Robotics and Computer–Integrated Manufacturing 89 (2024) 102775

Contents lists available at ScienceDirect

Robotics and Computer-Integrated Manufacturing

journal homepage: www.elsevier.com/locate/rcim

Partially observable deep reinforcement learning for multi-agent strategy optimization of human-robot collaborative disassembly: A case of retired electric vehicle battery

Jiaxu Gao [a], Guoxian Wang [a], Jinhua Xiao [a,*], Pai Zheng [b], Eujin Pei [c]

[a] School of Transportation and Logistics Engineering, Wuhan University of Technology, Wuhan 430063, China
[b] Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hong Kong, China
[c] Brunel University London, College of Engineering, Design & Physical Sciences, UB8 3PH, United Kingdom

ARTICLE INFO

Keywords:
Human-robot collaboration
Multi-agent deep reinforcement learning
Partially observable markov decision process
Disassembling sequence planning

ABSTRACT

The burgeoning electric vehicle (EV) industry has precipitated a commensurate surge in the consumption of EV batteries, which are currently labor-intensive and inefficient for the recycling and disassembly of EV batteries. However, it is a potential trend to enhance the efficacy and safety of the disassembly of EV batteries based on human-robot collaboration (HRC) method. Because of the uncertainty of retired EV battery disassembly and the inefficiency of the existing disassembling sequence, it is difficult to be fully accomplish through HRC disassembly. The collaborative disassembly of EV batteries by humans and robots can be conceptualized as agents engaging with and learning from the environment, and modeled as a multi-agent Markov game process. This paper aims to address the challenge of HRC in the disassembly of EV batteries by recognizing the dual attributes of partial observability and non-smoothness in the suitable disassembly scenario. A partially observable multi-agent reinforcement learning environment is constructed, incorporating the structural aspects of the EV battery and the disassembly task. The framework is extended to the QMIX-HRC algorithm on the QMIX architecture (as a value-based multi-agent deep reinforcement learning algorithm), specifically designed to tackle the sequence problem in human-robot collaborative disassembly of EV batteries. The optimization results would yield a task sequence to offer maximal global co-benefit during the exploration iteration, facilitating a reduction in labor costs and an enhancement of co-efficiency. The viability of the QMIX-HRC disassembly strategy would be verified through the eventual disassembly sequence of a simulated battery pack through a real human-robot collaborative disassembly station.

1. Introduction

With the rapid expansion of the global new energy market, the production and utilization of EV batteries, being an essential energy storage component of new energy equipment, has witnessed a substantial annual upsurge [1]. The capacity of EV batteries used over a specific timeframe under normal operating conditions will gradually decline [2]. Upon reaching a capacity reduction of less than 80 % of the initial capacity, a determination will be made to undertake varying degrees of echelon utilization or proceed with disassembly and recycling activities based on the specific conditions of EV batteries [3]. In order to avoid to potential environmental pollution, it is imperative to standardize the large-scale recycling if disassembled EV batteries [4]. Compared with other retired electrical and electronic equipment, disassembled EV batteries can exhibit different levels of hazardousness, intricate and diverse structures, as well as considerable disparities under possible damage [5]. The design and implementation of exemplary recycling and disassembling programs are imperative for addressing the structural complexities, material composition, and recycling characteristics specific to EV batteries [6].

However, it is important to note that the disassembly of a power battery cannot be considered a simple reversal of the assembly process. Therefore, the practical operability of the disassembly tasks must be considered in disassembly sequence planning of the EV battery [7]. As shown in Fig. 1, the EV battery disassembly process can be classified into two major stages: The disassembly of the battery pack and the

* Corresponding author.
E-mail address: xiaojinh@whut.edu.cn (J. Xiao).

https://doi.org/10.1016/j.rcim.2024.102775
Received 1 November 2023; Received in revised form 6 March 2024; Accepted 12 April 2024
Available online 20 April 2024

disassembly of the battery module [8]. In industrial practice, the battery cell represents the final component subjected to meticulous disassembly [9]. In order to implement echelon utilization of modules and battery cells, it is necessary to comply with the standards to maximize economic benefits and environmental conservation [10]. The disassembly of EV batteries are predominantly reliant on the disassembly of human operation ways [11]. The key barriers encountered during the disassembly process can be categorized into the following two categories: EV battery recycling quality and substantial variations of EV battery structures [12]. During the normal use, recycling and transportation of EV batteries, the potentially uneven quality will yield during the recycling of power batteries. These challenges can impede the ability to achieve proper disassembly of EV packs within a standardized and intelligent remanufacturing system [13].

The contemporary tendency in industrial intelligence pertains to the utilization of robotic systems to assist human operators with task-intensive and labor-intensive work [14]. Consequently, endeavors that entail flexibility but do not necessitate a significant labor component are still preferred to be performed by human operators, while relatively loaded or repetitive tasks are more suitable for assignment to robotic systems [15,16]. By considering the optimization of disassembly task allocation, HRC is progressively gaining prominence as the prevailing approach in the intelligent manufacturing field. In the practical HRC scenarios, a significant limitation arises from the restricted range of the robot's sensors, which restricts the observability to only a partial representation of the environment rather than the complete environment information [17]. During task execution, human tend to concentrate on present operations, resulting in a blind field of vision that limits their ability to perceive the complete information about the environment. To summarize, the collaborative disassembly scenario involving human-robot interaction is characterized by inherent limitations in observability. The collaborative disassembly environment involving humans and robots serves as a complex setting that is jointly explored by multiple agents. The actions of both humans and robots dynamically modify the state of the environment in real-time, leading to non-smooth characteristics of the overall environment. As a result, the human-robot collaborative environment exhibits the dual features of partial observability and non-smoothness. These features will affect the efficiency of the optimization algorithm and make it unable to meet the actual requirements.

This study examines the task allocation and sequencing of EV batteries during the disassembly process in a HRC setting. We propose a reinforcement learning environment with partial observability to provide a realistic basis for agent exploration. The construction of the reinforcement learning environment depends on the structural analysis and task classification results of the EV battery to be disassembled. Traditional reinforcement learning algorithms lack consideration of partial observability, rendering them poorly suited for the current HRC environment. In this study, we apply the dual-agent QMIX algorithm to determine the optimal strategy for humans and robots to collaboratively complete the battery disassembly process. Our approach aims to address the requirements of battery disassembly while enhancing the efficiency of robot-assisted operations and reducing physical fatigue for human workers. The structure of this paper is as follows: Section 2 provides an overview of the current challenges associated with EV battery disassembly and the specific difficulties encountered in achieving effective HRC disassembly. Section 3 focuses on the optimization methods of the existing EV battery disassembly strategies and the application of reinforcement learning in HRC scenarios. Section 4 is the construction of the reinforcement learning environment and algorithmic computing process; Section 5 demonstrates the battery disassembly experiment completed using optimization strategies on a real HRC workstation using the optimization strategy to achieve the battery disassembly experiment. Finally Section 6 provides the conclusion and future work.

2. Related work

Recent research efforts have primarily focused on several key aspects pertaining to the disassembly of EV batteries. These include the environmental sustainability and cost-effectiveness of the disassembly process, ensuring the safety of disassembly operations, as well as optimizing the disassembly strategy. The central objective of these studies is to establish industry standards that can effectively standardize the recycling and remanufacturing process of EV batteries. This chapter provides a comprehensive summary of the current progress in three key research areas: research on disassembly strategies of EV batteries, research on disassembly sequence planning, and research on reinforcement learning for human-robot collaborative scenarios.

2.1. Disassembly strategy of EV battery

The research on disassembly strategies for EV batteries primarily emphasizes the optimization of the disassembly sequence within the context of intelligent industry, where collaborative efforts between robots and humans are leveraged. The intention is to address the limitations associated with the conventional manual disassembly process, which is inherently inefficient and burdensome in terms of recycling costs. By employing collaborative disassembly strategies involving both humans and robots, the aim is to enhance the economic feasibility of EV battery recycling. Rastegarpanah et al. [18] proposed using a behavioral tree model to control the EV battery disassembly task so that the robot can automate the identification of sorted battery parts. Unlike the mainstream adoption of vision systems for target detection, this system uses LiDAR for robot path planning. Xiao et al. [19] established an information interaction system in EV battery recycling and disassembly based on the STEP standard, improving information-sharing efficiency in the disassembly process. The division of labor between humans and robots is further discussed, emphasizing the critical decision-making function of humans in human-robot collaborative scenarios. Gloser-Chahoud et al. [20] addressed the inefficiency and waste of

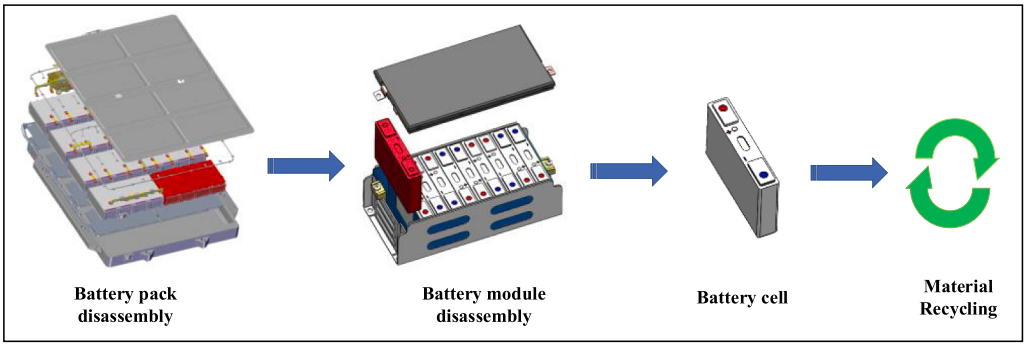

Fig. 1. EV Battery disassembly and recycling process.

resources in the current manual disassembly of batteries and discussed establishing a highly automated production line with intelligent industrial technology to achieve flexible disassembly. This initiative improves the closed-loop circularity of the entire recycling system, reduces the cost of the actual process, and facilitates the recycling and remanufacturing of different parts. Hellmuth et al. [21] proposed multi-dimensional evaluation criteria to assess the automation potential of each step in battery disassembly. Disassembly operations that reach a specific automation score will be assigned preferentially to robotic procedures, while humans will be reserved for functions such as lifting that the evaluated disassembly task can be used to separate different EV battery types more scientifically and efficiently. The current mainstream solution to the battery disassembly problem is to use computer vision technology for target detection and localization. Martin et al. [22] used computer vision techniques to identify and localize targets for disassembly. In the actual disassembly operation, there is a need for human intervention in the case of high-precision vision system-based robots in the face of complex situations that still do not have sufficient adaptability. The traditional disassembly strategy planning system with a CAD model as the primary reference does not apply to the EV battery recycling scenario because of actual and model mismatch characteristics. Li et al. [23] presented active detection to locate the screws at a specific angle in response to the poor practical results of previous battery disassembly bolt identification. This solves the problem that simulated planar datasets cannot be accurately applied to irregular screws in actual production processes and improves the efficiency and stability of robotic disassembly of batteries. Wegener et al. [24] proposed the establishment of a robot-assisted workstation dominated by human operation. When dealing with the disassembly of EV batteries, the operational tasks that are more complex and require better decision-making capabilities are still performed by humans. The robot's role as a support should share simple and repetitive tasks and reduce operator fatigue. By observing the inconvenience of human operation in the battery recycling process, Liu et al. [25] proposed using a robot equipped with a high-rotation cutting wheel to cut the battery casing in a fixed position. The original operation mode is adjusted to human beings are responsible for handling, positioning, sorting battery components, etc., and the robot is responsible for cutting. This human-robot collaborative approach improves the efficiency of the entire process and maximizes the protection of people from the dangers of short-circuiting the cut cells. Zhang et al. [26] proposed the migration of efficient tools from other industry fields to EV battery disassembly applications, including special robot actuators and special tooling for human. It further suggests that battery recycling should combine semi-automated mechanical disassembly of EV batteries and chemical recycling methods to meet the recycling industry's current requirements.

