machines

MDPI

Article

Using Graph-Enhanced Deep Reinforcement Learning for Distribution Network Fault Recovery

Yueran Liu, Peng Liao * and Yang Wang

College of Electrical Engineering, Sichuan University, Chengdu 610065, China; 2022141440102@stu.scu.edu.cn (Y.L.); fwang@scu.edu.cn (Y.W.)
* Correspondence: sculiaopeng@163.com

Abstract

Fault recovery in distribution networks is a complex, high-dimensional decision-making task characterized by partial observability, dynamic topology, and strong interdependencies among components. To address these challenges, this paper proposes a graph-based multi-agent deep reinforcement learning (DRL) framework for intelligent fault restoration in power distribution networks. The restoration problem is modeled as a partially observable Markov decision process (POMDP), where each agent employs graph neural networks to extract topological features and enhance environmental perception. To address the high-dimensionality of the action space, an action decomposition strategy is introduced, treating each switch operation as an independent binary classification task, which improves convergence and decision efficiency. Furthermore, a collaborative reward mechanism is designed to promote coordination among agents and optimize global restoration performance. Experiments on the PG&E 69-bus system demonstrate that the proposed method significantly outperforms existing DRL baselines. Specifically, it achieves up to 2.6% higher load recovery, up to 0.0 p.u. lower recovery cost, and full restoration in the midday scenario, with statistically significant improvements ($p < 0.05$ or $p < 0.01$). These results highlight the effectiveness of graph-based learning and cooperative rewards in improving the resilience, efficiency, and adaptability of distribution network operations under varying conditions.

Keywords: fault recovery; deep reinforcement learning; multiple agents; partially observable Markov decision process; distributed energy resources

check for updates

Academic Editor: Davide Astolfi

Received: 22 May 2025
Revised: 15 June 2025
Accepted: 21 June 2025
Published: 23 June 2025

Citation: Liu, Y.; Liao, P.; Wang, Y. Using Graph-Enhanced Deep Reinforcement Learning for Distribution Network Fault Recovery. *Machines* **2025**, *13*, 543. https://doi.org/10.3390/machines13070543

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The increasing frequency of extreme natural disasters has posed significant challenges to the safe and stable operation of power systems [1–3]. As a critical energy transmission network connecting the power grid to end users, the distribution network plays a fundamental role in national economic and social development. However, during large-scale power outages, traditional restoration strategies often rely on support from the transmission system, limiting the speed and flexibility of recovery efforts. This dependency hampers the ability to respond swiftly and mitigate the impact of disasters [4,5]. Therefore, leveraging local resources within the distribution network to enable autonomous and efficient fault restoration has become a critical research focus for enhancing power system security [6–9].

Following a widespread blackout, effectively utilizing distributed energy resources (DERs) to restore critical loads can significantly shorten outage durations for essential users, ensure the stable operation of critical infrastructure, and reduce economic losses [10].

Moreover, as power systems shift toward a high penetration of renewable energy, localized restoration strategies that integrate DERs are crucial for enhancing national energy security and maintaining social stability [11,12].

The primary objective of the distribution network fault restoration problem is to optimize line switching states, load restoration strategies, and DER output dispatch while satisfying operational constraints and power capacity limitations, thereby maximizing the restored load [13]. This problem is typically formulated as a mixed-integer optimization problem, such as mixed-integer linear programming [14,15], mixed-integer quadratic programming [16,17], or mixed-integer second-order cone programming [18,19], and solved using optimization solvers. However, these mathematical programming-based methods face several limitations. First, they heavily rely on an accurate mathematical model of the distribution network, making them less adaptable to complex and dynamic environments, thereby limiting their generalization and scalability. Second, constructing precise network models is computationally expensive, and different modeling approaches and mathematical formulations significantly impact solver efficiency and convergence. As a result, applying such methods in real-time decision-making scenarios remains a challenge.

Recently, deep reinforcement learning (DRL) has emerged as a promising approach for solving complex decision-making and optimization problems, particularly in high-dimensional, continuous decision spaces [20–22]. Unlike traditional optimization methods, DRL continuously refines its strategy through interactions with the environment, maximizing cumulative rewards without requiring an explicit mathematical model [23–25]. This capability makes DRL highly adaptable to complex and uncertain decision problems, particularly those in dynamic, hard-to-model power system operations [26,27]. For example, the authors of [28] proposed a distribution network fault restoration method based on deep Q-network (DQN), incorporating Monte Carlo tree search to accelerate training. Similarly, the authors of [29] employed DQN for real-time distribution system restoration and introduced a hybrid imitation learning approach to improve training efficiency. However, DQN is primarily designed for discrete action spaces and struggles with large-scale, multi-dimensional decision-making in distribution systems. To address this limitation, Du and Wu [30] proposed a fault restoration method based on deep deterministic policy gradient (DDPG), which is capable of handling continuous action spaces and coordinating multiple DERs for restoration. Further advancements have been made with multi-agent DRL [31], where a multi-agent DDPG (MADDPG) significantly improved computational efficiency and restoration performance compared to single-agent DRL approaches [32].

While DRL has made progress in distribution network fault restoration, several critical challenges remain.

- Ignoring the spatial topology of the distribution network, limiting global situational awareness. Distribution networks naturally exhibit a graph topology, where network connectivity directly influences power flow distribution. Neglecting these spatial features can hinder the agent's ability to fully understand the global network state, thereby affecting the effectiveness and scalability of restoration strategies.
- A high-dimensional discrete action space leading to exponential growth in computational complexity. In the fault restoration process, an agent must determine the switching status of multiple circuit breakers simultaneously. As the number of circuit breakers increases, the possible action combinations grow exponentially, making it difficult for the agent to efficiently explore the optimal strategy in such a vast action space. This "curse of dimensionality" often results in slow training convergence and poor decision quality.
- Lack of effective cooperation mechanisms among multiple agents, affecting global restoration performance. Fault restoration in distribution networks involves the

coordinated optimization of multiple physical components, naturally presenting a multi-agent interaction problem. However, most existing studies design independent reward functions for each agent, failing to capture the cooperative relationships among different decision units. As a result, agents may struggle to coordinate their actions effectively, leading to suboptimal overall restoration performance.

To address these challenges, this paper proposes a graph-based multi-agent DRL approach for optimizing distribution network fault restoration. The fault recovery problem is formulated as a partially observable Markov decision process (POMDP), with graph learning techniques introduced to enhance environmental feature representation. To mitigate the curse of dimensionality in the action space, we propose an action decomposition strategy that treats each circuit breaker's switching decision as an independent binary classification task. This transforms a high-dimensional action space into multiple lower-dimensional subspaces, significantly improving learning efficiency. Furthermore, each agent is enhanced by a graph-based neural network to extract meaningful features from graph-structured data. Finally, a collaborative reward mechanism is designed to promote cooperative behavior among agents during the restoration process. Table 1 presents a comparative summary of the proposed method and several recent DRL-based fault restoration approaches, highlighting the key methodological differences and advantages.

Table 1. Comparison of Recent DRL Methods.

Approach	Network Topology Modeling	Action Space Handling	Multi-Agent Cooperation	Reward Design	Learning Algorithm
Reference [28]	No	Discrete	No	Local	DQN
Reference [29]	No	Discrete	No	Local	DQN
Reference [30]	No	Continuous	No	Local	DDPG
Reference [31]	No	Continuous	Yes	Local	MADDPG
Our Paper	Yes	Action decomposition	Yes	Global	MADDPG + GCN

Compared to existing approaches, the proposed method enhances fault recovery performance by incorporating spatial topology through GCNs, simplifying decision-making via action decomposition, and promoting agent collaboration with a global reward design. Built upon the MADDPG framework, it offers a scalable and effective solution for fault recovery in modern distribution networks. The core innovations and contributions of this work are summarized as follows.

