Robotics and Computer–Integrated Manufacturing 95 (2025) 103025

Contents lists available at ScienceDirect

Robotics and Computer-Integrated Manufacturing

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

Full length article

Towards human-centric manufacturing: A reinforcement learning method for physical exertion alleviation in HRCA

Yingchao You, Ze Ji *

School of Engineering, Cardiff University, Queen's Buildings, The Parade, Cardiff, CF24 3AA, Wales, UK

ARTICLE INFO

Keywords:
Assembly
Human–robot collaboration
Physical exertion
Reinforcement learning
DuelingDQN

ABSTRACT

The advancement of the manufacturing system towards more human-centric, emphasising not only efficiency but also the well-being of workers. However, task planning in human–robot collaborative assembly (HRCA) remains challenging, when considering the physical exertion alleviation of workers, due to the complexities of physical exertion estimation and variations in human assembly operations. Different from conventional methods, this paper proposes a task planning method for physical exertion alleviation of workers in HRCA by leveraging the reinforcement learning (RL) method to train a policy. Initially, a musculoskeletal model-based method driven by human movement data to assess workers' physical exertion is integrated into this work. Then, the policy is trained in a DuelingDQN-AM framework, utilising a carefully designed reward function informed by the estimated physical exertion of workers. The effectiveness of this approach has been validated through a simulation experiment and a proof-of-concept real assembly experiment. Simulation experiment results demonstrate the advantages of DuelingDQN-AM over other methods in terms of convergence speed and stability across multiple cycles and products of varying complexity. Additionally, real-world experiment results show that the RL strategy reduced physical exertion by 15.63% compared to the baseline method.

1. Introduction

The advancement of human-centric manufacturing shows the focus on workers' safety, comfort, and health rather than only the process and efficiency when designing manufacturing systems. However, manufacturing workers still suffer from musculoskeletal disorders (MSD), which are mainly caused by physical fatigue. A study shows that around 53.1% of workers in automobile manufacturing workers have MSD conditions [1]. This data shows that there is a long way to go to manage the physical exertion of the workers to reduce the MSD rate.

Human–robot collaboration (HRC) is an important manufacturing paradigm that combines the strength of humans and robots, advancing the flexibility of the manufacturing system. Task planning plays an important role in the HRC, which can allocate subtasks to agents in sequence to complete a shared goal to ensure efficient resource utilisation and optimal performance. The factors considered in the task planning methods are diverse and may include agents' capabilities, task characteristics, efficiency, safety, and so on [2]. However, to our knowledge, it is rare for work to consider the reduction of physical exertion of human operators when planning tasks in HRCA.

Many task planners are proposed for HRCA in the literature [3]. Task planning in HRC is proven to be an NP-hard problem [4]. To solve such a problem, heuristic methods [5] were proposed to find optimal solutions rather than the exact solution within an acceptable time. Besides, some works adapt behaviour trees-based methods [6,7] to model the assembly process and obtain an allocation plan. Behaviour trees belong to a knowledge model, which is designed by the domain experts manually. The RL method can learn a policy by interacting with the environment, which has achieved some results on graph optimisation problems [8], travelling salesman problem [9], etc. Additionally, the RL method offers adaptability and generalisation capabilities in dynamic and uncertain environments, inherent in dynamic scheduling problems, where traditional methods often struggle [10]. Based on the above analysis, we believe that RL has the potential to address the challenge of physical exertion alleviation problem in HRCA.

In mitigating workers' physical exertion in HRCA, one of the primary challenges is the absence of real-time physical exertion estimation methods for workers. Existing methods, such as the Borg RPE scale [11], rely on subjective user feedback, which can lead to inaccurate measurements and interruptions in workflow. Ergonomic approaches, like the Rapid Upper Limb Assessment [12], also struggle to capture individual differences in human physical exertion accurately. Another challenge lies in the variability of human operations. Workers often have personal preferences in how they perform assembly tasks,

* Corresponding author.
E-mail addresses: Youy4@cardiff.ac.uk (Y. You), JiZ1@cardiff.ac.uk (Z. Ji).

https://doi.org/10.1016/j.rcim.2025.103025
Received 15 December 2024; Received in revised form 23 February 2025; Accepted 29 March 2025
Available online 16 April 2025

which can vary between individuals [13]. To avoid disrupting these preferences, this paper assumes that the task planner does not enforce specific work content on workers. The uncertainty in worker operations adds complexity to task planning and makes alleviating worker physical exertion more challenging.