2.2. Disassembling sequence planning research

Alfaro-Algaba et al. [27] used the Audi A3 EV battery as a research object, pursuing the balance between economic cost and environmental protection capability in disassembly. The entire recycling strategy is selected based on the system's assessment of the battery state and the recycling needs of each component. In different states of the battery, the process and sequence of disassembly will be adjusted accordingly to ensure that the economic benefits of recycling are maximized. Tan et al. [28] proposed a modified method based on the robotic arm for human-robot collaborative disassembly workstation and special disassembly tools, which further analyzes and optimizes the disassembly process from multiple perspectives, including safety and cost, pointing out that an improved automated disassembly framework is suitable for current battery disassembly scenarios. Sabri et al. [29] designed a human-robot collaborative workstation for EV battery disassembly, combining special disassembly equipment and an adaptive planner. On this basis, the optimal disassembly strategy can be defined with the optimal operation sequence, the optimal operation depth, and the optimal recycling economic strategy, which suggests that optimizing the disassembly strategy can also be fed back into the design and manufacturing process of EV battery. Wang et al. [30] introduced a novel approach in the context of optimizing the disassembly sequence of products with interlocking structures. Specifically, this approach entails the development of a new assembly matrix expression method that effectively captures the intricate relationships among the parts to be disassembled. This method has been designed to enhance the accuracy with which the structure of these parts is represented during the conversion of the primary solid model into a virtual structure. In addition to the factors typically addressed in traditional sequence optimization, Kheder et al. [31] incorporated an assessment of the maintainability of wearing parts. This was conducted as part of the disassembly sequence planning process using the ant colony algorithm. A comparative analysis was then carried out, contrasting the effectiveness of this method with that of other heuristic algorithms, thus substantiating its efficacy. Additionally, Laili et al. [32] identified that the remanufacturing disassembly process is susceptible to encountering unique circumstances that hinder full automation. To address this issue, they proposed the utilization of a backup operation method, allowing for flexible adjustments to the robot's task sequence. Furthermore, the researcher suggested employing the dual-selection multi-objective evolutionary algorithm to optimize the model, thus enhancing the robotic system's capability to execute automatic disassembly and ultimately improving overall performance.

2.3. Reinforcement learning for human-robot collaborative applications

The applications of reinforcement learning in HRC are categorized into security protection, policy selection, and path planning. The optimization of a robot's motion trajectory is a classical application of reinforcement learning. Essentially, theoretical task planning and action planning are processes of reinforcement learning to seek optimal strategies for an agent. HRC method is generally regarded as a reinforcement learning interaction between two agents to support a co-trained environment and a multi-intelligent agent. During training, the robotic agent will gradually learn how to assist the human better to accomplish complex tasks.

As reviewed many literatures, Yu et al. [33] visualized the task execution in a tessellated grid model with assembly trees and constraint relationships fully mapped in the 2D world. The article uses the deep Q-network (DQN) algorithm to iteratively solve task allocation and sequence optimization problems in a checkerboard grid environment. It compares the algorithm's solving ability in various multi-agent scenarios, i.e., human-robot interaction between humans and several robots. Tsiakas et al. [34] proposed a Human-centric Cyber-Physical Systems (HCPS) approach. The method focuses on operators and uses reinforcement learning algorithms and multi-sensor data from actual production to train operators' behavior and reduce human safety risks of real output. Liu et al. [35] used One-Hot encoding to compress the observed state information of the agent into the convolutional neural network of the DQN. It increases the algorithm's convergence speed significantly compared to traditional methods and is suitable for task-level decision-making within the industrial domain. Zhang et al. [36] proposed the Actor-Critic architecture in reinforcement learning to build dual goal and evaluation network models in a human-robot collaborative environment. The article improves the algorithm's robustness by artificially adding noise to the actions of an agent representing a human being, fitting the instability that characterizes the presence of a human being in a natural environment. In actual experiments, the optimal sequence generated by the algorithm will be further visualized to guide the operator to complete the in-order operation directly. Roveda et al. [37] proposed Model-Based Reinforcement Learning (MBRL) methods to reduce human fatigue and safety hazards. The learned model will be used in MPC+CEM (Model Prediction Controller for Cross-Entropy Method) to minimize human effort further.

In the experimental part of the article, it is discussed that human-robot collaborative planning aims to enable the robot to cooperate with assistance according to the human movement flexibly. Agarwal et al. [38] optimized the robot control part directly based on DQN. The optimization of this article focuses on improving the adaptability of the robot manipulation system and reducing the impact of human interference on the robot, thus facilitating the collaborative mode of human-robot interaction. Shafti et al. [39] proposed human-robot collaborative tasks in the real world that can only be accomplished by humans and robots working together. This is not only a human-robot collaborative operation problem in the industry but also extends to other fields that require robot assistance, such as rehabilitation robots. Reinforcement learning can drive humans and robots to relate to each other, with robots better adapting their strategies to human behavioral states. Khamassi et al. [40] proposed active exploratory reinforcement learning for environments with change. The algorithm can be based on discrete action spaces for continuous parameterization and extended to long sequence tasks and continuous action spaces. A summary of the research applied to human-robot collaborative disassembly sequence optimization is shown in the Table 1.

Through the existing research, HRC method is a better candidate solution for the disassembly of EV batteries. Robots with different end-effectors can assist people in completing various processes in disassembling EV batteries, including more dangerous cutting operations, etc., which can significantly improve overall work efficiency. Reinforcement learning has been widely used to solve decision-planning for robots in various fields, especially for some dynamic planning problems with good adaptability, and the current applications of reinforcement learning algorithms for planning in HRC scenarios are predominantly in part assembly. Compared with the specific assembly task, the disassembly task has greater uncertainty and execution difficulty. HRC disassembly of EV batteries or other end of life (EOL) products for task decision analysis in the current research mainly focuses on some heuristic algorithms in conjunction with the study of disassembly hybrid graphs. Reinforcement learning algorithms are more flexible than heuristic algorithms and have a good effect in solving complex environments, etc. The application of reinforcement learning algorithms to HRC disassembly of decision-making and planning can be beneficial in solving the practical issues in the EV battery recycling.

3. Multi-agent reinforcement learning

3.1. Principles of reinforcement learning

Reinforcement learning is an end-to-end machine learning methodology that explores the intricate interplay between immediate decision-making and long-term rewards, aiming to identify the most optimal course of action at each state through iterative training. This process is accomplished through the agent's interaction with the environment to achieve a set goal. It is common in reinforcement learning to presume that the environment explored by the agent is endowed with Markov properties, such that the dynamics of the environment can be effectively represented in the form of a Markov Decision Process (MDP). The Markov property entails that the state of a system at a given time is solely determined by the state of the preceding moment and remains independent of the states of earlier moments, thus eliminating any influence from the historical system states. A model consistent with Markov property can constitute a Markov decision process, usually represented as a tuple $\{S,A,R,\gamma,P\}$. S is the set of state space; A is the set of action space; R is the reward function, γ is the discount factor, which takes the value between 0 and 1, and its value indicates the importance of the whole decision-making process on the long-term gain and the current gain; P is the state transfer function, whose content is the probability of transferring between two states. The agent's strategy is represented using π.

$$\pi(a|s) = P(A_t = a|S_t = s) \tag{1}$$

The iterative process is as follows: The agent senses the state of the current environment S_t, makes an action decision A for the environment, and gets the new environment state S_{t+1} and the benefit R_t from the action. The agent selects an action strategy which, based on the updated environment, maximizes its overall gain. There are two main types of iterative methods commonly used in reinforcement learning: strategy-based iteration and value-based iteration. The central iterative of value-based algorithm formula can be mathematically expressed as:

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha\left[R_t + \gamma \max_a Q(s_{t+1},a) - Q(s_t,a_t)\right] \tag{2}$$

In the formula, $Q(s,a)$ denotes action value function, α denotes learning rate. All action Q-values are stored in a table as a matrix, which is usually impossible when faced with a high-dimensional state environment. This would employ a neural network to fit the function Q using the method of function approximation, expressed as $Q_w(s,a)$. This is the core concept of deep reinforcement learning (DRL), which stands for algorithms such as DQN. At present, many studies on the application of reinforcement learning in sequence optimization use DQN algorithms. Nevertheless, DQN may not be entirely suitable for practical industrial applications due to two main reasons. Firstly, the empirical replay buffer area in DQN has a limited capacity to store data, which may not be sufficient for complex real-world scenarios with a large amount of data. Secondly, DQN require complete knowledge of the agent's total state space, which may be challenging to obtain in practical industrial settings.

3.2. Partial observability

In real-world industrial production settings, agents, such as operators or robots, frequently encounter observational limitations. For instance, robots may not have the ability to perceive the precise positions of individuals in all areas, and humans may have blind spots in their field of vision. In this context, POMDP is more appropriate than MDP, which accounts for the inherent partial observability in the system, allowing for more realistic representation and decision-making in situations where complete information about the system state is not available. As shown in Fig. 2, unlike the MDP structure, POMDP uses observation O to express the state obtained by an agent after observing the environment. The POMDP is usually represented by a six-tuple set (S,A,R,P,γ,O). The agent makes a corresponding action decision each time it gets an observation, and the observation conforms to a probability distribution $o_t \sim O(s_t)$. Coping with POMDP, the traditional DRL network does not work well because the conventional model fits the Q-function $Q(s,a|\theta)$, and the Q-function to be equipped in the POMDP is $Q(o,a|\theta)$. To address this challenge, recurrent neural networks (RNN) network architecture

Table 1
The reviews on HRC disassembling sequence optimization.