1. Graph-based state representation: To capture the topological structure and interdependencies among network components, we model the distribution network as a graph and employ a graph-based neural network for feature extraction. This enables agents to effectively learn spatial correlations between network nodes, enhancing their global situational awareness and improving fault restoration effectiveness.
2. Action decomposition strategy: To address the computational challenges of high-dimensional action spaces, we propose an action decomposition approach, treating each circuit breaker's switching operation as an independent binary classification problem. By leveraging the continuity of neural network outputs, this method simplifies discrete action selection, reducing computational complexity while improving convergence and action efficiency.
3. Collaborative reward mechanism: To enhance cooperation among multiple agents, we develop a collaborative reward mechanism that promotes information sharing and coordination. By designing a global reward function that accounts for the overall

system restoration performance, agents not only optimize their individual actions but also contribute to achieving a globally optimal restoration strategy.

The rest of the paper is organized as follows. Section 2 formulates the power distribution network fault recovery problem as a POMDP, where the state representation of the distribution network is enhanced using graph-structured data. Section 3 provides a detailed description of the multi-agent DRL framework. Section 4 presents a case study, including comparative analysis and performance validation. Finally, Section 5 concludes the paper.

2. Partially Observable Markov Decision Process Formulation

DRL enables agents to learn and optimize their decision-making strategies through continuous interaction with the environment, as illustrated in Figure 1.

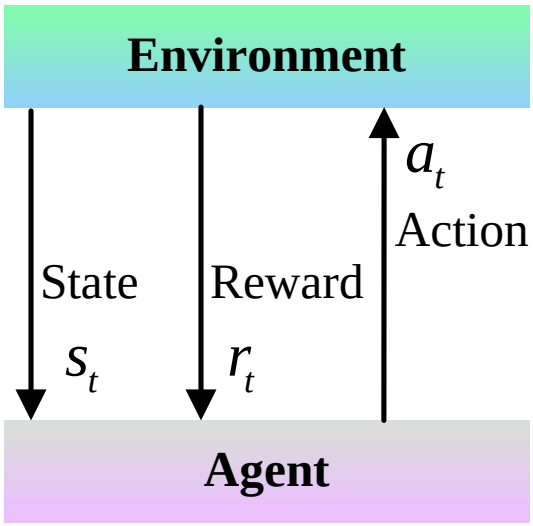


Figure 1. The Markov decision process.

At time step t, the environment is in state s_t, and the agent selects an action a_t based on its policy. The environment then transitions to a new state s_{t+1} and provides an immediate reward r_t as feedback. The agent iteratively refines its decision-making by continuously interacting with the environment, exploring different actions, and updating its policy to maximize cumulative rewards, ultimately achieving optimal control.

The fault recovery process in a power distribution network is a complex sequential decision-making problem, involving multiple agents interacting with a dynamic environment. Specifically, the system consists of two types of agents.

- The network reconfiguration agent (NRA), responsible for optimizing network topology to isolate faults and restore network connectivity.
- The power scheduling agent (PSA), which schedules DERs to minimize restoration costs.

This multi-agent decision-making process is formulated as a POMDP, characterized by five key components: the state space, observation space, action space, state transition function, and reward function. To facilitate understanding and provide an intuitive overview of the underlying structure, a flowchart of the POMDP framework is presented in Figure 2. The diagram illustrates the interactions among these five components and highlights their roles in the decision-making process.

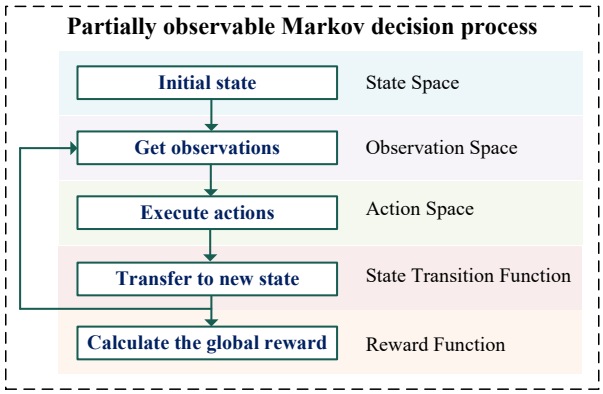


Figure 2. Flowchart of the POMDP framework.

A detailed mathematical formulation of the POMDP framework, including precise definitions of each component, is provided in the following subsections (see Table A1 for the mathematical symbol definition table for POMDP).

2.1. State Space

The state space $\mathcal{S}$ represents the set of all possible system states. In the distribution network fault recovery problem, the system state at each time step includes power flow conditions, DER generation levels, and load demands. Considering the inherent topology of the power grids, we enhance state representation by encoding it as a graph-structured feature set. This allows graph neural networks to effectively capture topological dependencies and improve feature extraction.

Thus, the state at each time step t is formulated as

$$s_t = G_t(N, E, \mathbf{F}) \tag{1}$$

where G represents the graph representation of the power grid, N is the set of nodes, E is the set of edges, and $\mathbf{F}$ denotes the node feature matrix encoding attributes such as power flow conditions and DER availability.

2.2. Observation Space

The observation space $\mathcal{O}$ represents the set of observations available to each agent. Since the distribution network state is represented as a graph, each agent's observations include the adjacency matrix and the node feature matrix.

- The adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ encodes the topological structure of the power grid, where each element x_{ij} is defined as

$$x_{ij} = \begin{cases} 1 & \text{if node } i \text{ is directly connected to node } j \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

- The node feature matrix describes node-specific attributes, where each row corresponds to a node, and each column represents a specific feature. It is defined as

$$\mathbf{F} \in \mathbb{R}^{n \times d} \tag{3}$$

where n is the number of nodes, and d is the number of features per node. The feature vector of node i is given by $f_i \in \mathbb{R}^d$.

To support effective action generation, the node feature matrices for the NRA and PSA at time step t are defined as

$$\mathbf{F}_t^{\mathrm{NRA}} = \left[f_{1,t}^{\mathrm{NRA}}, f_{2,t}^{\mathrm{NRA}}, ..., f_{n,t}^{\mathrm{NRA}} \right] \tag{4}$$

$$\mathbf{F}_t^{\mathrm{PSA}} = \left[f_{1,t}^{\mathrm{PSA}}, f_{2,t}^{\mathrm{PSA}}, ..., f_{n,t}^{\mathrm{PSA}} \right] \tag{5}$$

where $f_{i,t}^{\mathrm{NRA}}$ and $f_{i,t}^{\mathrm{PSA}}$ are the feature vectors of node i at time step t, defined as

$$f_{i,t}^{\mathrm{NRA}} = \left[P_{i,t}^{\mathrm{load}}, P_{i,t}^{\mathrm{der,max}}, P_{i,t}^{\mathrm{der,min}} \right] \tag{6}$$

$$f_{i,t}^{\mathrm{PSA}} = [\mathcal{P}_{i,t}, \mathcal{V}_{i,t}, \mathcal{Y}_{i,t-1}] \tag{7}$$

where $P_{i,t}^{\mathrm{load}}$ denotes the power demand of the load at node i and time step t. $P_{i,t}^{\mathrm{der,max}}$ and $P_{i,t}^{\mathrm{der,min}}$ denote the maximum and minimum available power generation of the DER at node i and time step t, respectively. $\mathcal{P}_{i,t}$ and $\mathcal{V}_{i,t}$ correspond to the active power injection and voltage magnitude at node i and time step t, respectively. $\mathcal{Y}_{i,t-1}$ is a binary variable indicating whether the load at node i was restored at time step $t-1$.

2.3. Action Space

The action space $\mathcal{A}$ denotes the set of all possible actions that each agent can take at a given time step. The NRA modifies the grid topology by controlling the switching actions of circuit breakers. At each time step t, the action taken by the NRA is defined as

$$a_t^{\mathrm{NRA}} = \{\chi_t^{\tau} | \tau \in \Gamma\} \tag{8}$$

where $\chi_t^{\tau} \in [0,1]$ represents the probability of closing switch τ at the time step t, and Γ denotes the set of all controllable switches.

The PSA regulates the power output of DERs to balance supply and demand. The action taken by the PSA is formulated as

$$a_t^{\mathrm{PSA}} = \left\{ \ell_t^{\wp} | \wp \in \mho \right\} \tag{9}$$

where $\ell_t^{\wp} \in [0,1]$ represents the power generation scaling factor of DER $\wp$ at time step t, and $\mho$ denotes the set of DERs. The actual power output of DER $\wp$ is then computed as

$$\mathcal{P}_t^{\wp} = \ell_t^{\wp} \cdot P_t^{\wp,\max} \tag{10}$$

where $P_t^{\wp,\max}$ is the maximum available generation capacity of DER $\wp$ at time step t.