To address the above-mentioned limitations, this paper proposes an RL-based framework for mitigating the physical exertion of workers in HRCA. A human musculoskeletal method from our previous work [14] is adapted to estimate real-time physical exertion based on human movement data. Then, a DuelingDQN algorithm, combined with an action masking technique, is introduced to train a policy alongside a random policy to manage the human operation variations. The reward function of the DuelingDQN is carefully designed to integrate the estimated physical exertion values, guiding task allocation. The policy enables the robot to proactively undertake tasks that are most fatiguing for the human operator, thus, alleviating the human physical exertion. Finally, we conduct both simulation and real-world experiments to demonstrate the effectiveness of our method.

The contribution of this work is summarised as follows:

1. The problem of human physical exertion alleviation in HRCA is modelled as a multi-agent Markov decision process, and a novel RL-based framework is proposed to address this challenge.
2. An RL algorithm, DuelingDQN-AM, which is informed by a musculoskeletal-based physical exertion assessment method, is introduced for task planning in HRCA to mitigate workers' physical exertion.
3. A real-world assembly experiment involving multiple participants was conducted to comprehensively validate the proposed framework. The results demonstrate that the RL strategy reduces physical exertion by 15.63% compared to the baseline strategy.

2. Literature review

2.1. Task planning methods in HRC

HRC is an emerging manufacturing paradigm that leverages the strengths of human workers, such as dexterity, flexibility, perception, and intelligence, and robots, known for their precision, repeatability, and strength. This integration enhances the overall adaptability and flexibility of manufacturing systems. However, task planning in HRC remains a challenging issue due to constraints related to resources, agent capabilities, and assembly requirements, which has attracted significant research interest [15]. Addressing the complexities of task planning involves considering multiple factors, including resource availability, task specifications, operation time, ergonomics, safety, costs, product quality, workload, movement efficiency, and space utilisation.

In efforts to alleviate physical exertion on workers during HRC, several task planners have developed scheduling algorithms to optimise worker fatigue management [16–18]. For instance, Li et al. proposed a discrete bees algorithm for sequencing tasks, aimed at reducing the time required for disassembly tasks and mitigating fatigue-related inefficiencies [16]. However, these methods have largely been validated in simulations, with limited real-world evidence to support their practical effectiveness. Additionally, research has explored role allocation and co-manipulation in HRC [7,19,20]. While these studies primarily focus on the fatigue of movement primitives, they lack comprehensive integration into real-world assembly processes, failing to address the inherent complexities and uncertainties involved in such operations.

RL allows robots to acquire skills through interaction with their environment, thereby eliminating the need for a predefined knowledge model for task planning. Its primary objective is to learn the optimal policy by maximising cumulative rewards for solving the problem of Markov Decision Processes. RL has been successfully implemented across various robotic applications, such as autonomous navigation [21] and manipulation [22]. Given the inherent stochastic decision-making associated with assembly tasks, these processes can also be effectively modelled as Markov Decision Processes [23]. This research aims to reduce the physical exertion on workers in HRCA tasks while simultaneously ensuring that both task order requirements and the agent's capabilities. Since RL's ability to learn and adapt through trial and error, we believe it offers a promising alternative to traditional approaches, potentially yielding more efficient solutions to mitigate workers' physical exertion.

2.2. Physical exertion assessment techniques

Manual workers, particularly those involved in physically demanding tasks such as assembly line work, are at high risk of suffering musculoskeletal disorders due to prolonged, repetitive movements, especially of the hands and arms [24]. While implementing effective fatigue/physical exertion management strategies is essential to reduce these risks, current methods of assessing physical exertion, including subjective self-reports and objective measurements, cannot provide continuous, automatic and precise assessment of physical exertion. Subjective assessments, such as the Borg RPE and CR10 scales [11], rely on personal perceptions, leading to inaccurate results and disrupting the workflow.

