F., Ghassemzadeh, S., Rostami, N. & Tabar, V. S. A stochastic two-stage microgrid formation strategy for enhancing distribution network resilience against earthquake event incorporating distributed energy resources, parking lots and responsive loads. *Sustain. Cities Soc.* **101**, 105191 (2024).
3. Kamal, F., Chowdhury, B. H. & Lim, C. Networked microgrid scheduling for resilient operation. *IEEE Trans. Ind. Appl.* **60**, 2290–2301 (2023).
4. Zhang, W. et al. Spatial-temporal resilience assessment of distribution systems under typhoon coupled with rainstorm events. *IEEE Trans. Ind. Inform.* **21**, 188–197 (2024).
5. Shi, W., Liang, H. & Bittner, M. Data-driven resilience enhancement for power distribution systems against multishocks of earthquakes. *IEEE Trans. Ind. Inform.* **20**, 7357–7369 (2024).
6. Qin, H. & Liu, T. Resilience enhancement of multi-microgrid system of systems based on distributed energy scheduling and network reconfiguration. *J. Electr. Eng. Technol.* **19**, 2135–2147 (2024).
7. Monteiro, M. R., de Souza, A. Z. & Abdelaziz, M. Hierarchical load restoration for integrated transmission and distribution systems with multi-microgrids. *IEEE Trans. Power Syst.* **39**, 7050–7063 (2024).
8. Liang, Z., Qian, T., Korkali, M., Glatt, R. & Hu, Q. A vehicle-to-grid planning framework incorporating electric vehicle user equilibrium and distribution network flexibility enhancement. *Appl. Energy* **376**, 124231 (2024).
9. Zou, X., Wang, Y. & Strbac, G. A resilience-oriented pre-positioning approach for electric vehicle routing and scheduling in coupled energy and transport sectors. *Sustain. Energy, Grids Netw.* **39**, 101484 (2024).
10. Mondal, S., Ghosh, P. & De, M. Evaluating the impact of coordinated multiple mobile emergency resources on distribution system resilience improvement. *Sustain. Energy, Grids Netw.* **38**, 101252 (2024).
11. Kazemtarghi, A. & Mallik, A. A two-stage stochastic programming approach for electric energy procurement of ev charging station integrated with bess and pv. *Electr. Power Syst. Res.* **232**, 110411 (2024).
12. Xie, H., Gao, S., Zheng, J. & Huang, X. A three-stage robust dispatch model considering the multi-uncertainties of electric vehicles and a multi-energy microgrid. *Int. J. Electr. Power Energy Syst.* **157**, 109778 (2024).
13. Chen, L., He, H., Jing, R., Xie, M. & Ye, K. Energy management in integrated energy system with electric vehicles as mobile energy storage: An approach using bi-level deep reinforcement learning. *Energy* **307**, 132757 (2024).
14. Wang, Y., Qiu, D., He, Y., Zhou, Q. & Strbac, G. Multi-agent reinforcement learning for electric vehicle decarbonized routing and scheduling. *Energy* **284**, 129335 (2023).
15. Wang, Y., Qiu, D. & Strbac, G. Multi-agent deep reinforcement learning for resilience-driven routing and scheduling of mobile energy storage systems. *Appl. Energy* **310**, 118575 (2022).
16. Liu, L., Huang, Z. & Xu, J. Multi-agent deep reinforcement learning based scheduling approach for mobile charging in internet of electric vehicles. *IEEE Trans. Mob. Comput.* **23**, 10130–10145 (2024).
17. Masrur, H., Shafie-Khah, M., Hossain, M. J. & Senjyu, T. Multi-energy microgrids incorporating ev integration: optimal design and resilient operation. *IEEE Trans. Smart Grid* **13**, 3508–3518 (2022).
18. Ebadat-Parast, M., Nazari, M. H. & Hosseinian, S. H. Distribution system resilience enhancement through resilience-oriented optimal scheduling of multi-microgrids considering normal and emergency conditions interlink utilizing multi-objective programming. *Sustain. Cities Soc.* **76**, 103467 (2022).
19. Wang, W., Xiong, X., He, Y., Hu, J. & Chen, H. Scheduling of separable mobile energy storage systems with mobile generators and fuel tankers to boost distribution system resilience. *IEEE Trans. Smart Grid* **13**, 443–457 (2022).
20. Gan, W., Wen, J., Yan, M., Zhou, Y. & Yao, W. Enhancing resilience with electric vehicles charging redispatching and vehicle-to-grid in traffic-electric networks. *IEEE Trans. Ind. Appl.* **60**, 953–965 (2023).
21. Kong, L., Zhang, H., Xie, D. & Dai, N. Leveraging electric vehicles to enhance resilience of interconnected power-transportation system under natural hazards. *IEEE Trans. Transp. Electrif.* **11**(1), 1126–40 (2024).
22. Wen, J., Gan, W., Chu, C.-C., Jiang, L. & Luo, J. Robust resilience enhancement by ev charging infrastructure planning in coupled power distribution and transportation systems. *IEEE Trans. Smart Grid* **16**, 491–504 (2024).
23. Lu, Z., Xu, X., Yan, Z. & Shahidehpour, M. Multistage robust optimization of routing and scheduling of mobile energy storage in coupled transportation and power distribution networks. *IEEE Trans. Transp. Electrif.* **8**, 2583–2594 (2021).
24. Zhang, X., Wang, Z. & Lu, Z. Multi-objective load dispatch for microgrid with electric vehicles using modified gravitational search and particle swarm optimization algorithm. *Appl. Energy* **306**, 118018 (2022).
25. Kumar, B. A. et al. Hybrid genetic algorithm-simulated annealing based electric vehicle charging station placement for optimizing distribution network resilience. *Sci. Rep.* **14**, 7637 (2024).
26. Shaheen, H. I., Rashed, G. I., Yang, B. & Yang, J. Optimal electric vehicle charging and discharging scheduling using metaheuristic algorithms: V2g approach for cost reduction and grid support. *J. Energy Stor.* **90**, 111816 (2024).
27. Abid, M. S. et al. A novel multi-objective optimization based multi-agent deep reinforcement learning approach for microgrid resources planning. *Appl. Energy* **353**, 122029 (2024).
28. Alqahtani, M. & Hu, M. Dynamic energy scheduling and routing of multiple electric vehicles using deep reinforcement learning. *Energy* **244**, 122626 (2022).
29. Gautam, M., Bhusal, N. & Ben-Idris, M. Postdisaster routing of movable energy resources for enhanced distribution system resilience: A deep reinforcement learning-based approach. *IEEE Ind. Appl. Mag.* **30**, 63–76 (2023).
30. Ding, Y., Chen, X. & Wang, J. Deep reinforcement learning-based method for joint optimization of mobile energy storage systems and power grid with high renewable energy sources. *Batteries* **9**, 219 (2023).
31. Zhao, C., Li, Y., Zhang, Q. & Ren, L. Low carbon economic energy management method in a microgrid based on enhanced d3qn algorithm with mixed penalty function. *IEEE Trans. Sustain. Energy* https://doi.org/10.1109/TSTE.2025.3528952 (2025).
32. Alqahtani, M., Scott, M. J. & Hu, M. Dynamic energy scheduling and routing of a large fleet of electric vehicles using multi-agent reinforcement learning. *Comput. Ind. Eng.* **169**, 108180 (2022).
33. Mao, S., Jin, J. & Xu, Y. Routing and charging scheduling for ev battery swapping systems: Hypergraph-based heterogeneous multiagent deep reinforcement learning. *IEEE Trans. Smart Grid* **15**(5), 4903–16 (2024).