Literature	Method	Disassembly objects
Li et al. [41]	BA (genetic algorithm and artificial bee colony algorithm)	gear pump
Xu et al. [42]	MDBA-Pareto(modified discrete Bees algorithm based on Pareto)	computer case
Guo et al. [43]	multi-layer chromosome coding method	automobile engine
Liao et al. [44]	multi-attribute utility function	computer
Chu et al. [45]	HPSO_QL (hybrid particle swarm optimization with the Q-learning algorithm)	EV battery
Fang et al. [46]	MO-MFO (multi-objective multi-fidelity optimisation)	product
Allagui et al. [47]	reinforcement learning	multi-axis system

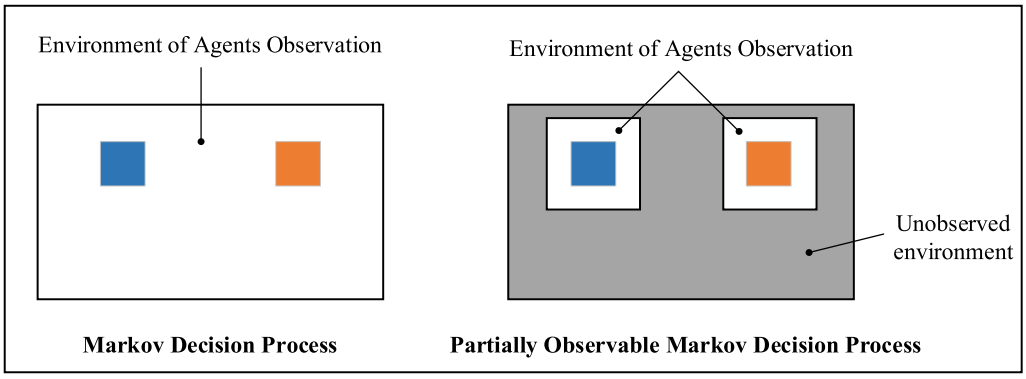

Fig. 2. Comparison of MDP and POMDP environments.

can effectively handle the POMDP.

Compared to the traditional network architecture, the DRL algorithm with RNN has more adaptability, which means that the algorithm has deep recurrent Q-network (DRQN). DRQN replaces the fully connected layer with an Long Short-Term Memory (LSTM) network in the RNN, and the output of the whole network is still each action corresponding to the $Q(s,a)$. The LSTM update process necessitates the input of a segment of observations within a continuous state, along with the associated action rewards.

3.3. Multi-agent game approach

When the number of agents is not fixed and multiple agents are involved, the field of multi-agent reinforcement learning (MARL) architectures emerges. With an increasing number of agents, there is a need to re-evaluate factors such as communication between agents, environment dynamics, and the determination of overall strategic reward values. In a multi-agent system, the relationship between the agents can also be referred to as a Markov game. The game system can be described as$\{n,S,A_1\cdots A_n,P,\gamma,R_1\cdots R_n\}$. Denotes the number of agents in the system; S represents the state of the entire system; P is the state transfer function; A_n and R_n defines the nth agent's action strategy and reward. In a Markov game, each agent formulates its respective strategy by assimilating information from state observations and making predictions about the behaviors of other agents. The types of games can be broadly categorized into zero-sum games under perfect competition, perfect collaboration, and mixed competition and collaboration. The HRC scenario can be classified as a holistic collaborative state, as it necessitates the joint efforts of both entities in order to successfully complete a task. The overall benefit derived from this scenario is contingent upon the cumulative benefits contributed by each agent involved. The current mainstream multi-agent reinforcement learning algorithms include value decomposition networks(VDN), QMIX and multi-agent deep deterministic policy gradient (MADDPG).

The QMIX network architecture is composed of Rashid et al. [48] based on the VDN algorithm, which is particularly applicable for fully cooperative multi-agent environments due to its distinct methodology of computing the value function. The QMIX architecture combines the Q-values of multiple agents by leveraging a single neural network, ultimately yielding the global aggregate Q-value. Above the specific network structure, QMIX can be categorized into mixing network (evaluation network, hybrid network) and hyper network for each Agent. The QMIX architecture have two mixing networks, one for the regular training network and the other for the target network, where the target network uses the parameters that have yet to be updated in time in the training network to improve the stability of the network. The evaluation network serves the purpose of generating individual Q-values, adopting a structure akin to that of the DRQN network. On the other hand, the hybrid network incorporates the Q-values from all agents to perform calculations. Furthermore, the super-network takes the system state S as input and produces a non-negative weight parameter for the hybrid network, ensuring the enforcement of monotonicity. QMIX considerd the global state as an additional reference when calculating the global Q-value to be effective in complex environments. The fundamental concept underlying QMIX aims to discover fully decentralized joint policies while ensuring consistency. To achieve this, it becomes imperative to ensure that the global Q-value aligns with the outcome of selecting the action with the maximum value. The formula is expressed as follows in the human-robot collaborative scenario:

$$\underset{A_u}{\operatorname{argmax}} Q_{tot}(\tau,A_u)=\left\{\begin{array}{l}\operatorname{argmax}_{A_R} Q_1(\tau_R,A_R)\\ \operatorname{argmax}_{A_H} Q_2(\tau_H,A_H)\end{array}\right\} \tag{3}$$

In the formula, τ denotes the observation.Q_{tot} denotes the total action value. The QMIX network undergoes updates following a methodology similar to the DQN update. The parameters are updated by computing the temporal difference (TD) error between the evaluation network and the target network:

$$TDerror = Q_{tot}(evaluate) - \gamma Q_{tot}(\text{target}) \tag{4}$$

Where Q_{tot}(evaluate) is calculated by the evaluate net, Q_{tot}(target) is calculated by the target net.

3.4. QMIX-HRC framework

The QMIX paradigm is extended to the human-robot collaboration scenario, namely QMIX-HRC. The structure of QMIX-HRC as shown in Fig. 3. In this diagram, Q_H and Q_R represent the action value of agent-human and agent-robot. It has two special modules: experience reply and target network. The primary concept behind experience replay involves the creation of an additional region where the current state, action, reward, and the subsequent state are stored at each moment in time. The target network is proposed to solve the instability of neural network training. The QMIX-HRC adopted paradigm for solving a multi-agent reinforcement learning environment is Centralized Training with Decentralized Execution (CTDE).

In the disassembly environment for HRC, the actions can be regarded as discrete, as the smallest unit of action corresponds to the execution of a single task. In this particular environment, the collaborative disassembly workstation involving both humans and robots can be conceptualized as a system featuring two distinct agents: the operator (human) and the robot. The two agents are required to autonomously allocate disassembly tasks towards a common disassembly objective. They aim to identify an optimal disassembly strategy that minimizes the overall disassembly time and reduces the level of complexity associated with disassembly. This objective is achieved through repeated observations and learning within the given environment. Subsequently, based on their respective optimal strategies, a comprehensive disassembly plan is generated. The foremost consideration in this process is evaluating the

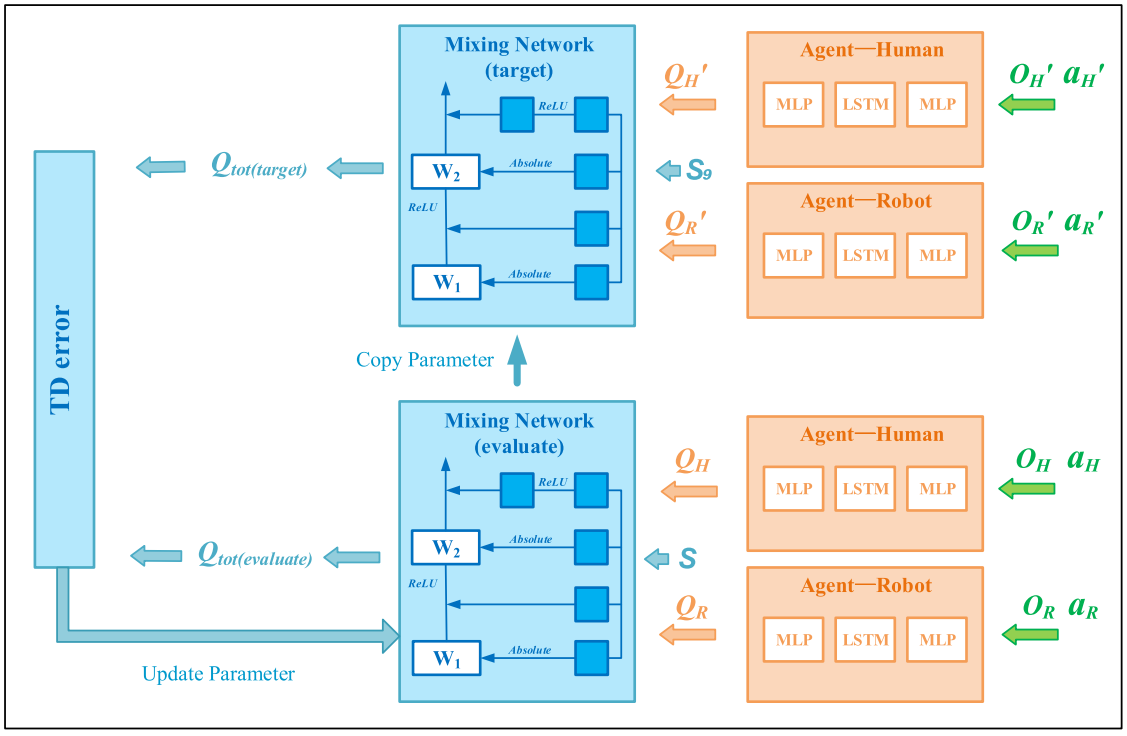

Fig. 3. The QMIX-HRC structure.

practical viability of the operation and anticipating potential challenges that may arise in real working conditions. This assessment is performed by incorporating constraints imposed by the physical space. In this regard, the standard CAD 3D model serves as a reference, enabling the initial formulation of a corresponding constraint matrix. Nonetheless, it is important to acknowledge that discrepancies are likely to exist between the actual battery packs to be processed and the standard model. As a consequence, such variations render the original constraint matrix ineffective and inapplicable. As a result, the process route must be modified accordingly, thus influencing the predetermined disassembly strategy. In order to address this challenge, it is necessary for the optimization model of the EV battery disassembly strategy to possess dynamic planning capabilities and be adaptable to the intricate conditions inherent to the artifact. The whole disassembly strategy framework is shown in Fig. 4.