2.4. State Transition Function

The state transition function models the probability distribution of transitioning from the current state s_t to the next state s_{t+1} given an action a_t.

In the distribution network fault recovery problem, the state transition is typically deterministic, meaning that the next state is uniquely determined by the applied actions. Therefore, the transition function can be expressed as a deterministic mapping, defined as

$$s_{t+1} = \rho(s_t, a_t) \tag{11}$$

The transition function $\rho(\cdot)$ is implemented using PandaPower, an open-source, Python-based power system analysis tool designed for simulating network reconfiguration and power flow dynamics in electrical grids.

PandaPower provides a high-level interface for power flow calculations, network topology manipulation, and fault scenario simulation. In our work, it enables the precise modeling of network reconfiguration processes, including switch operations, as well as accurate computation of resulting system states such as bus voltages, line loading, and supply status. By integrating PandaPower into our state transition function, we ensure that the effects of each recovery action are evaluated under realistic electrical and topological constraints, enhancing the fidelity and practical relevance of our POMDP-based formulation. A comprehensive description of PandaPower's architecture and capabilities is available in [33].

2.5. Reward Function

The objective of DRL is to maximize the long-term cumulative reward, defined as

$$R_t = \sum_{k=t}^{T} \gamma^{k-t} \cdot r_k \tag{12}$$

where R_t represents the discounted cumulative reward at time step t, r_k denotes the immediate reward received at time step k. $\gamma \in [0, 1]$ is the discount factor, which balances the trade-off between immediate and future rewards. T represents the terminal time step.

The NRA is incentivized to optimize network topology for efficient fault isolation and service restoration. Its reward function is given by

$$r_t^{\mathrm{nr}} = \frac{P_t^{\mathrm{res}} - P_{t-1}^{\mathrm{res}}}{\sum_{i \in N} P_{i,t}^{load}} \tag{13}$$

where P_t^{res} denotes the total load power restored after the action at time step t, and $P_t^{\mathrm{res}} - P_{t-1}^{\mathrm{res}}$ quantifies the incremental improvement in fault recovery compared to the previous time step.

The PSA is encouraged to optimize DER dispatch while minimizing operational costs. Its reward function is defined as

$$r_t^{\mathrm{ps}} = - \sum_{\wp \in \mho} \lambda_{\wp} \cdot P_{\wp,t}^{der} \tag{14}$$

where $\lambda_{\wp}$ represents the cost coefficient associated with DER $\wp$, and $P_{\wp,t}^{der}$ denotes the actual power generation of DER $\wp$ at time step t.

To encourage collaborative decision-making among agents and achieve system-wide optimality, we design a global reward function, given by

$$r_t^{\mathrm{global}} = \alpha \cdot r_t^{\mathrm{nr}} + \beta \cdot r_t^{\mathrm{ps}} + \kappa \cdot \wp_t \tag{15}$$

where r_t^{global} is the global reward at time step t, while r_t^{nr} and r_t^{ps} correspond to the rewards for network reconfiguration and power scheduling at time step t, respectively. The term $\wp_t$ acts as a penalty function to enforce system constraints during fault recovery. The coefficients α, β and κ balance the contributions of different reward components.

This global reward mechanism ensures that each agent's policy improvement contributes to maximizing the overall system performance rather than focusing on localized benefits. By aligning all agents toward a shared objective, this approach reduces conflicts among local policies and improves the overall efficiency of the fault recovery process.

3. Multi-Agent Deep Reinforcement Learning Framework

Based on the POMDP modeling of the distribution network fault recovery problem, we propose a novel multi-agent DRL framework, which includes a graph-enhanced Actor–Critic architecture and multi-agent training algorithm to achieve efficient autonomous fault recovery strategies.

3.1. Graph-Enhanced Actor–Critic Architecture

In this framework, the NRA and the PSA collaboratively interact to achieve fault recovery by network reconfiguration and optimized dispatching. Each agent makes independent decisions based on local observations and optimizes its policy using a global reward, ensuring an improved system-wide recovery performance.

The decision-making of each agent follows the actor–critic DRL architecture, which consists of an actor network and a critic network.

- Actor network, which determines actions based on local observations.
- Critic network, which evaluates action quality using global information.

However, conventional fully connected actor–critic networks may fail to capture the topological connectivity of the distribution network, making them ineffective for structural changes during fault recovery. To address this, we integrate a GCN into the actor–critic architecture, enhancing policy generation and evaluation capabilities.

- The actor network extracts graph-based latent features from local observations and maps them to optimal actions.
- The critic network leverages global graph representations to estimate Q-values more effectively.

The process of action generation and evaluation is illustrated in Figure 3. In the action generation stage, the raw observations from the actual power grid are first abstracted into graph data, consisting of an adjacency matrix and a node feature matrix. Next, a two-layer GCN is employed to extract node-level latent features. These extracted features are then mapped to the action space through the actor network. To facilitate decision-making, a Flatten layer compresses the graph-structured output of the GCN into a one-dimensional format. The transformed features are then activated using the ReLU function and passed through an FCN to generate actions. The final action values are obtained through an activation function. Similarly, in the policy evaluation stage, the critic network maps the concatenated graph data and action to a Q-value, using multiple FCNs and activation functions. This process enables the critic to effectively assess the quality of the current policy.

The GCN propagates information through graph structures by aggregating features from neighboring nodes to update the target node's feature representation. Each GCN layer performs localized message passing, where information is aggregated and transformed. The process is mathematically represented as

$$\mathcal{H}^{(l+1)} = \sigma\left(\tilde{D}^{-\frac{1}{2}}\tilde{A}\tilde{D}^{-\frac{1}{2}}\mathcal{H}^{(l)}W^{(l)}\right) \tag{16}$$

where $\mathcal{H}^{(l)}$ denotes the feature matrix at layer l, and $W^{(l)}$ is a trainable weight matrix. $\tilde{A}$ is the normalized adjacency matrix, and $\tilde{D}$ is the degree matrix for normalization. σ represents a non-linear activation function.

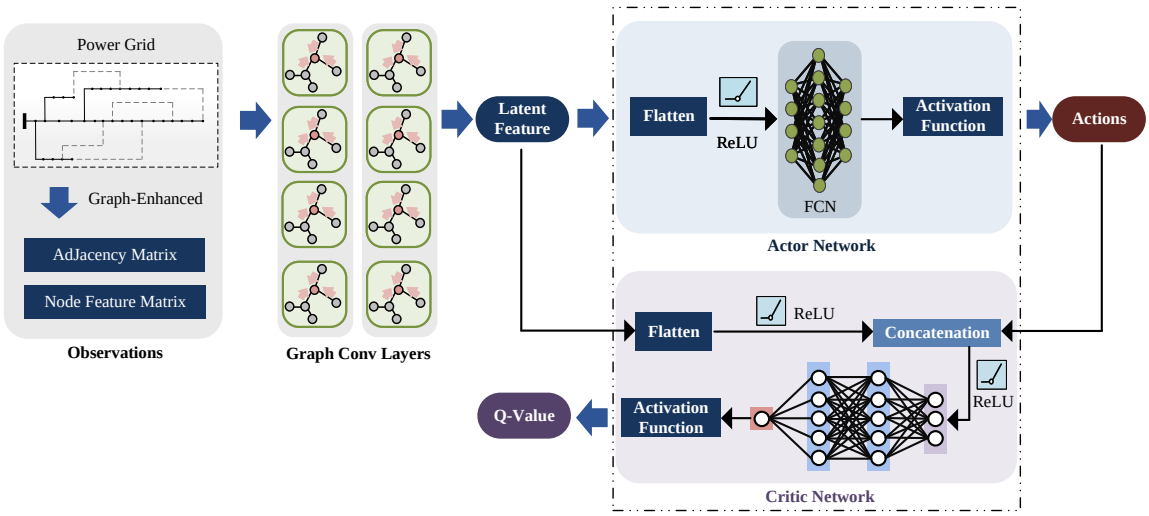


Figure 3. Action generation and evaluation process.