Objective methods, which include physiological, ergonomic, and biomechanical approaches, offer more reliable data but are influenced by various individual and environmental factors, limiting their accuracy. Physiological indicators like heart rate [25] and surface electromyography [26] are often used to estimate physical exertion, though individual physiological and biochemical factors, may impact the accuracy of results. Ergonomic techniques, such as Rapid Upper Limb and Rapid Entire Body Assessments [12], aim to minimise injury risk during repetitive tasks but often fail to account for human anatomical differences, leading to inconsistent assessments. Meanwhile, musculoskeletal models, like OpenSim [27], provide valuable insights by simulating bodily dynamics but are complex and resource-intensive, requiring extensive data input and high computational power [28]. Thus, in our previous work [14], we proposed a physical exertion assessment model capable of providing real-time muscle exertion estimation based on human movement data. We integrate this model into the current research to inform the reward function in our RL-based approach for physical exertion alleviation.

3. Problem formulation

In a robot–human assembly unit, we assume the presence of a robot r and an assembly worker h, collaborating sequentially to assemble n products P, each consisting of I parts. The agents operate in a turn-based manner, modelled as an agent-environment cycle game. The product's assembly sequence and task allocation are treated as stochastic decision-making processes. Thus, we model a product's human assembly operation process as a multi-agent Markov decision process, defined by a tuple $M = (S, A_h, A_r, P_a, \gamma, R)$. S represents the state space of the product assembly status. A are discrete operational action spaces performed by humans and robots on the product. $P_a(s'|s,a)$ is the transition probability from the state $s \in S$ to state $s' \in S$ under action a. $\gamma \in [0,1)$ is the discount factor. R represents the reward after the transition from state s to state s' by action $a \in A_r$. The reward function will be carefully designed, informed by the human worker's accumulated physical exertion.

We advocate for flexible HRC, where humans are not strictly obliged to follow the planner's suggested actions. This flexibility is important because different assembly workers may have varying preferences for assembly sequences, introducing an element of randomness. Therefore, in our work, we do not restrict the worker assembly sequence. To reduce workers' physical exertion in HRCA, our goal is to learn a policy π_r for the robot that prioritises tasks causing higher physical exertion for the worker, leaving less physically demanding tasks for the worker, thereby alleviating their physical exertion.

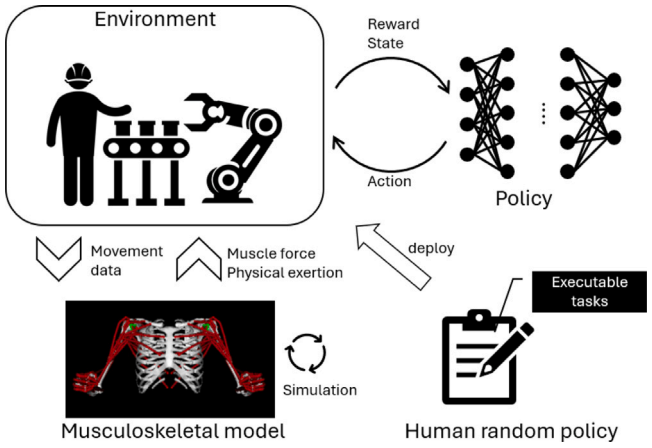

Fig. 1. The general framework of RL-based task planning method for worker's physical exertion alleviation in HRCA.

4. Method

A comprehensive HRCA framework for the proposed method has been developed to facilitate RL-based task planning for robots, with the primary objective of mitigating workers' physical exertion in HRCA, as depicted in Fig. 1. Central to this framework is a human–robot cell, where the agents collaborate on an assembly task. The movement of the human operator is monitored, and a musculoskeletal-based physical exertion estimation method is integrated, which provides an estimation of worker physical exertion by utilising the worker's movement data as input. The policy π_r, parametrised by a deep neural network, controls the robot at the task level. The training is performed through an iterative process, allowing agents to interact with the environment and learn optimal actions that maximise the cumulative reward R. To ensure π_r is adaptable to diverse worker behaviours, e.g. different assembly sequences, we implement a random policy π_h for the human in the training process, which randomly operates executable assembly tasks.