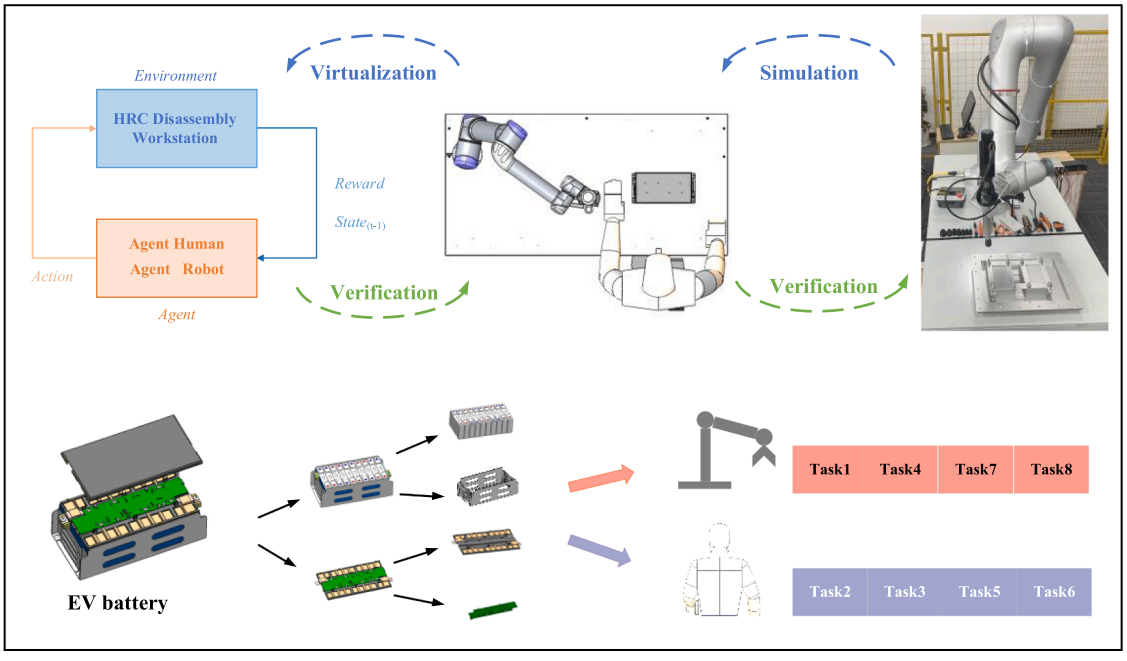

Fig. 4. Strategy framework for human-robot collaborative disassembly of batteries.

4. Modeling and simulations

4.1. Modeling environmental design

The reinforcement learning environment proposed in this study is a redeveloped and designed version of the well-established Gridworld environment, specifically tailored to accommodate the requirements of the research. The environment is divided into two distinct areas: the task execution area and the waiting area. The planning of the task execution area is driven by the actual constraints imposed by the module, ensuring realistic and relevant task allocation within this specific area. Each task block is arranged from top to bottom according to the order of task execution. Tasks at the same level are concurrent tasks and can be executed simultaneously; tasks at the next level need to wait until all tasks at this level are completed before they can begin to be selected for completion. A single agent, either a robot or an operator, is exclusively assigned to complete a specific task. In the event that there are no alternative tasks available at the current level for the idle agent, it will enter the waiting area and await the completion of all tasks before proceeding to the next level. Due to the potential limitations of the robot in independently completing the task of loosening screws, each task involving the loosening and removal of screws is subdivided into sub-tasks based on the number of screws involved. This approach facilitates the monitoring of the agent's progress in terms of the number of screws successfully completed. In practical scenarios, the presence of screws that have not been effectively removed can lead to the subsequent tasks being halted. Consequently, the sequence of tasks generated may no longer be viable. In the case where sub-tasks related to screws are not fully executed, each task is classified as temporarily faulty. These faulty tasks are uniformly referred to the operator for secondary verification and assistance in completion. Its logical framework is shown in Fig. 5.

The task execution area is defined by individual task blocks, which are determined based on the cost-benefit analysis of different agents performing the same task. In this environment, the two agents are allowed to explore freely, independently select tasks, and generate their own task sequences. Additionally, a standard total task sequence is generated based on their combined efforts. During the training process, the dual agents employ a fully cooperative strategy within the framework of a Markov game. The primary objective of the system is to achieve overall optimization by maximizing the total benefits obtained by both agents. The disassembly process of battery modules exhibits distinct characteristics compared to other disassembly objects and assembly segments. Specifically, there is a notable absence of a significant number of concurrent sequence tasks during the disassembly of battery modules. Moreover, the battery module's structure demonstrates distinct characteristics such as a well-defined hierarchical division and explicit constraints. These features contribute to its unique structural organization. Therefore, it is necessary to design the environment in a manner that takes into account the specific characteristics of the battery module. The determination of the optimal task sequence is based on optimizing task allocation, minimizing waiting time, and achieving efficient completion of the final task allocation as shown in Tables 2 and

Table 2
Comparison of reinforcement learning algorithms.

Property	DQN	QMIX	MADDPG
Iterative mode	Value-based	AC	AC
Whether to support multi-agent system	by adding structure	support	support
Whether to require discrete action	discrete	discrete	discrete or continuous
Performance in partially observable environments	relatively poor	preferable	normal
Whether it has an Actor-Critic structure	not have	possess	possess
Requirements for multi-agent relationships		collaboration	competition or collaboration

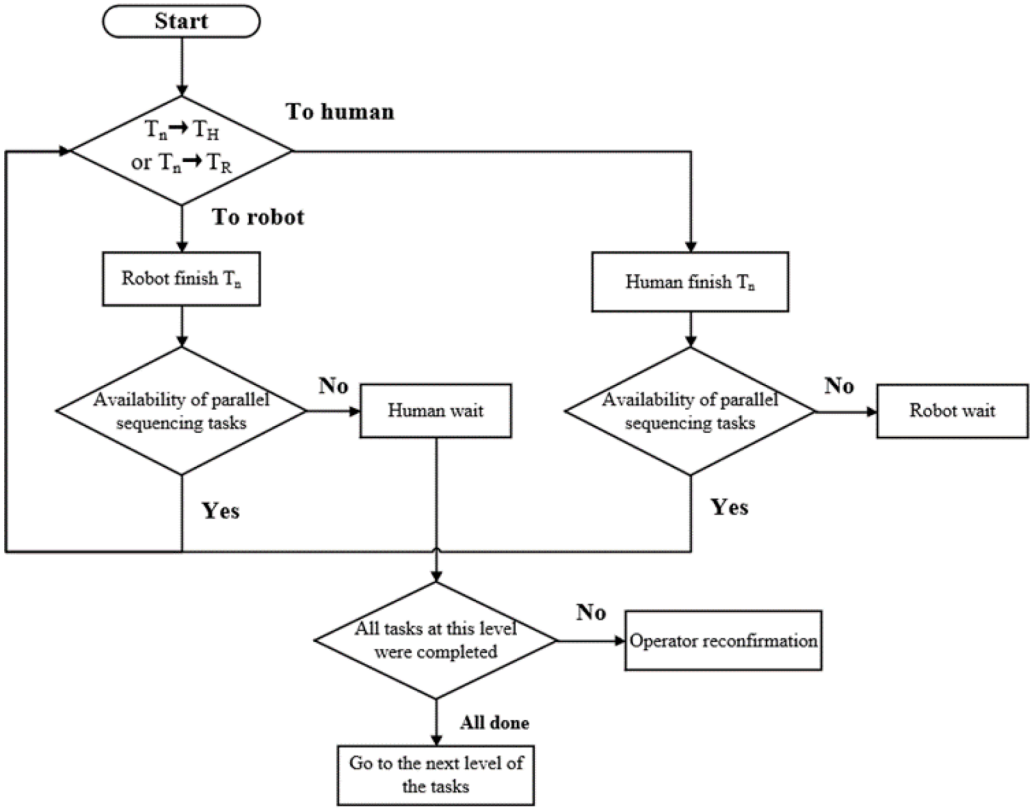

Fig. 5. Reinforcement learning environment logic.

Table 3
Symbol descriptions.

Notation	Explanation
A	Action space of an agent
a	Action of agent
S	State space
s	State of the environment
Obs	Observe space of an agent
π	Policy
π^*	Optimal strategy
$Aval\_A$	Available action of agent
γ	Discount rate
Env	Dual agent environment
$Cost$	The cost of the action consumption
P	State transition functions
ε	Greedy
θ	The parameter of evaluating network
θ^-	The parameter of the target network
lr	Learning rate
$Agent\_human$	The agent that represents a human
$Agent\_robot$	The agent that represents a robot
T_H	The task is done by the person
T_R	The task is done by the robot
M_p	disassembly priority matrix
R	Reward space

3

4.2. EV battery structure analysis

This study focuses on disassembly at the module level that the 3D model of the module is separated from Samsung MS372P5s. Initially, the study conducts a correlation analysis among the constituent elements, followed by an analysis of disassembly priority and direction interference. The relevance constraints reveal a direct relationship between two particular components. Based on the structural analysis of the three-dimensional model, it is determined that there exist connection relationships between the following components: part 1 (upper cover of the module) and part 8 (thick side shell), which are connected via a snap-fastener; part 2 (printed circuit board, PCB) and part 3 (internal support frame of the module), which are connected using screws; part 3, part 4 (pole head)and part 5 (pole piece), which are also connected by screws. The correlation constraint matrix M_C can be generated by summarizing the correlation constraints of each part. The priority analysis reveals a specific order of disassembly among these components, which is obtained through blast disassembly and interference analysis. Specifically, part 1 takes precedence over part 2, as the part 2 cannot be removed without removing the cover. Similarly, part 2 takes precedence over part 5, as the pole piece cannot be removed without the part2 being disassembled. Moreover, part 3 has a higher priority than part 6 (battery cell), as the module core cannot be dismantled without removing the part 3, part 5 has a higher priority than part 4, as the pole head cannot be detached without removing the pole piece. Lastly, it is observed that in order to dismantle the module core, the part 8 must be removed along with the part 6. The priority constraint matrix can be summarized by the priority relationship between parts. Please refer to Fig. 6 and Table 4 for further clarification. Through interference analysis, and disassembly priority matrix M_P and Correlation matrix M_C can be obtained by collation:

$$M_p = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad (5)$$

Table 4
The direction of interference analysis.