Hierarchical propagation in the GCN allows information diffusion across the network topology, as illustrated in Figure 4. In the first GCN layer, node 1 aggregates features from its direct neighbors (e.g., nodes 0, 2, and 4). In the second GCN layer, as each node has already aggregated information from its neighbors, node 1 indirectly receives information from second-hop neighbors (e.g., nodes 3 and 5).

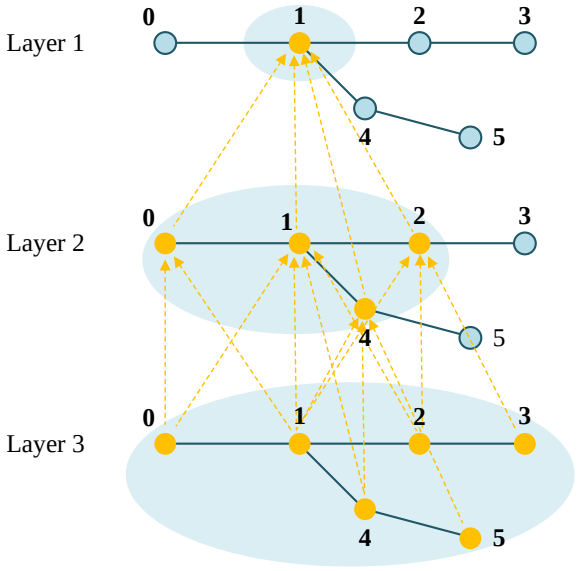


Figure 4. Multilayer GCN information dissemination.

This structure is particularly advantageous for fault recovery, as it allows agents to effectively capture topological changes and interactions among multiple nodes, significantly improving decision-making.

By leveraging graph-based feature extraction, the proposed architecture enhances the agent's capability to capture the dynamically evolving topology of the distribution network. As a result, it significantly improves policy learning and optimizes autonomous fault recovery strategies.

3.2. Multi-Agent Training Algorithm

To effectively train the agents in the graph-enhanced actor–critic architecture, we adopt the MADDPG algorithm, which is well-suited for the cooperative yet independent

decision-making nature of NRA and PSA in the distribution network fault recovery process. MADDPG extends the conventional DDPG to multi-agent settings, enabling centralized training with decentralized execution.

- During training, each agent's critic network has access to global state information and the actions of all agents. This helps in learning more stable and coordinated policies by leveraging system-wide observations.
- During execution, each agent makes decisions independently using only its local observations, ensuring practical applicability in real-world deployment.

The critic function for agent i is, therefore, trained using a centralized Q-function $Q_i(s, a_1, a_1, \ldots, a_n)$, where s represents the global system state, and $a_1, a_1, \ldots, a_n$ denote the actions of all agents.

Each agent updates its policy using gradient ascent based on the deterministic policy gradient, as shown in

$$\nabla_{\theta_i} \mathcal{J}(\theta_i) = \mathbb{E}_{s,a\sim\mathcal{D}}\left[\nabla_{\theta_i} \pi_i(a_i \mid o_i) \nabla_{a_i} Q_i(s, a_1, a_1, \ldots, a_n)\right] \tag{17}$$

where π_i is the policy of agent i, and θ_i denotes the parameters of the actor network. o_i represents the local observation of agent i. Q_i is the centralized Q-function incorporating the actions of all agents. $\mathcal{D}$ is the experience replay buffer, which stores past transitions for stable training.

The critic network is updated using the mean squared Bellman error, as shown in

$$L(\psi_i) = \mathbb{E}_{s,a,r,s'}\left[\left(Q_i(s, a_1, a_1, \ldots, a_n) - y\right)^2\right] \tag{18}$$

where ψ_i denotes the parameters of critic network of agent i, y is the target value, which is computed as

$$y = r^{global} + \gamma Q_i'(s', a_1', \ldots, a_n') \tag{19}$$

where Q_i' is the estimated Q-value, s' represents the next state, and $a_1, a_1, \ldots, a_n$ denotes the next actions of all agents.

To improve coordination between the NRA and PSA, we introduce an experience-sharing mechanism.

- NRA shares topology changes with the PSA, allowing the PSA to adjust power dispatch accordingly.
- PSA provides load restoration feedback to the NRA, enabling it to optimize reconfiguration decisions.
- A global reward function (as defined in Section 2.5) ensures that both agents align their objectives to maximize system-wide restoration efficiency.

This training strategy enhances scalability, stability, and cooperation, making it well-suited for real-world autonomous fault recovery.

For a detailed list of mathematical symbols used in this section, please refer to Table A2, which provides the mathematical symbol definition table for DRL.

4. Case Study

This section evaluates the effectiveness of the proposed graph-enhanced multi-agent DRL approach using the PG&E 69-bus distribution system. The evaluation focuses on training performance, comparative analysis with baseline methods, and testing under realistic fault recovery scenarios. The distribution network environment is modeled using PandaPower, which simulates the electrical behavior of the network. The DRL agents and main training routines are implemented in Python 3.10 using PyTorch 1.11.0.

4.1. Experimental Setup

The PG&E 69-bus system, as shown in Figure 5, consists of 69 nodes and 78 distribution lines, including 74 normally closed feeders and 5 normally open tie-lines. Specifically, these 5 tie-lines are part of the standard benchmark configuration and are widely adopted in the literature [34–38] as normally open switches that can be closed for network reconfiguration or restoration purposes.

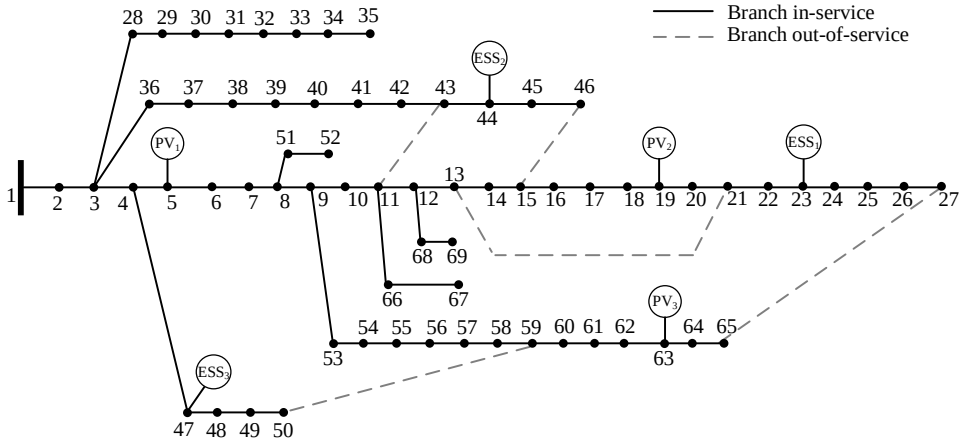


Figure 5. Layout of the PG&E 69-bus distribution system.

Node 1 serves as the substation node, supplying all downstream nodes with a nominal voltage of 12.66 kV. The line parameters and load profiles are adopted from reference [39]. The system is integrated with DERs, including photovoltaic (PV) systems and energy storage systems (ESSs). The PV units are considered uncontrollable, while the ESS units are controllable and participate in dispatch operations. Specifically, PV systems are installed at buses 5, 19, and 63, while ESS units are deployed at buses 23, 44, and 47. Note that the locations of the ESS units were not optimized, but were strategically selected to reflect realistic deployment scenarios and to provide broad spatial coverage across the network. This choice aligns with the study's goal of evaluating the proposed control strategy under a fixed DER deployment.

- Bus 23 is located at the end of a radial feeder, downstream of PV generation at bus 19. This placement supports voltage stability by mitigating voltage rise from upstream PV injection.
- Bus 44 is situated near the midpoint of another main feeder and adjacent to a normally open tie-switch, allowing effective interaction with reconfiguration operations and system flexibility.
- Bus 47 is located at the far end of the third radial feeder. This feeder includes a long radial section ending at node 50, with limited local generation. Deploying ESSs here helps alleviate downstream voltage drops and loss concentration, especially during peak loads.