4.1. Physical exertion estimation

Static optimisation is a critical simulation-based method for muscle force estimation based on a worker's musculoskeletal model. The model is customised by scaling a general bimanual upper arm model [29] using the anthropometric data of the worker, which reflects individual differences. The Muscle force or activation at each specific moment is calculated based on the observed human motion, where the motion is captured using inertial measurement units (IMUs). In our work, a human muscle force-based method, leveraging static optimisation is adapted to assess worker physical exertion.

In addition to using IMUs, we employed a vision-based method [14] to determine whether a component is being held by the worker. If the component is held, its gravitational force is applied to the hand joints. Furthermore, we assume that contact forces during the assembly process were not considered in this study.

Static optimisation requires a heavy computation burden, making it unsuitable for real-time analysis. Using a neural network as a surrogate model for real-time analysis is a common approach. To efficiently estimate muscle forces on the upper body of a human operator, we employ an IK-BiLSTM-AM-based surrogate model, which integrates inverse kinematics (IK), bidirectional long short-term memory (BiLSTM), and an attention mechanism (AM), as proposed in our previous work [14]. This model approximates the complex biomechanics simulation for muscle force estimation.

A muscle force-based physical exertion estimation model [30] is then applied, which estimates physical exertion using historical force data via a first-order differential equation. The mathematical formulation of this model is as follows:

$$\frac{de_m(t)}{dt} = \begin{cases} (1 - e_m(t))\frac{f_m(t)}{c_m} & f_m(t) \geq f_{\text{th}} \\ -e_m(t)\frac{R}{c_m} & f_m(t) < f_{\text{th}} \end{cases} \tag{1}$$

where $e_m(t)$ represents the physical exertion of human muscle m, while $f_m(t)$ denotes the force exerted by muscle m at time t. The recovery coefficient R, set to 0.5, indicates the recovery rate from fatigue. The threshold of muscle force for muscle m is denoted by f_{th}. The capability coefficient of muscle m, denoted by c_m, reflects the muscle's resistance to fatigue.

4.2. DuelingDQN policy

In the HRCA framework, we adopt a DuelingDQN policy with action masking techniques (DuelingDQN-AM) for reliable task planning for physical exertion alleviation. DuelingDQN is an improved version of standard DQN [31], designed to perform well in environments with sparse reward structures. The rationale behind using DuelingDQN lies in its Q-value estimation mechanism, which is composed of two components: the state-value function and the advantage function. This design enables the model to evaluate the relative importance of actions a within a given state s, while simultaneously identifying the value of the state itself. The state-value stream in DuelingDQN enhances learning by decoupling the evaluation of states from the actions, allowing for a more focused and efficient representation of state values in the decision-making process. Given that our problem involves a multi-step, sparse reward scenario, this approach is considered a promising approach for physical exertion alleviation in HRCA.

State-value function, the first part of the DuelingDQN Q-value, represents the value of being in a particular state, regardless of the action taken. The advantage function represents the advantage of taking a specific action in that state relative to other actions. The Q value in DuelingDQN is represented as:

$$Q(s,a) = V(s) + \left(A(s,a) - \frac{1}{|A|} \sum_{a'} A(s,a') \right) \tag{2}$$

where $Q(s,a)$ is the Q-value for taking action a in state s; $V(s)$ is the value function that estimates the value of being in state s; $A(s,a)$ is the advantage function, which estimates the advantage of taking action a in state s; $|A|$ is the total number of actions. The training process of DuelingDQN for workers' physical exertion is illustrated in Fig. 2.