Disassembly task	Components	Interference direction	Human disassembly method	Robot disassembly tool
Remove the upper cap	Upper cap	$+Z$	Hand	Mechanical grippers chuck
Remove the M6 screws	M6 screws	$+Z$	Hand	Mechanical grippers
Remove Part Collection 1	Structural frame,Pole piece,PCB	$+Z$	Hand	Mechanical grippers chuck
Remove the M4 screws	M4 screws	$+Z$	Hand	Mechanical grippers
Remove the PCB	PCB	$+Z$	Hand	Mechanical grippers chuck
Remove the Pole piece	Pole piece	$+Z$	Hand	Mechanical grippers chuck

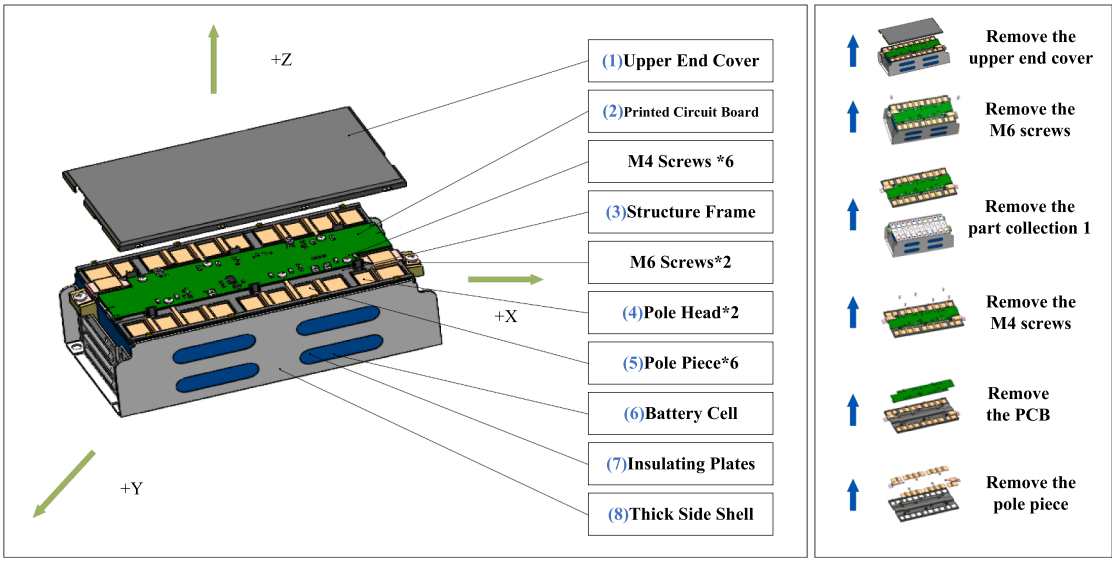

Fig. 6. Disassembly of directional interference.

$$M_C = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{6}$$

4.3. Evaluation analysis

The Disassembly task list is as follows: 1. To remove the upper-end cover of the battery module; 2. to unscrew the structural M6 screws; 3. To remove the structural M6 screws; 4. To remove the PCB, pole piece and structural frame; 5. To unscrew the PCB M4 screws; 6. To remove the PCB M4 screws; 7. To remove the PCB; 8. To remove the pole piece; 9. To remove the battery cell; and 10. To remove the insulating plates. The only task that belongs to the first level is to remove the upper cover of the module, which protects the internal devices of the module, and the other parts can be disassembled only after the upper cover is removed. The task belonging to the second level is loosening the two M6 screws securing the internal structural frame; the third level is tasked with removing all the loose screws and organizing them uniformly. At this point, the first stage of the disassembly task is initially completed. The battery module can continue to be disassembled into two parts, component group 1 is the PCB, pole piece, structure frame, and component group 2 is the cell, insulating plate, and module housing. To manufacture concurrent tasks as much as possible, both parts can be disassembled simultaneously in subsequent task operations. In this case, the process of refixation will not be talked about for the time being. At this point the fourth level is tasked with separating component group 1 from component group 2. The fifth level of tasks is loosening the six M4 screws that hold the PCB and internal structural frame in part set 1. Again this task will be divided into six sub-tasks, and any sub-task not completed will be adjusted to manual assistance; The mission of the same fifth level is to remove the core, which, due to the differences in fixation in this step, is temporarily considered in this case as not having been glued and fixed and can be removed directly. There are two tasks in Level 6, the first is to remove the six M4 screws that hold the PCB board and internal structural frame in Component Group 1, and the second is to remove the insulating plates in Component Group 2. The seventh level is tasked with removing the PCB in component group 1, and the eighth level is tasked with removing the sink in component group 1. At this point, the disassembly of the module is complete, with all parts disassembled into their most minor units and all fasteners removed. The disassembly process consists of 10 specific tasks, which can be turned into eight layers according to particular constraints, and the disassembly process needs to be executed from top to bottom in the order of the layers. As shown in Fig. 7, the reinforcement learning environment (dual agents) is established based on task analysis. The entire environment can be seen as a large grid world, and the process of two-agent walking represents the execution of tasks. The action space is a four-dimensional space composed of four actions, downward, left, right, and stay put. The observation space is a 3 × 3 grid environment around an agent.

The mission evaluation system employs a comprehensive approach that integrates multiple indicators for evaluation purposes. In order to determine the cost of each task, it is necessary to consider the difficulty level associated with completing the task for different operational roles, the time required to complete the task, and the inherent characteristics of the task. The complexity of the task is primarily influenced by factors such as the level of precision required for the operation, whether specialized tools are utilized, and the number of actions needed to successfully complete the task. In cases where the task exhibits characteristics of complexity, flexibility, and uncertainty, the time cost for human workforce to complete the task is relatively low. Conversely, for tasks characterized by repetitiveness, a high degree of danger, and physical exertion, the time cost for a robot to complete the task is relatively low. The comprehensive set of evaluation factors encompass various aspects, including the task's complexity, flexibility, uncertainty in actual production, level of danger, and physical labor, among others. Within this system, all disassembly tasks are categorized into three main groups: tasks that are most suitable for human completion, tasks that are most suitable for robot completion, and tasks that can be completed by either humans or robots. To effectively adapt to the varying degrees of difficulty for different roles in accomplishing these tasks, penalty coefficients α and β are assigned to each respective category. The value of α can range from 1 to 99 in different tasks, where 99 is a task that the robot cannot complete independently. The value of β ranges from 1 to 5 in different tasks, and 5 is the most labor-intensive task. The results of the task evaluation are shown in Table 5.

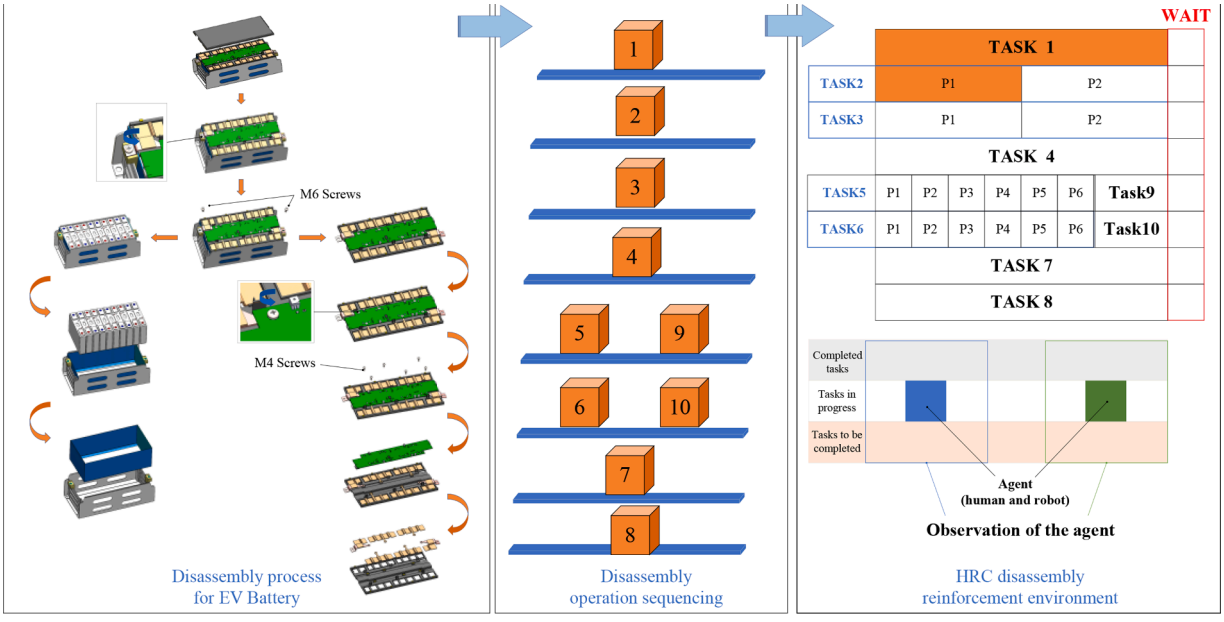

Fig. 7. The procedure of reinforcement learning environments.

Table 5
Evaluation analysis.

Task	Action role	Disassembly method	Cost
Remove the upper cap	Human	Hand	-β*2
	Robot	Mechanical grippers	-α*2
Loosen the M6 screws	Human	Electric screwdriver	-β*8
	Robot	Servo tightening machine	-α*8
Remove the M6 screws	Human	Hand	-β*12
	Robot	Mechanical grippers	-α*12
Remove Part Collection	Human	Hand	-β*4
	Robot	Mechanical grippers	-α*4
Loosen the M4 screws	Human	Electric screwdriver	-β*7.2
	Robot	Servo tightening machine	-α*7.2
Remove the M4 screws	Human	Hand	-β*12
	Robot	Mechanical grippers	-α*12
Remove the PCB	Human	Hand	-β*8
	Robot	Mechanical grippers	-α*8
Remove the Pole piece	Human	Hand	-β*28
	Robot	Mechanical grippers	-α*28
Remove the Battery cells	Human	Hand	-β*6
	Robot	Mechanical grippers	-α*6
Remove the Insulating gaskets	Human	Hand	-β*20
	Robot	Mechanical grippers	-α*20

4.4. Results and discussion

This deep reinforcement learning project is constructed based on Pytorch. The whole project is deployed on a computer with an NVIDIA RTX3070–8Gb GPU. The interpreter is Python 3.8, and the GPU is driven with NVIDIA CUDA 11.6. The experimental parameter settings are shown in Table 6. The pseudocode of the algorithm is shown in Fig. 8.

Upon completion of the training process using the selected parameters, the agent has two possible ways to conclude a round of iterations. The first is by successfully completing all the assigned tasks within the given parameters. Alternatively, the agent may also conclude the round if it exceeds the specified upper limit of steps during the iteration process. After a certain number of iterations, the training effect is shown in Fig. 9. The figure illustrates the progression of the total cost incurred by the two agents as they complete the tasks in each round of iteration. During the initial phase of training, the inability of the two agents to locate an optimal path to execute the task was observed, primarily due to the presence of a penalty factor leading to excessively high costs. During the intermediate phase of training, the two agents progressively develop a strategic approach, leading to a gradual decrease and eventual convergence of the total cost. This process ultimately results in the identification of an optimal strategy. Notably, In the partially observable environment, QMIX-HRC algorithm has better iterative stability than the traditional DQN algorithm. The comparison between QMIX-HRC and DQN algorithms is shown in Table 7, where two main metrics are shown, namely convergence speed and convergence stability. In terms of convergence speed QMIX-HRC algorithm converges to the vicinity of the optimal solution after 16 iterations, and DQN algorithm converges to the vicinity of the optimal solution after 24 iterations, and QMIX-HRC

Table 6
Parameters of the algorithm.