These placements ensure coverage across all major branches of the distribution network and create a challenging yet realistic scenario for testing the proposed algorithm. By fixing the ESS locations, the study maintains a clear focus on evaluating the control and optimization performance, without conflating results with the impacts of storage siting.

Table 2 provides an overview of the distributed energy resources. Figure 6 presents the time-series PV output over two consecutive days, with the first day utilized for training the agents and the second day reserved for testing and performance evaluation on unseen data.

Table 2. Specifications of Distributed Energy Resources.

Node	Type	Capacity	Control Capability
5	PV1		No
19	PV2	See Figure 6	No
63	PV3		No
23	ESS1	300 kW	Yes
44	ESS2	250 kW	Yes
47	ESS3	150 kW	Yes

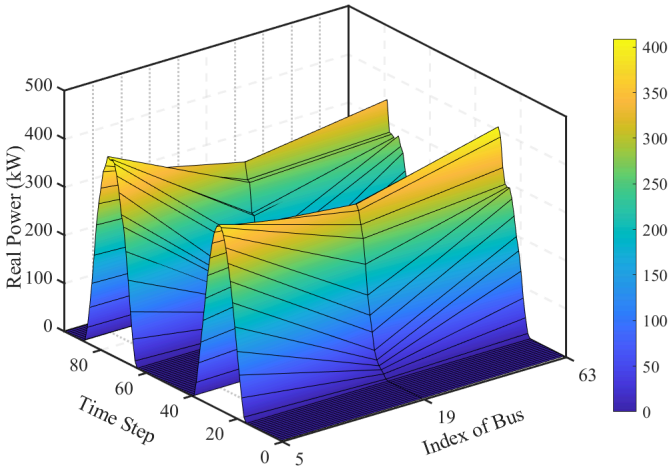


Figure 6. PV Generations under each time step.

The agents are trained to develop a full-day fault recovery policy. The NRA performs reconfiguration operations by controlling line switches, while the PSA optimizes ESS dispatch strategies to optimize restoration cost. Considering that PV generation is sampled at 30 min intervals, the decision-making process is discretized into 48 time steps, each corresponding to a half-hour period. The key hyperparameters used for training the agents are summarized in Table 3.

Table 3. Hyperparameter Settings.

Parameter	Value
Learning rate of Actor	1×10^{-4}
Learning rate of Actor	1×10^{-3}
Discount factor	0.9
Replay buffer size	5000
Batch size	64
GCN layers	2

4.2. Training Performance Comparison

The training task is designed around a fault occurring on line 5–6, which is located close to the main power supply and upstream of most tie-switches. The agents are trained to learn effective fault recovery strategies at any hour of the day. The following four algorithms are compared.

- DQN: designed for discrete action spaces, it enables only reconfiguration-based recovery through NRA, without continuous dispatch capability.
- DDPG: supports continuous action spaces; trains both NRA and PSA independently.
- MADDPG: facilitates centralized training with decentralized execution, enabling coordinated recovery through NRA and PSA.

- Proposed Method: builds upon MADDPG by incorporating GCNs for enhanced topology-aware decision-making and introducing a collaborative reward structure to improve agent coordination.

Figure 7 presents the training reward curves of the four algorithms, offering a comparative view of their convergence behaviors and effectiveness in learning fault recovery strategies within a distribution network environment. The x-axis represents the number of training episodes, while the y-axis denotes the mean reward (MR) obtained per episode. The shaded regions reflect the reward standard deviation (RSD), indicating the variability in performance.

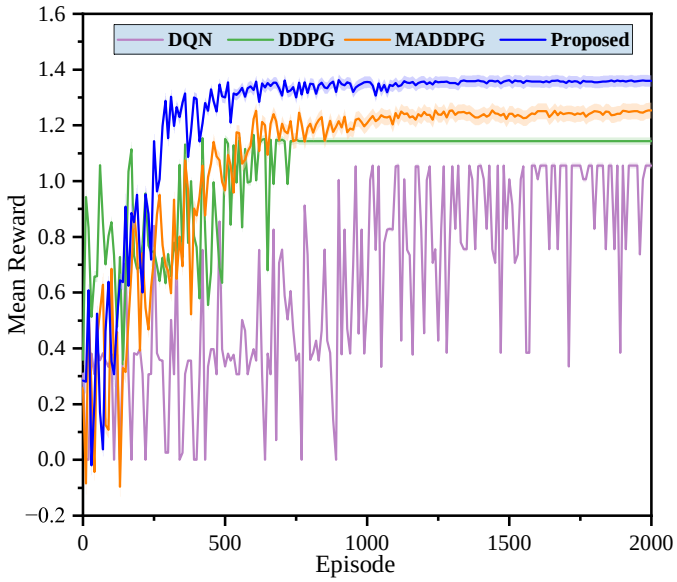


Figure 7. Comparison of training reward curves for different algorithms.

As shown in Figure 7, DQN exhibits limited learning capability. Constrained by its discrete action space, it can only perform network reconfiguration through NRA and lacks the ability to engage PSA for coordinated energy dispatch. This results in poor convergence, with the reward curve remaining unstable and at a relatively low level. DDPG demonstrates improved training stability over DQN, as it supports continuous action spaces, allowing both NRA and PSA to learn more expressive control policies. Specifically, NRA benefits from our proposed action decomposition strategy for effective line-switching, while PSA leverages the continuous space to optimize energy storage system (ESS) scheduling. Nevertheless, because each agent is trained independently without explicit inter-agent coordination, the learning process converges more slowly and levels off at a moderate reward. This suggests that a lack of cooperation may prevent the agents from discovering globally optimal policies. MADDPG significantly outperforms both DQN and DDPG by employing centralized training with decentralized execution. With access to global information during training, the agents can develop coordinated strategies, resulting in faster convergence and higher final rewards. These outcomes underscore the importance of agent collaboration in achieving efficient system-wide recovery. The proposed method achieves the best overall performance, with the fastest convergence rate and the highest final reward. By incorporating GCNs into the actor–critic architecture, the agents gain a deeper understanding of network topology and dynamic interactions. This topology-aware feature extraction substantially enhances their decision-making accuracy. Furthermore, the integration of a collaborative reward mechanism strengthens the cooperation between NRA and PSA, leading to more robust and cost-effective recovery strategies. The resulting

reward curve is notably smooth and stable, reflecting superior learning efficiency and strong generalization capabilities.

To further evaluate the individual contributions of key components in our proposed framework, we conduct an ablation study focusing on two main modules, the GCN used for spatial feature extraction, and the collaborative reward mechanism for multi-agent coordination. The following four variants are compared.

- Without GCN and Collaborative Reward: This baseline removes both the GCN and the collaborative reward mechanism. Agents are trained using FCNs and independent local rewards, which corresponds to the standard MADDPG algorithm.
- Without GCN: The graph-based feature extractor is replaced with a standard FCN, removing the ability to capture latent spatial features.
- Without Collaborative Reward: The collaborative reward mechanism is removed, and agents are trained by local rewards instead.
- Full Proposed Method: This is the complete version of our approach, incorporating both the GCN and the collaborative reward mechanism.

Figure 8 presents the MR curves during training for each ablation variant. As shown, the full version of our proposed method achieves the highest final MR and the fastest convergence, demonstrating the effectiveness of both the GCN-based feature extractor and the collaborative reward design. The variant without the collaborative reward ("Without Collaborative Reward") shows a clear decline in performance. Although the GCN enables agents to extract spatial features effectively, the absence of inter-agent reward alignment reduces coordination between the NRA and PSA, leading to suboptimal actions that may appear locally reasonable but fail to optimize system-level efficiency. In the variant without the GCN ("Without GCN"), performance degrades more noticeably, while the collaborative reward still promotes coordination, the lack of topology-aware representations limits each agent's understanding of network structure, resulting in slower convergence and lower final MR. The baseline configuration that removes both the GCN and the collaborative reward ("Without GCN and Collaborative Reward") performs the worst among all variants. This setup effectively degenerates into the standard MADDPG framework, where agents are trained independently using simple FCNs and local rewards. As a result, learning is slower, reward signals are more volatile, and final performance is substantially lower.