4.3. Action masking techniques

Action masking is a technique of RL that masks invalid actions in the action space, in order to improve training efficiency and reliability. In our case, certain actions are strictly prohibited in specific states due to factors such as task order requirements and the robot's capabilities. Task order requirements are defined by the logical sequence of assembly actions, which must comply with the geometric constraints inherent in the product design, thereby ensuring that each assigned action aligns with the established procedural constraints. The agent's capability, generally dictated by factors such as the robot arm's workload and the gripper's grasping capacity, imposes further limitations. Therefore, we introduce a mask $M(s,a)$ to account for these constraints. Thus, the Q value function is modified as follows:

$$Q_{\text{masked}}(s,a) = M(s,a) \cdot \left(V(s) + \left(A(s,a) - \frac{1}{|A|} \sum_{a'} A(s,a') \right) \right) + (1 - M(s,a)) \cdot \epsilon \tag{3}$$

where $M(s,a) = 0$ if action a is invalid in state s; $M(s,a) = 1$ if the action a is valid; ϵ is a small number. $\epsilon = -10^5$ in our case.

And-Or graph G is a common way to represent the product assembly task, decomposing complex assembly tasks into task units and defining

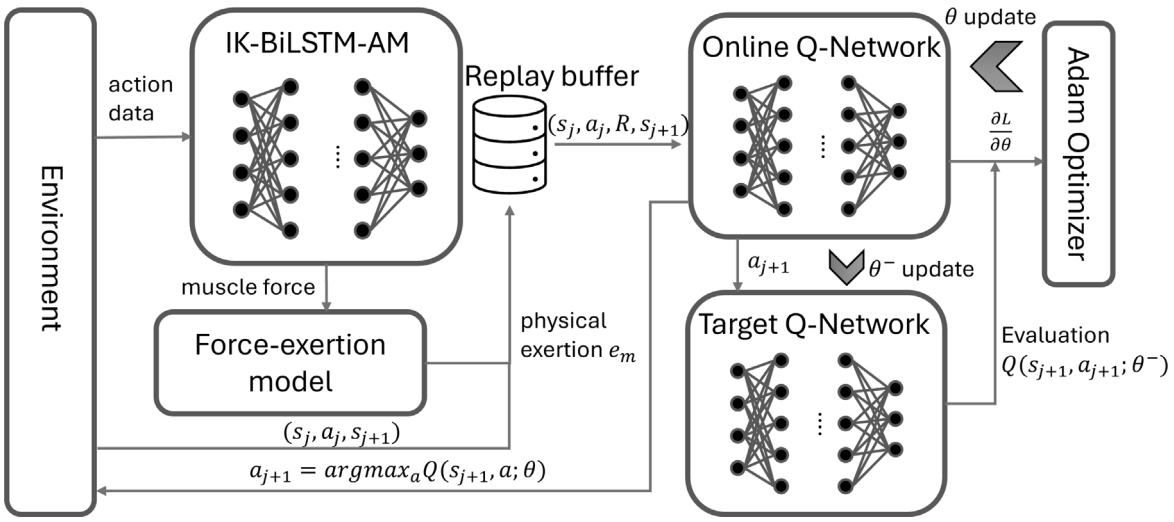

Fig. 2. The training process of the DuelingDQN with action masking techniques for alleviating workers' physical exertion is illustrated. An IK-BiLSTM-AM network, along with a force-exertion model, is employed to estimate physical exertion, using human motion data as input. The policy is trained using both the online Q-network and target Q-network, with the Adam optimiser facilitating updates. This trained policy enables the robot to execute optimal actions that minimise workers' physical exertion during HRCA.

the task constraint between them. An example of And-Or graph is shown in Fig. 4. Thus, the AND-OR graph is used to mask unavailable actions that violate task order constraints.

Based on the graph G, the executable task in the state s is defined: a collection of task units with no dependent tasks that have not yet been executed [32]. We define $E(s,a)$, where $E(s,a) = 1$ if action a is executable in state s; $E(s,a) = 0$ if it is not.

The robot's capability to perform action a is denoted as $C(a)$, where $C(a) = 1$ if the robot can perform action a; $C(a) = 0$ if it cannot. Using both the robot's capabilities and the assembly sequence requirements, we define the action mask as:

$$M(s,a) = \neg(E(s,a) \& C(a)) \tag{4}$$

where $M(s,a) = 1$ if action a is either not executable in the state s or beyond the robot's capabilities, ensuring adherence to task constraints and robot limitations.

4.4. Observation space, action space, and reward function

The observation space, action space and reward function for physical exertion alleviation using RL techniques are carefully defined in this section.