Parameter	Value
Number of agents	2
Action space dim	4
Episode limit	5000
γ	0.99
Learning rate	0.0005
Batch size	32
Q-mix hidden dim	32
Hyper layers number	1
LSTM hidden dim	64
Update frequency of the target network	200
Max train steps	5000
Buffer size	5000

QMIX-HRC Algorithm

Input: Obsevre space, State space, Action space, Learning rate, Episode limit, Buffer size, γ

Output: Cost, Strategy

For **N** in (Episode limit) **:**

> All parameters are initialized
>
> Get state, agent_human observe, agent_robot observe
>
> Get available action list from Env
>
> Choose person_action , robot_action:
>
> > **If random < γ:**
> > Random select action from avalible action list
> > **else:**
> > Choose max Q-value action from avalible action list
>
> Get reward from Env
>
> Store state,agent_human observe, agent_robot observe avalible action list, reward in buffer D
>
> **If len(D) ≧ Buffer size :**
> Sample random same transition from different episode from buffer D
>
> Update parameters
>
> State=State$_{(t+1)}$, avail action list= avail action list $_{(t+1)}$
>
> Copy the θ to θ^{-}

Fig. 8. The procedure of the QMIX-HRC algorithm.

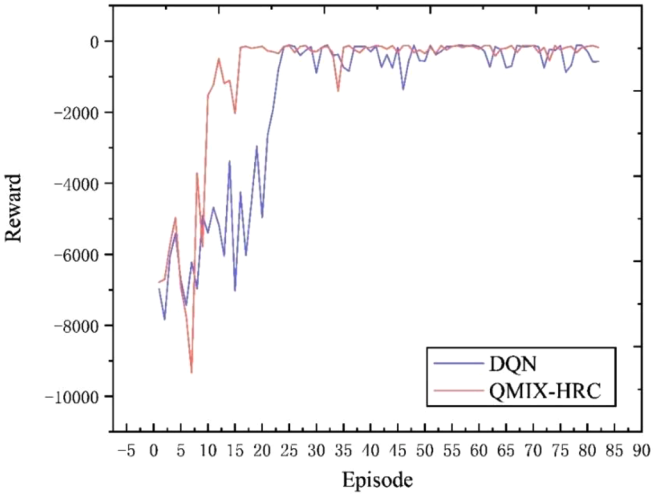

Fig. 9. QMIX-HRC algorithm results.

Table 7
Comparison of algorithm results.

Index	QMIX-HRC	DQN
Rate of convergence	**16**	24
Convergent stability	**1.07**	2.32

algorithm's convergence speed is faster than that of DQN; in terms of convergence stability, we use the average of the gains after the algorithm converges to the optimal solution attachment to do the error analysis with the optimal solution gains, and we get the QMIX-HRC's The average error of QMIX-HRC is 1.07, which is much lower than the average error of DQN of 2.32, proving that the convergence stability of QMIX-HRC algorithm is better.

The sequential disassembly task assignments for achieving optimal disassembly are illustrated as shown in Fig. 10. Step 1 involves the manual disassembly of the module's upper cover. Subsequently, in Step 2, the robot is responsible for loosening the M6 screws that secure the module support frame. In Step 3, it is the human operator's task to verify and retrieve the M6 screws. Moving forward, Step 4 entails the manual disassembly of the module, encompassing the structural frame, PCB, and pole pieces. Simultaneously, the robot is engaged in Step 5, assisting in removing the M4 screws that secure the PCB while the core is manually extracted from the module. Step 6 consists of manually removing the insulating shims inside the module, followed by Step 7, which involves the manual removal of the previously extracted M4 screws. Continuing with the disassembly process, Step 8 necessitates the manual removal of the PCB, while Step 9 concludes with the manual removal of the pole pieces. Upon completion of Step 9, all components of the module have been disassembled into their smallest units, thereby signifying the fulfillment of the disassembly task for the module.

In contrast to conventional reinforcement learning environments employed in the context of addressing collaborative task problems between humans and robots, the utilization of learning environments characterized by partial observability offers a closer approximation to real-world industrial scenarios. In the context of addressing problems involving partial observability, the incorporation of RNN facilitates the retention of information over extended durations, thereby enhancing the network's capacity to predict Q-values effectively. A comparative analysis reveals that the utilization of RNN within the DRQN framework results in faster convergence speed and improved convergence accuracy when compared to the traditional DQN architecture. These superior characteristics render the DRQN network with RNN more suitable for application in industrial settings.

5. Disassembly experiment analysis

The optimization strategy in this paper is validated on a real human-robot disassembly station with simulated EV battery workpiece disassembly. This station consists of the ROKAE collaborative robot, AMX226 plc, DDK servo tightening machine, and Intel D435i depth camera. The configuration of the experimental setup is illustrated in Fig. 11. Given the inherent risks associated with the disassembly of an actual battery

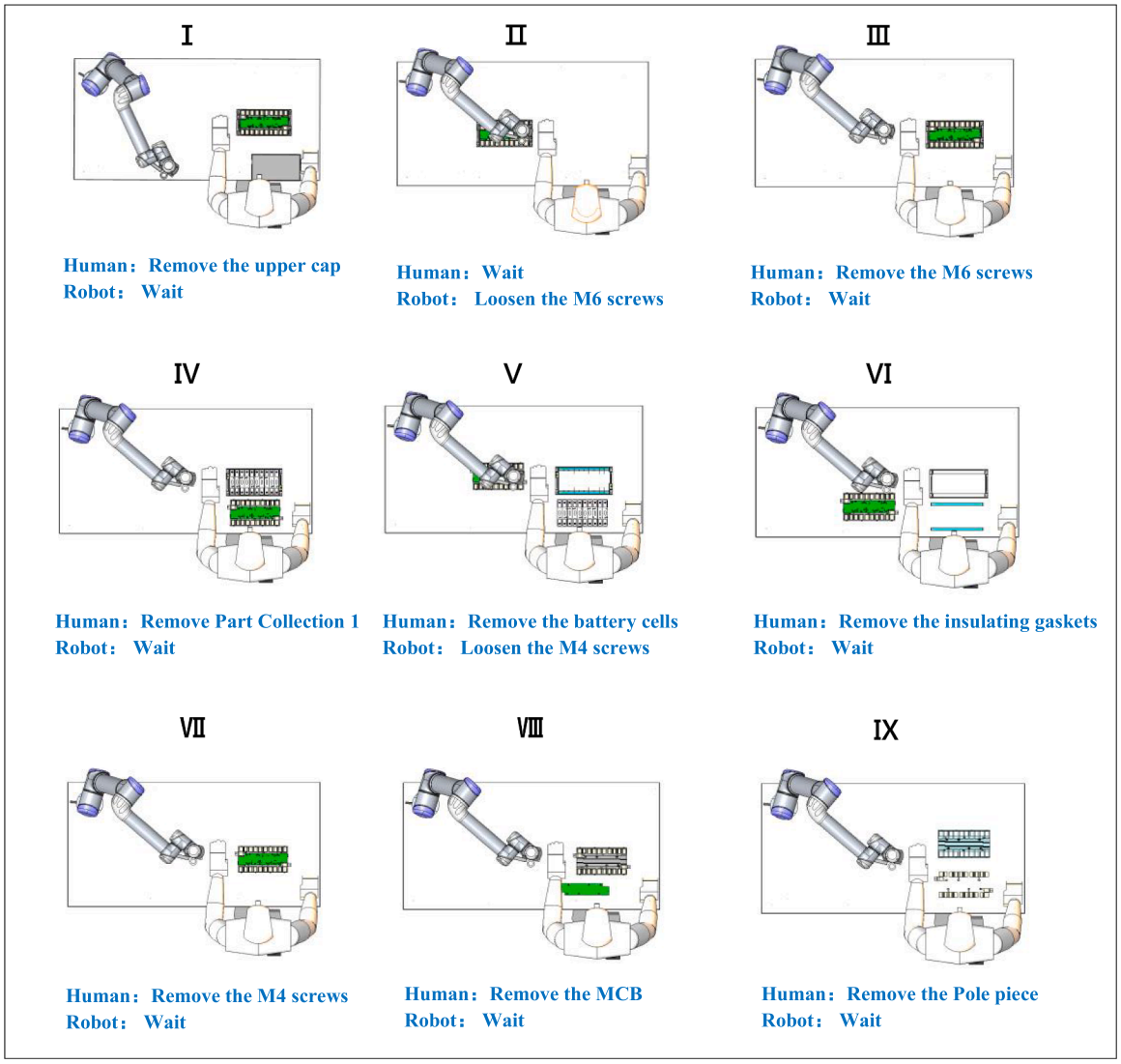

Fig. 10. Demonstration of the optimal disassembly sequence.

Fig. 11. Human-robot collaborative disassembly workstation.

pack, precautions were taken to ensure the safety of the experimental procedure. Specifically, the disassembly process was designed based on a realistic structural model of a standardized battery pack shell, thus mitigating potential hazards. As shown in Fig. 12, the disassembly object consists of the following parts: the battery pack upper case, the battery pack lower case, the battery module upper cover, the battery module lower cover, the cells, and the fixings. The disassembly objective of this experiment focuses on the part of removing the top cover of the battery module, and the specific tasks are as follows: 1. To loosen all the fastening bolts on the upper cover of the battery module; 2. To remove all the fastening bolts on the upper cover of the battery module; 3. in this experiment, the fastening bolts on the top cover of the battery module use hexagonal bolts, and the servo tightening machine has a series of sizes of hexagonal sockets with it to adapt to the different specifications of the bolts. As depicted in the illustration, the workbench can be delineated into three distinct zones: 1) the robot work area, which is off-limits to operators when the robot is in its operational state. 2) the operator's work area, which permits human intervention during the execution of concurrent task sequences, and 3) the tool area, designated for storing various disassembly tools, accessible to both the robot and human personnel for tool replacement purposes.