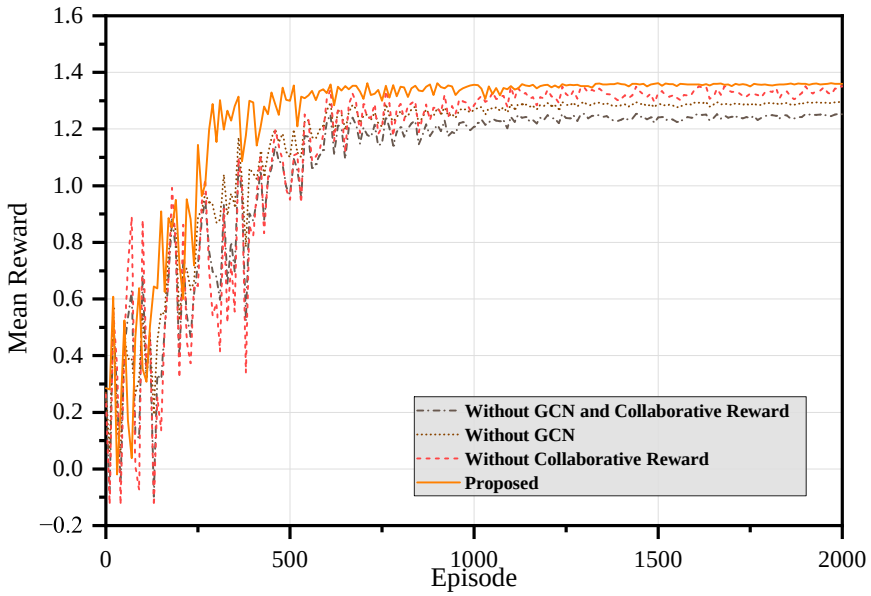


Figure 8. Comparison of training reward curves for each ablation variant.

These results highlight the complementary roles of GCN and the collaborative reward in enhancing both the representational power and the cooperative behavior of agents. The ablation study confirms that each component individually contributes to the performance gain, while their combination yields the most robust and efficient recovery strategy.

4.3. Testing and Validation

In our study, although the network model is based on a benchmark system, the power output profiles of DERs were derived from actual historical data. This hybrid setup enhances the realism of our simulations and closely mirrors the behavior of practical distribution systems with DERs. After completing training, we further validated the proposed method using a set of test scenarios in which the DER power profiles used during testing were different from those seen during training. The purpose of this validation is to assess the agents' ability to generalize the learned policies and to execute effective, time-sensitive fault recovery strategies on fault scenarios occurring at different times throughout a typical day.

To validate the effectiveness of the learned strategy, we introduce line outages at three representative time points, corresponding to morning, noon, and evening, when there is a significant difference in electricity demand and distributed generation. These test cases aim to verify the robustness and adaptability of the learned policies across a diverse set of operating conditions. In each scenario, a line fault is assumed to occur, prompting the agents to take immediate recovery actions. The NRA responds by altering the status of sectionalizing and tie switches to isolate the fault and restore service to as many loads as possible. Simultaneously, the PSA adjusts the dispatch of ESSs to supplement power supply and minimize recovery cost. Among the baseline methods compared during training, the standardized MADDPG is selected as the benchmark for testing.

Table 4 presents a statistical comparison between the proposed method and the standard MADDPG across three representative time-based testing scenarios. For each time point, uncertainty is introduced into the load demand, and 10 independent experiments are conducted to evaluate the robustness of both approaches. Results are reported as mean ± standard deviation, and paired *t*-tests are performed to assess the statistical significance of improvements over the MADDPG baseline.

Table 4. Statistical Comparison of Fault Recovery Performance.

Time	Method	Load Recovery (%)	Cost (p.u.)	Switches Opened	Switches Closed	*p*-Value *
08:00	Proposed	99.3 ± 0.4	0.42 ± 0.03	54–55, 14–15	50–59, 11–43	<0.01
	MADDPG	97.6 ± 0.6	0.49 ± 0.05	12–13, 55–56	11–43, 27–65	
12:00	Proposed	100.0 ± 0.0	0.33 ± 0.02	09–53	50–59, 65–27	<0.05
	MADDPG	99.1 ± 0.5	0.35 ± 0.04	54–55	15–46, 65–27	
18:00	Proposed	96.8 ± 0.3	0.45 ± 0.04	25–26, 64–65, 09–10	15–46, 50–59	<0.01
	MADDPG	94.2 ± 0.5	0.52 ± 0.06	54–55, 64–65	11–43, 50–59	

* *p*-values are obtained from two-sample *t*-tests comparing load recovery rates between the Proposed method and MADDPG.

At 08:00, both PV output and load demand are relatively low. Following a fault on line 5–6, which causes power loss to part of the downstream network, the NRA agent consistently responds by closing switches 50–59 and 11–43, while opening switches 54–55 and 14–15 to maintain voltage levels within acceptable bounds. The PSA's actions vary slightly across 10 runs due to fluctuations in load conditions. On average, the proposed method achieves a load recovery rate of 99.3% ± 0.4% and a recovery cost of 0.42 ± 0.03 p.u., both significantly better than MADDPG ($p < 0.01$).

At 12:00, PV generation reaches its peak while system load remains at a moderate level. The NRA agent exploits this abundance of renewable energy by closing switches 50–59 and 65–27 to redirect supply from PV-equipped nodes, while opening switch 09–53 to avoid network meshing that could introduce operational risks. Owing to the ample power supply, this scenario consistently achieves 100.0% ± 0.0% load restoration with the lowest recovery cost of 0.33 ± 0.02 p.u., which is again statistically superior to MADDPG ($p < 0.05$). The proposed method's advantage in recovery cost highlights the agents' ability to maximize solar energy utilization while minimizing reliance on energy storage systems. This scenario further demonstrates the effectiveness of coordinated decision-making between the NRA and PSA.

At 18:00, PV generation drops significantly due to reduced solar irradiance, while load demand peaks during the evening hours. Under these challenging conditions, the agents restore service by closing switches 15–46 and 50–59 to reconnect downstream loads with upstream sources, and by opening switches 25–26 and 64–65 to prevent voltage violations caused by excessive downstream demand. Additionally, switch 09–10 is opened to eliminate potential loops in the network topology. Despite the high-load, low-generation scenario, the proposed method achieves 96.8% ± 0.3% load recovery at a cost of 0.45 ± 0.04 p.u., demonstrating strong adaptability and resilience. Statistical tests confirm the superiority of the proposed method in both recovery rate and cost compared to MADDPG ($p < 0.01$).

These statistically significant improvements confirm that the incorporation of GCNs and collaborative rewards enables more accurate topology perception and better agent coordination, especially in complex operational conditions.

4.4. Practical Deployment and Engineering Feasibility

To assess the practical applicability of the proposed method, we discuss how it can be integrated into real-world distribution network operations. Although this study uses the PG&E 69-bus benchmark system, the method is designed with deployment feasibility in mind and does not rely on assumptions that are infeasible in modern power systems.

The proposed DRL-based framework serves as a decision-support tool rather than a fully autonomous control mechanism. In current utility environments, switching operations are typically authorized and executed by human operators. Our method complements this structure by offering real-time action recommendations that operators can evaluate and implement during restoration or reconfiguration events.

From an engineering perspective, the framework is lightweight and user-oriented.

- Input requirements include only standard operational data such as switch status, node connectivity, load levels, and fault locations, which are typically available through existing SCADA or DMS platforms.
- Output consists of switching actions or sequences, which can be interpreted and validated by engineers without the need for algorithmic knowledge.
- Inference speed is within real-time operational requirements once the model is trained, enabling timely response in fault scenarios.

One of the key advantages of the DRL framework lies in the decoupling of training and deployment. The computationally intensive training phase is performed offline and only once per system configuration or policy update. Table 5 summarizes the training time and average inference latency of the proposed method on an NVIDIA RTX 4070 GPU (NVIDIA Corporation, Santa Clara, CA, USA).