4.4.1. Observation space

The observation space of the robot consists of the status of the product being assembled and the accumulated physical exertion of the human operators. The status of the parts involved in the task is defined as s_i, and the accumulated physical exertion of muscle m is defined as e_m. Consequently, the observation space is defined as O.

$$O = [s_1, \dots s_i, \dots, s_I, e_1, \dots e_m \dots, e_M] \tag{5}$$

where s_i denotes the state of the ith part p_i. The state of the parts could be assigned with 3 cases: 0 for the parts that are to be assembled, 1 for those being assembled, and 2 for those that are already assembled. The physical exertion e_m is derived by Eq. (1). In our case, $M = 20$ selected muscles, located in the arms, shoulders, back, and chest on the right side of the upper body, are monitored.

4.4.2. Action space

The robot's action space is defined as A_r, including the operation of the product's parts. However, the robot has limitations in its assembly capabilities. Besides the action masking techniques, we penalise the policy if an allocated task exceeds the robot's capabilities.

$$A_r = [a_1, \dots, a_i, \dots, a_I, a_{\text{idle}}] \tag{6}$$

where a_i is the action that assembles the corresponding parts or tools p_i. a_{idle} means the agent is idle.

4.4.3. Reward function

The design of the reward function is paramount to the overall effectiveness of the RL policy, as it directly influences the agent's decision-making process. In this work, the reward function is crafted with several critical factors, namely the task order requirements, the agent's capabilities, and the physical exertion experienced by the human collaborator. Additionally, we punish the robot while the human collaborator conducts tasks that impose fatiguing physical exertion on them, aiming to reduce their workload. With these considerations in mind, the reward function, denoted as R, is designed to guide the agent towards optimal performance.

$$R = \begin{cases} r_c & \text{if an assembly task is completed.} \\ r_s & \text{if an assembly step is completed.} \\ \xi(\vartheta_i - \vartheta_{i-1}) & \text{if the physical exertion of the human worker increases} \end{cases} \tag{7}$$

where r_c is a completion reward if an assembly task is done; r_s is a reward if an assembly step is done.

In assembly tasks that require coordination among multiple muscle groups, relying on a single metric often fails to fully capture the complexity of physical exertion. For repetitive and intricate activities, integrating signals from multiple muscle groups offers a new perspective, enabling a more accurate assessment of overall fatigue caused by localised muscle fatigue [33]. Furthermore, peak muscle loading plays a critical role in musculoskeletal disorders, highlighting the necessity of considering maximum muscle exertion, as emphasised in [34].

This dual perspective underscores the importance of evaluating both overall and peak muscle exertion when assessing worker fatigue. Accordingly, Eq. (8) is formulated as follows:

$$\vartheta_i = \frac{1}{M}\sum_{m=1}^{M} e_m + \max_m e_m \tag{8}$$

which combines the average muscle physical exertion and max muscle physical exertion across M selected muscles. The accumulated increase