As depicted in Fig. 12, following the application of the optimized

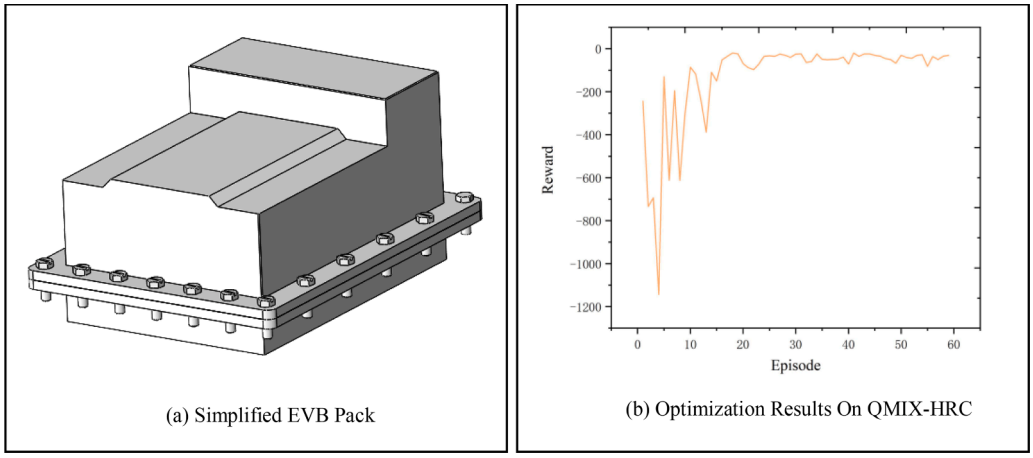

Fig. 12. Simulated EV battery structure and its process optimization results.

disassembly strategy, all initially allocated unscrewing tasks are delegated to the collaborative robot. The robot utilizes a servo-tightening machine to perform the disassembly process on each of the eight objects targeted for disassembly. The human operator is tasked with collecting the loosened bolts, verifying the completion of the disassembly process, and subsequently removing the unsecured top cover of the battery module. A rudimentary calculation of labor hours demonstrates that employing a robot-assisted personnel approach for the bolt removal task results in a reduction in overall labor hours, when compared to the algorithm-based method of exclusively manual execution. As shown in Fig. 13 (A-D), after the program is set up, the robot acquires the coordinates of all 8 points and disassembles them one by one in order, as shown in Figs. 13(E) and (F), it manually collects all the nuts and confirms whether the quantity is correct or not; as shown in Fig. 13(G), it manually removes the top cover of the unconstrained battery module from the Z-direction, and, as shown in Fig. 13(H), it finally disassembles it into a single part. The demonstration exemplifies the feasibility of disassembling an EV battery pack at a human-robot collaborative workstation. Also, it proves that human-robot collaborative disassembly is characterized by high efficiency, high safety, and low personnel fatigue intensity compared with current battery disassembly methods. The comparison of disassembling methods can be described as shown in Table 8 from experiments and some references [21,42,49].

In the process of experiment and analysis, we opted to employ a virtualized battery system as a means to circumvent the potential hindrances stemming from the inherent quality issues of the physical battery pack when subjected to real-world operating conditions. In actual battery disassembly procedures, the removal of bolt fixations is followed by the use of supplementary tools for tasks such as adhesive detachment and other related operations. This type of process necessitates a significant level of flexibility and maneuverability. However, notable potential for automation lies in the disassembly of bolt fixations, which are prominent in both the battery pack and module disassembly stages The presented disassembly procedure showcased herein exemplifies a quintessential bolt disassembly operation pertinent to the broader battery disassembly process, and concurrently represents a task that is not only inherently tedious and labor-intensive, but also time-consuming when carried out through the conventional manual disassembly methodology [49].

Table 8
Comparison of disassembling methods.

Compare metrics	Purely manual disassembling method	Optimized human-robot collaborative disassembling method
Bolt loosening time (Single)	5 s	4.5 s
Bolt loosening time (All)	120 (15×8) seconds	76 (9.5 × 8)seconds
Total disassembling time	about 340 s	about 280 s
Safety	The risk comes from cell leakage, sharp edges	The risk comes from human-robot collisions
Total disassembling difficulty(0~1)	0.6	0.47
The labor intensity of human	Carry, locating, clasped	Carry

6. Conclusions

To overcome the challenges associated with the optimization of HRC disassembly sequences in EV battery systems, this study proposes a reinforcement learning framework called QMIX-HRC. By constructing a reinforcement learning environment that accounts for partial observability, we transform the HRC work scenario into a multi-agent game process. The QMIX-HRC addresses the optimization difficulties caused by non-stationarity and partial observability in human-robot disassembly. This allows each agent to select the strategy that maximizes the overall benefit, thus ensuring optimal disassembling strategies for both human and robot. Compared with the commonly used DQN algorithm, the QMIX-HRC framework we use has better convergence stability. Through the exploration of distinct reinforcement learning environments customized for various EV battery categories, it is possible to ascertain the most effective disassembly sequences for each battery type. By using this system, we effectively solve the problem of low disassembling efficiency in current retired power batteries, while also providing flexibility in handling instability factors through environmental changes. In order to prove the effectiveness of the algorithm, we took an example for a disassembling experiment using a simulated power battery pack on the HRC disassembling workstation, and the experimental results show that the task sequence optimized by the algorithm reduces the working time and reduces the labor intensity

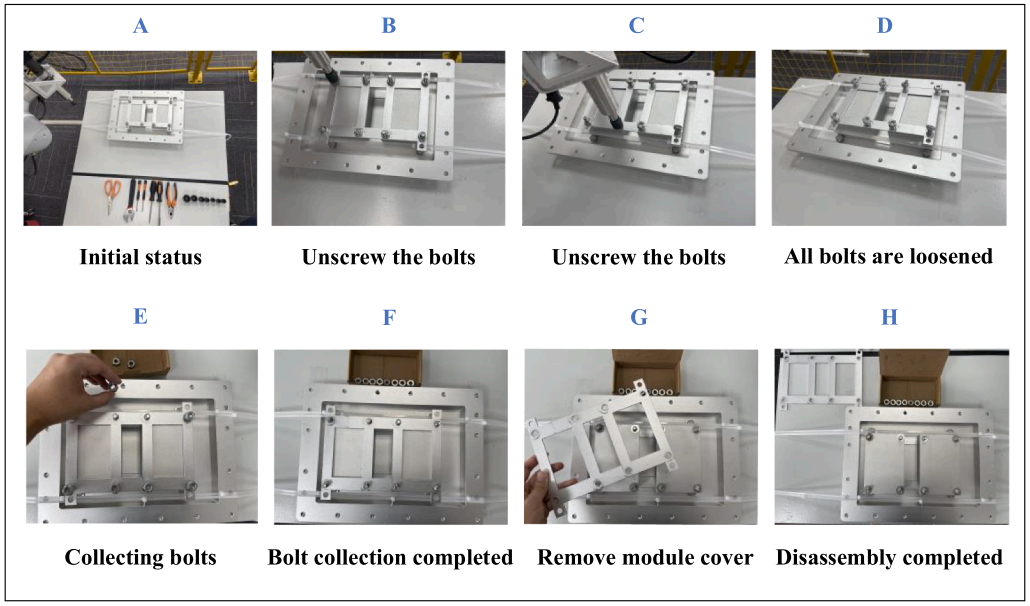

Fig. 13. Experimental demonstration based on a battery pack model.

compared with the existing methods. Future research will primarily focus on optimizing execution strategies between HRC workstations composed of multiple robots, as well as optimizing more complex task execution sequences. This includes the development of additional end-effectors to facilitate robot-assisted collaboration with humans.

CRediT authorship contribution statement

Jiaxu Gao: Conceptualization, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – original draft. **Guoxian Wang:** Writing – review & editing. **Jinhua Xiao:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Pai Zheng:** Writing – review & editing. **Eujin Pei:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

The authors are very grateful to all the anonymous reviewers for the valuable opinions and suggestions on the improvement of our paper. This research is supported by National Natural Science Foundation of China, No. 52305551.

Author Agreement Statement

We the undersigned declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We understand that the Corresponding Author is the sole contact for the Editorial process. He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs.

References

[1] L. Wang, X. Wang, W. Yang, Optimal design of electric vehicle battery recycling network – From the perspective of electric vehicle manufacturers, Appl. Energy. 275 (2020) 21, https://doi.org/10.1016/j.apenergy.2020.115328.

[2] L.H. Saw, Y. Ye, A.A.O. Tay, Integration issues of lithium-ion battery into electric vehicles battery pack, J. Clean. Prod. 113 (2016) 1032–1045, https://doi.org/10.1016/j.jclepro.2015.11.011.

[3] C. Liu, J. Lin, H. Cao, Y. Zhang, Z. Sun, Recycling of spent lithium-ion batteries in view of lithium recovery: a critical review, J. Clean. Prod. 228 (2019) 801–813, https://doi.org/10.1016/j.jclepro.2019.04.304.

[4] A. Tripathy, A. Bhuyan, R. Padhy, L. Corazza, Technological, organizational, and environmental factors affecting the adoption of electric vehicle battery recycling, IEEE Trans. Eng. Manag. (2022), https://doi.org/10.1109/TEM.2022.3164288.

[5] G. Harper, R. Sommerville, E. Kendrick, L. Driscoll, P. Slater, R. Stolkin, A. Walton, P. Christensen, O. Heidrich, S. Lambert, A. Abbott, K. Ryder, L. Gaines, P. Anderson, Recycling lithium-ion batteries from electric vehicles, Nature 575 (2019) 75–86, https://doi.org/10.1038/s41586-019-1682-5.

[6] L. Yun, D. Linh, L. Shui, X. Peng, A. Garg, M. Loan, P. Le, S. Asghari, Resources, Conservation & Recycling Metallurgical and mechanical methods for recycling of lithium-ion battery pack for electric vehicles, Resour. Conserv. Recycl. 136 (2018) 198–208, https://doi.org/10.1016/j.resconrec.2018.04.025.

[7] K. Wegener, S. Andrew, A. Raatz, K. Dröder, C. Herrmann, Disassembly of electric vehicle batteries using the example of the Audi Q5 hybrid system, Procedia CIRP 23 (2014) 155–160, https://doi.org/10.1016/j.procir.2014.10.098.

[8] C. Herrmann, A. Raatz, S. Andrew, J. Schmitt, Scenario-based development of disassembly systems for automotive lithium ion battery systems, Adv. Mater. Res. 907 (2014) 391–401, https://doi.org/10.4028/www.scientific.net/AMR.907.391.

[9] J. Xiao, C. Jiang, B. Wang, A review on dynamic recycling of electric vehicle battery: disassembly and echelon utilization, Batteries 9 (2023) 57, https://doi.org/10.3390/batteries9010057.

[10] H. Zhang, J. Huang, R. Hu, D. Zhou, H. ur R. Khan, C. Ma, Echelon utilization of waste power batteries in new energy vehicles: review of Chinese policies, Energy 206 (2020) 118178, https://doi.org/10.1016/j.energy.2020.118178.

[11] L. Zhou, A. Garg, J. Zheng, L. Gao, K. Oh, Battery pack recycling challenges for the year 2030: recommended solutions based on intelligent robotics for safe and efficient disassembly, residual energy detection, and secondary utilization, Energy Storage 3 (2021), https://doi.org/10.1002/est2.190.

[12] J. Xiao, N. Anwer, W.D. Li, B. Eynard, Dynamic Bayesian network-based disassembly sequencing optimization for electric vehicle battery, CIRP J. Manuf. Sci. Technol. 38 (2022) 824–835, https://doi.org/10.1016/j.cirpj.2022.07.010.