As shown in Table 5, the offline training is completed in less than 1 h, and the online inference latency is sufficiently low for real-time application in utility operations. Once trained, the DRL agent can generalize to unseen fault scenarios without the need for retraining, ensuring scalability across various DER/load conditions.

Table 5. Training Time and Inference Latency of the Proposed Method.

Metric	Value
Training Time	47.21 min
Average Inference Time	0.034 s per decision

To further improve accessibility and industry adoption, a graphical user interface module is planned for future development. This interface would allow system operators to interactively query the DRL agent, visualize switching options, and make informed decisions without requiring expertise in reinforcement learning.

Overall, this study demonstrates that the proposed method is not only technically effective but also operationally viable in modern distribution networks, supporting a smooth integration into existing utility workflows.

5. Conclusions

In this study, we proposed a graph-based multi-agent DRL framework for fault recovery in distribution networks. Experimental evaluations on the PG&E 69-bus system demonstrated that incorporating graph convolutional networks significantly enhanced agents' global situational awareness. The action decomposition strategy efficiently utilized the continuous action space, improving both decision-making speed and accuracy. Moreover, the collaborative reward mechanism, based on a global performance-oriented reward function, not only optimized individual agent behavior but also facilitated the emergence of globally optimal restoration strategies.

Despite its promising performance, the method has several limitations. First, the reliability of the DRL agent in real-world deployment depends on the quality and availability of real-time data, as well as the stability of communication infrastructure—factors not fully reflected in the simulation environment. Second, while the PG&E 69-bus system is widely used in academic research, it does not capture the full operational complexity of large-scale, dynamic utility networks.

Future work will focus on scaling the approach to more complex and dynamic systems, incorporating real-time uncertainty modeling, and addressing cybersecurity and interoperability concerns to support practical deployment in smart grid environments.

Author Contributions: Conceptualization, Y.L. and P.L.; methodology, Y.L.; software, Y.L.; validation, Y.L., P.L., and Y.W.; formal analysis, Y.L.; investigation, Y.L.; resources, P.L.; data curation, Y.L.; writing—original draft preparation, Y.L.; writing—review and editing, P.L.; visualization, Y.L.; supervision, Y.W.; project administration, P.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data are contained within the article.

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

Appendix A

For the reader's convenience, two notation tables are provided in this appendix. Table A1 lists the symbols and descriptions used in the context of POMDP, while Table A2 summarizes the notation relevant to DRL.

Table A1. Mathematical symbol definition table for POMDP.

Symbol	Definition
$\mathcal{S}$	State space
$\mathcal{O}$	Observation space
$\mathcal{A}$	Action space
$\rho(\cdot)$	The transition function
s_t	The state of distribution system at time step t
G_t	The graph representation of the distribution system at time step t
N, E	The sets of nodes and edges, respectively
Γ	The set of all controllable switches
$\mho$	The set of DERs
$\mathbf{A}$	The adjacency matrix
$\mathbf{F}$	The node feature matrix
$\mathbf{F}_t^{\text{NRA}}$, $\mathbf{F}_t^{\text{PSA}}$	The node feature matrices for NRA and PSA at time step t, respectively
$f_{i,t}^{\text{NRA}}$, $f_{i,t}^{\text{PSA}}$	The feature vectors of node i for NRA and PSA, respectively
$P_{i,t}^{\text{load}}$	The power demand of the load at node i and time step t
$P_{i,t}^{\text{der,max}}$, $P_{i,t}^{\text{der,min}}$	The maximum and minimum available power generation of the DER at node i and time step t, respectively
$\mathcal{P}_{i,t}$	The active power injection at node i and time step t
$\mathcal{V}_{i,t}$	The voltage magnitude at node i and time step t
$\mathcal{Y}_{i,t-1}$	A binary variable indicating whether the load at node i was restored at time step $t-1$
a_t^{NRA}, a_t^{PSA}	The actions for NRA and PSA at time step t, respectively
χ_t^{τ}	The probability of closing switch τ at the time step t
$\ell_t^{\wp}$	The power generation scaling factor of DER $\wp$ at time step t
$\mathcal{P}_t^{\wp}$	The actual power output of DER $\wp$
$P_t^{\wp,\text{max}}$	The maximum available generation capacity of DER $\wp$ at time step t
R_t	The discounted cumulative reward at time step t
γ	The discount factor
r_t^{nr}, r_t^{ps}	The rewards for network reconfiguration and power scheduling at time step t, respectively
P_t^{res}	The total load power restored after the action at time step t
$\lambda_{\wp}$	The cost coefficient associated with DER $\wp$
$P_{\wp,t}^{der}$	The actual power generation of DER $\wp$ at time step t
r_t^{global}	The global reward at time step t
α, β, κ	The coefficients used to balance the contributions of different reward components
$\wp_t$	The penalty function to enforce system constraints during fault recovery

Table A2. Mathematical symbol definition table for DRL.

Symbol	Definition
$\mathcal{H}^{(l)}$	The feature matrix at layer l
$W^{(l)}$	A trainable weight matrix
$\tilde{A}$	The normalized adjacency matrix
$\tilde{D}$	The degree matrix
σ	A non-linear activation function
π_i	The policy of agent i
θ_i	The parameters of actor network
o_i	The local observation of agent i at current time step
Q_i	The centralized Q-function at current time step
$\mathcal{D}$	The experience replay buffer, which stores past transitions for stable training
ψ_i	The parameters of critic network of agent i

Table A2. *Cont.*

y	The target Q-value at current time step
Q_i'	The estimated Q-value at the next time step
s'	The state at the next time step
$a_1', a_1', \ldots, a_n'$	The actions of agents at the next time step