34. Sepehrzad, R., Khodadadi, A., Adinehpour, S. & Karimi, M. A multi-agent deep reinforcement learning paradigm to improve the robustness and resilience of grid connected electric vehicle charging stations against the destructive effects of cyber-attacks. *Energy* **307**, 132669 (2024).
35. Sepehrzad, R., Faraji, M. J., Al-Durra, A. & Sadabadi, M. S. Enhancing cyber-resilience in electric vehicle charging stations: A multi-agent deep reinforcement learning approach. *IEEE Trans. Intell. Transp. Syst.* **25**, 18049–18062 (2024).
36. Wu, X., Li, J., Su, C., Fan, J. & Xu, M. A deep reinforcement learning based hierarchical eco-driving strategy for connected and automated hevs. *IEEE Trans. Veh. Technol.* **72**, 13901–13916 (2023).
37. Qiu, D., Wang, Y., Zhang, T., Sun, M. & Strbac, G. Hybrid multiagent reinforcement learning for electric vehicle resilience control towards a low-carbon transition. *IEEE Trans. Ind. Inform.* **18**, 8258–8269 (2022).
38. Shen, F. et al. Transactive energy based sequential load restoration of distribution systems with networked microgrids under uncertainty. *IEEE Trans. Smart Grid* **15**, 2601–2613 (2023).
39. Du, Y. & Wu, D. Deep reinforcement learning from demonstrations to assist service restoration in islanded microgrids. *IEEE Trans. Sustain. Energy* **13**, 1062–1072 (2022).
40. Sun, Y., Chen, Z., Li, Z., Tian, W. & Shahidehpour, M. Ev charging schedule in coupled constrained networks of transportation and power system. *IEEE Trans. Smart Grid* **10**, 4706–4716 (2018).
41. Hussain, A. & Musilek, P. Fairness and utilitarianism in allocating energy to evs during power contingencies using modified division rules. *IEEE Trans. Sustain. Energy* **13**, 1444–1456 (2022).

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant 62302154 and Grant 62306108, in part by the Natural Science Foundation of Hubei Province under Grant 2024AFB882, in part by the National Natural Science Foundation of China Cultivation Project under Grant 2024pygpzk09, in part by the Doctoral Scientific Research Foundation of Hubei University of Technology under Grant XJ2022006701, in part by the Key Research and Development Program of Hubei Province under Grant 2023BEB024; in part by Open Project Funding of the Key Laboratory of Intelligent Sensing System and Security (Hubei University), Ministry of Education under Grant KLISSS202404.

Author contributions

Yuxin Wu wrote the manuscript and performed the data analysis; Ting Cai performed the formal analysis; Xiaoli Li performed the validation.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to T.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.