[13] M. Choux, W.S. Pripp, F. Kvalnes, M. Hellström, To shred or to disassemble–a techno-economic assessment of automated disassembly vs. shredding in lithium-ion battery module recycling, Resour. Conserv. Recycl. 203 (2024) 107430, https://doi.org/10.1016/j.resconrec.2024.107430.

[14] S. Hjorth, D. Chrysostomou, Human–robot collaboration in industrial environments: a literature review on non-destructive disassembly, Robot. Comput. Integr. Manuf. 73 (2022) 102208, https://doi.org/10.1016/j.rcim.2021.102208.

[15] S.A. Green, M. Billinghurst, X. Chen, J.G. Chase, Human-robot collaboration: a literature review and augmented reality approach in design, Int. J. Adv. Robot. Syst. 5 (2008) 1–18, https://doi.org/10.5772/5664.

[16] A.D. Dragan, S. Bauman, J. Forlizzi, S.S. Srinivasa, Effects of robot motion on human-robot collaboration, in: ACM/IEEE Int. Conf. Human-Robot Interact. 2015-March, 2015, pp. 51–58, https://doi.org/10.1145/2696454.2696473.

[17] J. Xiao, J. Gao, N. Anwer, B. Eynard, Multi-agent reinforcement learning method for disassembly sequential task optimization based on human-robot collaborative disassembly in electric vehicle battery recycling, J. Manuf. Sci. Eng. Trans. ASME. 145 (2023) 121001, https://doi.org/10.1115/1.4062235.

[18] A. Rastegarpanah, H.C. Gonzalez, R. Stolkin, Semi-autonomous behaviour tree-based framework for sorting electric vehicle batteries components, Robotics 10 (2021) 1–18, https://doi.org/10.3390/robotics10020082.

[19] J. Xiao, W. Li, Y. Lv, G. Du, Disassembly information interoperability for electric vehicle battery in remanufacturing based on STEP standards, Procedia CIRP 104 (2021) 1873–1877, https://doi.org/10.1016/j.procir.2021.11.316.

[20] S. Glöser-Chahoud, S. Huster, S. Rosenberg, S. Baazouzi, S. Kiemel, S. Singh, C. Schneider, M. Weeber, R. Miehe, F. Schultmann, Industrial disassembling as a key enabler of circular economy solutions for obsolete electric vehicle battery systems, Resour. Conserv. Recycl. 174 (2021), https://doi.org/10.1016/j.resconrec.2021.105735.

[21] J.F. Hellmuth, N.M. DiFilippo, M.K. Jouaneh, Assessment of the automation potential of electric vehicle battery disassembly, J. Manuf. Syst. 59 (2021) 398–412, https://doi.org/10.1016/j.jmsy.2021.03.009.

[22] E. Martinez-Laserna, I. Gandiaga, E. Sarasketa-Zabala, J. Badeda, D.I. Stroe, M. Swierczynski, A. Goikoetxea, Battery second life: hype, hope or reality? A critical review of the state of the art, Renew. Sustain. Energy Rev. 93 (2018) 701–718, https://doi.org/10.1016/j.rser.2018.04.035.

[23] J.R. Li, L.P. Khoo, S.B. Tor, A novel representation scheme for disassembly sequence planning, Int. J. Adv. Manuf. Technol. 20 (2002) 621–630, https://doi.org/10.1007/s001700200199.

[24] K. Wegener, W. Hua, F. Dietrich, K. Dröder, S. Kara, Robot assisted disassembly for the recycling of electric vehicle batteries, Procedia CIRP 29 (2015) 716–721, https://doi.org/10.1016/j.procir.2015.02.051.

[25] J. Liu, Z. Zhou, D.T. Pham, W. Xu, C. Ji, Q. Liu, Robotic disassembly sequence planning using enhanced discrete bees algorithm in remanufacturing, Int. J. Prod. Res. 56 (2018) 3134–3151, https://doi.org/10.1080/00207543.2017.1412527.

[26] J. Zhang, B. Li, A. Garg, Y. Liu, A generic framework for recycling of battery module, Energy Res 42 (2018) 3390–3399, https://doi.org/10.1002/er.4077.

[27] M. Alfaro-Algaba, F.J. Ramirez, Techno-economic and environmental disassembly planning of lithium-ion electric vehicle battery packs for remanufacturing, Resour. Conserv. Recycl. 154 (2020) 104461, https://doi.org/10.1016/j.resconrec.2019.104461.

[28] W.J. Tan, C.M.M. Chin, A. Garg, L. Gao, A hybrid disassembly framework for disassembly of electric vehicle batteries, Int. J. Energy Res. 45 (2021) 8073–8082, https://doi.org/10.1002/er.6364.

[29] S. Baazouzi, F.P. Rist, M. Weeber, K.P. Birke, Optimization of disassembly strategies for electric vehicle batteries, Batteries 7 (2021), https://doi.org/10.3390/batteries7040074.

[30] Y. Wang, F. Lan, J. Liu, J. Huang, S. Su, C. Ji, Z. Zhou, Interlocking problems in disassembly sequence planning, Int. J. Prod. Res. 59 (15) (2021) 4723–4735, https://doi.org/10.1080/00207543.2020.1770892.

[31] M. Kheder, M. Trigui, N. Aifaoui, Optimization of disassembly sequence planning for preventive maintenance, Int. J. Adv. Manuf. Technol. 90 (2017) 1337–1349, https://doi.org/10.1007/s00170-016-9434-2.

[32] Y. Laili, X. Li, Y. Wang, L. Ren, X. Wang, Robotic disassembly sequence planning with backup actions, IEEE Trans. Automat. Sci. Eng. 19 (3) (2021) 2095–2107, https://doi.org/10.1109/TASE.2021.3072663.

[33] T. Yu, J. Huang, Q. Chang, Optimizing task scheduling in human-robot collaboration with deep multi-agent reinforcement learning, J. Manuf. Syst. 60 (2021) 487–499, https://doi.org/10.1016/j.jmsy.2021.07.015.
[34] K. Tsiakas, M. Papakostas, M. Papakostas, M. Bell, R. Mihalcea, S. Wang, M. Burzo, F. Makedon, An interactive multisensing framework for personalized human robot collaboration and assistive training using reinforcement learning, in: ACM Int. Conf. Proceeding Ser. Part F1285, 2017, pp. 423–427, https://doi.org/10.1145/3056540.3076191.
[35] Z. Liu, Q. Liu, L. Wang, W. Xu, Z. Zhou, Task-level decision-making for dynamic and stochastic human-robot collaboration based on dual agents deep reinforcement learning, Int. J. Adv. Manuf. Technol. 115 (2021) 3533–3552, https://doi.org/10.1007/s00170-021-07265-2.
[36] R. Zhang, Q. Lv, J. Li, J. Bao, T. Liu, S. Liu, A reinforcement learning method for human-robot collaboration in assembly tasks, Robot. Comput. Integr. Manuf. 73 (2022) 102227, https://doi.org/10.1016/j.rcim.2021.102227.
[37] L. Roveda, J. Maskani, P. Franceschi, A. Abdi, F. Braghin, L. Molinari Tosatti, N. Pedrocchi, Model-based reinforcement learning variable impedance control for human-robot collaboration, J. Intell. Robot. Syst. Theory Appl. 100 (2020) 417–433, https://doi.org/10.1007/s10846-020-01183-3.
[38] R. Agarwal, D. Schuurmans, M. Norouzi, An optimistic perspective on offline reinforcement learning, in: 37th Int. Conf. Mach. Learn. ICML 2020. PartF16814, 2020, pp. 92–102.
[39] A. Shafti, J. Tjomsland, W. Dudley, A.A. Faisal, Real-world human-robot collaborative reinforcement learning, IEEE Int. Conf. Intell. Robot. Syst. (2020) 11161–11166, https://doi.org/10.1109/IROS45743.2020.9341473.
[40] M. Khamassi, G. Velentzas, T. Tsitsimis, C. Tzafestas, Active exploration and parameterized reinforcement learning applied to a simulated human-robot interaction task, in: Proc. - 2017 1st IEEE Int. Conf. Robot. Comput. IRC 2017, 2017, pp. 28–35, https://doi.org/10.1109/IRC.2017.33.
[41] K. Li, Q. Liu, W. Xu, J. Liu, Z. Zhou, H. Feng, Sequence planning considering human fatigue for human-robot collaboration in disassembly, Procedia CIRP 83 (2019) 95–104, https://doi.org/10.1016/j.procir.2019.04.127.
[42] W. Xu, Q. Tang, J. Liu, Z. Liu, Z. Zhou, D.T. Pham, Disassembly sequence planning using discrete Bees algorithm for human-robot collaboration in remanufacturing, Robot. Comput. Integr. Manuf. 62 (2020) 101860, https://doi.org/10.1016/j.rcim.2019.101860.
[43] L. Guo, Z. Zhang, X. Zhang, Human–robot collaborative partial destruction disassembly sequence planning method for end-of-life product driven by multi-failures, Adv. Eng. Inf. 55 (2023) 101821, https://doi.org/10.1016/j.aei.2022.101821.
[44] H.Y. Liao, Y. Chen, B. Hu, S. Behdad, Optimization-based disassembly sequence planning under uncertainty for human–robot collaboration, J. Mech. Des. 145 (2) (2023) 022001, https://doi.org/10.1115/1.4055901.
[45] M. Chu, W. Chen, Human-robot collaboration disassembly planning for end-of-life power batteries, J. Manuf. Syst. 69 (2023) 271–291, https://doi.org/10.1016/j.jmsy.2023.06.014.
[46] Y. Fang, Z. Li, S. Wang, X. Lu, Multi-objective multi-fidelity optimisation for position-constrained human-robot collaborative disassembly planning, Int. J. Prod. Res. (2023) 1–18, https://doi.org/10.1080/00207543.2023.2251064.
[47] A. Allagui, I. Belhadj, R. Plateaux, M. Hammadi, O. Penas, N. Aifaoui, Reinforcement learning for disassembly sequence planning optimization, Comput. Ind. 151 (2023) 103992, https://doi.org/10.1016/j.compind.2023.103992.
[48] T. Rashid, M. Samvelyan, C.S. De Witt, G. Farquhar, J. Foerster, S. Whiteson, Monotonic value function factorisation for deep multi-agent reinforcement learning, J. Mach. Learn. Res. 21 (2020) 1–51.
[49] R. Gerbers, K. Wegener, F. Dietrich, K. Dröder, Safe, flexible and productive human-robot-collaboration for disassembly of lithium-ion batteries, Recycl. Lithium-Ion Batter.: Lithorec Way (2018) 99–126, https://doi.org/10.1023/A:1008089230047.