References

1. Zhao, Y.; Lin, J.; Song, Y.; Xu, Y. A robust microgrid formation strategy for resilience enhancement of hydrogen penetrated active distribution networks. *IEEE Trans. Power Syst.* **2023**, *39*, 2735–2748. [CrossRef]
2. Ghosh, P.; De, M. Probabilistic quantification of distribution system resilience for an extreme event. *Int. Trans. Electr. Energy Syst.* **2022**, *2022*, 3838695. [CrossRef]
3. Cai, S.; Xie, Y.; Zhang, M.; Jin, X.; Wu, Q.; Guo, J. A stochastic sequential service restoration model for distribution systems considering microgrid interconnection. *IEEE Trans. Smart Grid* **2023**, *15*, 2396–2409. [CrossRef]
4. Zahraoui, Y.; Alhamrouni, I.; Basir Khan, M.R.; Mekhilef, S.; P. Hayes, B.; Rawa, M.; Ahmed, M. Self-healing strategy to enhance microgrid resilience during faults occurrence. *Int. Trans. Electr. Energy Syst.* **2021**, *31*, e13232. [CrossRef]
5. Liu, S.; Chen, C.; Jiang, Y.; Lin, Z.; Wang, H.; Waseem, M.; Wen, F. Bi-level coordinated power system restoration model considering the support of multiple flexible resources. *IEEE Trans. Power Syst.* **2022**, *38*, 1583–1595. [CrossRef]
6. Sabouhi, H.; Doroudi, A.; Fotuhi-Firuzabad, M.; Bashiri, M. Electricity distribution grids resilience enhancement by network reconfiguration. *Int. Trans. Electr. Energy Syst.* **2021**, *31*, e13047. [CrossRef]
7. Wang, L.; Chen, B.; Ye, Y.; Chongfuangprinya, P.; Yang, B.; Zhao, D.; Hong, T. Enhancing distribution system restoration with coordination of repair crew, electric vehicle, and renewable energy. *IEEE Trans. Smart Grid* **2024**, *15*, 3694–3705. [CrossRef]
8. Keshavarz Ziarani, H.; Hosseinian, S.H.; Fakharian, A. Providing a New Multiobjective Two-Layer Approach for Developing Service Restoration of a Smart Distribution System by Islanding of Faulty Area. *Int. Trans. Electr. Energy Syst.* **2024**, *2024*, 9687002. [CrossRef]
9. Choopani, K.; Hedayati, M.; Effatnejad, R. Self-healing optimization in active distribution network to improve reliability, and reduction losses, switching cost and load shedding. *Int. Trans. Electr. Energy Syst.* **2020**, *30*, e12348. [CrossRef]
10. Fan, B.; Liu, X.; Xiao, G.; Xu, Y.; Yang, X.; Wang, P. A Memory-Based Graph Reinforcement Learning Method for Critical Load Restoration With Uncertainties of Distributed Energy Resource. *IEEE Trans. Smart Grid* **2025**, *16*, 1706–1718. [CrossRef]
11. Liu, W.; Ding, F. Hierarchical distribution system adaptive restoration with diverse distributed energy resources. *IEEE Trans. Sustain. Energy* **2020**, *12*, 1347–1359. [CrossRef]
12. Taheri, B.; Safdarian, A.; Moeini-Aghtaie, M.; Lehtonen, M. Distribution system resilience enhancement via mobile emergency generators. *IEEE Trans. Power Deliv.* **2020**, *36*, 2308–2319. [CrossRef]
13. Alobaidi, A.H.; Fazlhashemi, S.S.; Khodayar, M.; Wang, J.; Khodayar, M.E. Distribution service restoration with renewable energy sources: A review. *IEEE Trans. Sustain. Energy* **2022**, *14*, 1151–1168. [CrossRef]
14. Sun, X.; Xie, H.; Bie, Z.; Li, G. Restoration of high-renewable-penetrated distribution systems considering uncertain repair workloads. *Csee J. Power Energy Syst.* **2022**, *11*, 150–162.
15. Nazemi, M.; Dehghanian, P.; Lu, X.; Chen, C. Uncertainty-aware deployment of mobile energy storage systems for distribution grid resilience. *IEEE Trans. Smart Grid* **2021**, *12*, 3200–3214. [CrossRef]
16. Zhao, J.; Wang, H.; Wu, Q.; Hatziargyriou, N.D.; Shen, F. Optimal generator start-up sequence for bulk system restoration with active distribution networks. *IEEE Trans. Power Syst.* **2020**, *36*, 2046–2057. [CrossRef]
17. Erenoğlu, A.K.; Erdinç, O. Real-time allocation of multi-mobile resources in integrated distribution and transportation systems for resilient electrical grid. *IEEE Trans. Power Deliv.* **2022**, *38*, 1108–1119. [CrossRef]
18. Li, G.; Yan, K.; Zhang, R.; Jiang, T.; Li, X.; Chen, H. Resilience-oriented distributed load restoration method for integrated power distribution and natural gas systems. *IEEE Trans. Sustain. Energy* **2021**, *13*, 341–352. [CrossRef]
19. Zhang, L.; Wang, C.; Liang, J.; Wu, M.; Zhang, B.; Tang, W. A coordinated restoration method of hybrid AC/DC distribution network for resilience enhancement. *IEEE Trans. Smart Grid* **2022**, *14*, 112–125. [CrossRef]
20. Long, Y.; Lu, Y.; Zhao, H.; Wu, R.; Bao, T.; Liu, J. Multilayer deep deterministic policy gradient for static safety and stability analysis of novel power systems. *Int. Trans. Electr. Energy Syst.* **2023**, *2023*, 4295384. [CrossRef]
21. Chai, R.; Niu, H.; Carrasco, J.; Arvin, F.; Yin, H.; Lennox, B. Design and experimental validation of deep reinforcement learning-based fast trajectory planning and control for mobile robot in unknown environment. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *35*, 5778–5792. [CrossRef] [PubMed]
22. Bian, R.; Jiang, X.; Zhao, G.; Liu, Y.; Dai, Z. A Scalable and Coordinated Energy Management for Electric Vehicles Based on Multiagent Reinforcement Learning Method. *Int. Trans. Electr. Energy Syst.* **2024**, *2024*, 7765710. [CrossRef]

23. Fan, B.; Liu, X.; Xiao, G.; Yang, X.; Chen, B.; Wang, P. Enhancing Adaptability of Restoration Strategy for Distribution Network: A Meta-Based Graph Reinforcement Learning Approach. *IEEE Internet Things J.* **2024**, *11*, 25440–25453. [CrossRef]
24. Zhao, J.; Li, F.; Sun, H.; Zhang, Q.; Shuai, H. Self-attention generative adversarial network enhanced learning method for resilient defense of networked microgrids against sequential events. *IEEE Trans. Power Syst.* **2022**, *38*, 4369–4380. [CrossRef]
25. Nikkhah, M.H.; Lotfi, H.; Samadi, M.; Hajiabadi, M.E. Energy hub management considering demand response, distributed generation, and electric vehicle charging station. *Int. Trans. Electr. Energy Syst.* **2023**, *2023*, 9042957. [CrossRef]
26. Ye, Y.; Wang, H.; Chen, P.; Tang, Y.; Strbac, G. Safe deep reinforcement learning for microgrid energy management in distribution networks with leveraged spatial–temporal perception. *IEEE Trans. Smart Grid* **2023**, *14*, 3759–3775. [CrossRef]
27. Alatawi, M.N. Optimization of Home Energy Management Systems in smart cities using bacterial foraging algorithm and deep reinforcement learning for Enhanced Renewable Energy Integration. *Int. Trans. Electr. Energy Syst.* **2024**, *2024*, 2194986. [CrossRef]
28. Bedoya, J.C.; Wang, Y.; Liu, C.C. Distribution system resilience under asynchronous information using deep reinforcement learning. *IEEE Trans. Power Syst.* **2021**, *36*, 4235–4245. [CrossRef]
29. Zhang, Y.; Qiu, F.; Hong, T.; Wang, Z.; Li, F. Hybrid imitation learning for real-time service restoration in resilient distribution systems. *IEEE Trans. Ind. Inform.* **2021**, *18*, 2089–2099. [CrossRef]
30. Du, Y.; Wu, D. Deep reinforcement learning from demonstrations to assist service restoration in islanded microgrids. *IEEE Trans. Sustain. Energy* **2022**, *13*, 1062–1072. [CrossRef]
31. Mocanu, E.; Mocanu, D.C.; Nguyen, P.H.; Liotta, A.; Webber, M.E.; Gibescu, M.; Slootweg, J.G. On-line building energy optimization using deep reinforcement learning. *IEEE Trans. Smart Grid* **2018**, *10*, 3698–3708. [CrossRef]
32. Fan, B.; Liu, X.; Xiao, G.; Kang, Y.; Wang, D.; Wang, P. Attention-Based Multiagent Graph Reinforcement Learning for Service Restoration. *IEEE Trans. Artif. Intell.* **2024**, *5*, 2163–2178. [CrossRef]
33. Thurner, L.; Scheidler, A.; Schäfer, F.; Menke, J.H.; Dollichon, J.; Meier, F.; Meinecke, S.; Braun, M. Pandapower—An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems. *IEEE Trans. Power Syst.* **2018**, *33*, 6510–6521. [CrossRef]
34. Chen, X.; Wu, W.; Zhang, B. Robust restoration method for active distribution networks. *IEEE Trans. Power Syst.* **2015**, *31*, 4005–4015. [CrossRef]
35. Chen, K.; Wu, W.; Zhang, B.; Sun, H. Robust restoration decision-making model for distribution networks based on information gap decision theory. *IEEE Trans. Smart Grid* **2014**, *6*, 587–597. [CrossRef]
36. Guan, W.; Tan, Y.; Zhang, H.; Song, J. Distribution system feeder reconfiguration considering different model of DG sources. *Int. J. Electr. Power Energy Syst.* **2015**, *68*, 210–221. [CrossRef]
37. Luo, W.; Li, C.; Ju, S. Multisource cooperative restoration strategy for distribution system considering islanding integration. *Power Syst. Technol.* **2022**, *46*, 1485–1495.
38. Liu, K.y.; Sheng, W.; Liu, Y.; Meng, X. A network reconfiguration method considering data uncertainties in smart distribution networks. *Energies* **2017**, *10*, 618. [CrossRef]
39. Chakravorty, M.; Das, D. Voltage stability analysis of radial distribution networks. *Int. J. Electr. Power Energy Syst.* **2001**, *23*, 129–135. [CrossRef]

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