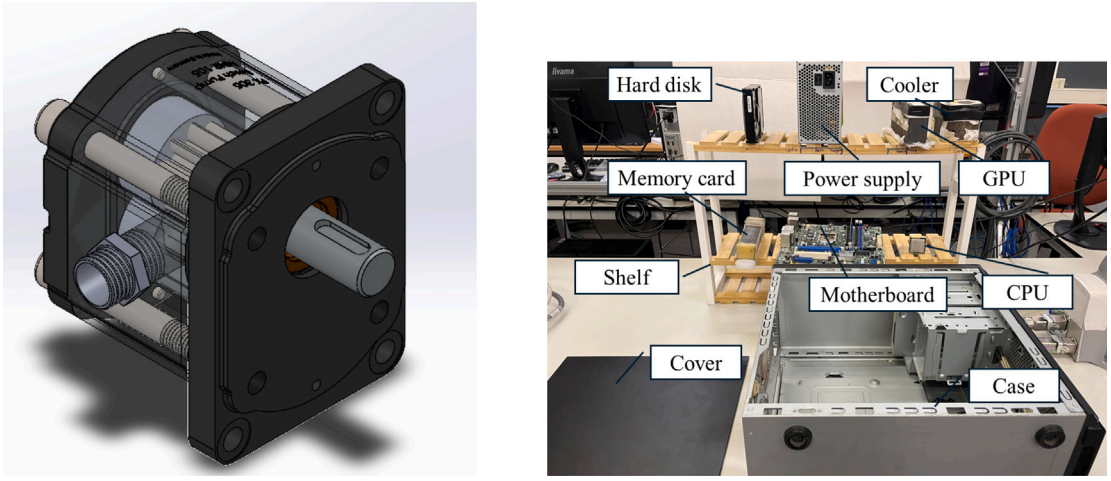

Fig. 3. The products with various complexities that are used in the experiment: (a) pump (36 parts), (b) PC desktop (9 parts).

at the step i in muscle exertion, $\vartheta_i - \vartheta_{i-1}$, is used as a penalty for the robot if the human worker's physical exertion rises. ξ is the weight variable.

4.5. Model training

Overall, the Dueling DQN algorithm is a reinforcement learning method, where data is collected through repeated interactions with the environment to train the model. The virtual environment is modelled as a Markov Decision Process, where the human and robot sequentially assemble parts in a turn-based manner (Section 3). A musculoskeletal model-based method (Section 4.1) is used to estimate human physical exertion based on movement data. This empirical estimation informs the reward function, guiding the policy towards minimising exertion. The Dueling DQN agent explores the environment by selecting actions based on an ϵ-greedy policy. A random policy is used to simulate human behaviour, ensuring adaptability to different assembly sequences. Each interaction generates a dataset containing state, action, and reward. The agent stores these state–action-reward transitions in a replay buffer and updates the Q-network through mini-batch sampling.

5. Experiment

This section introduces the experiments conducted to validate the proposed method. The evaluation consists of two experiments: (1) Experiment 1 makes a comparison study between DuelingDQN-AM and state-of-the-art methods in RL in a simulation environment, and (2) Experiment 2 is a proof-of-concept real assembly experiment to demonstrate the method's effectiveness in alleviating physical exertion involving multiple participants.

5.1. Experiment 1

5.1.1. Experiment setup

The objective of Experiment 1 is to evaluate the performance of the proposed DuelingDQN-AM on products of varying complexities and different cycles n of assembly products in a simulation environment. The performance is compared with state-of-the-art methods. Additionally, we assess the suitability of other baseline RL methods, which have been successfully applied in other domains, for addressing the fatigue alleviation challenges.

The products used in this experiment are illustrated in Fig. 3. Product complexity is defined by the number of parts, with a PC desktop consisting of 9 parts classified as an easy assembly task, and a pump consisting of 36 parts regarded as a difficult task. Muscle physical exertion is simulated, with values randomly generated between 0 and 0.3 across 20 muscles, encompassing the arms, shoulders, chest, and back of the upper body. They are: Triceps brachii: Long head, lateral head, medial head. Biceps brachii: Long head, short head. Deltoid: Anterior, middle, posterior. Pectoralis major: Clavicular head, sternal head, ribs head. Latissimus dorsi: Thoracic fibers, lumbar fibers, iliac fibers. Extensor carpi radialis longus. Extensor carpi radialis brevis. Extensor carpi ulnaris. Flexor carpi ulnaris. Pronator teres. Pronator quadratus.

Table 1
The parameters of DuelingDQN-AM for policy training.

Parameters	Value
Discount factor	0.9
Exploration rate	1 decay to 0.1
Batch size	64
Learning rate	1e−3
Estimation step	3
Target update frequency	300
Network architecture	multilayer perceptron
Hidden layer sizes	[128, 128, 128, 128]
weight variable ξ	20
r_c	10
r_s	−1

The parameters for DuelingDQN-AM are provided in Table 1. Each method was trained for 300 epochs for the PC desktop and 500 epochs for the pump task, with each epoch consisting of **1000** steps. All computations were performed on an NVIDIA GeForce RTX 3080 GPU.

5.1.2. Baselines

In our research, we carefully selected representative algorithms from the RL field for a comparative study. Given the discrete nature of the action space and observation space of the physical exertion alleviation problem, we chose prevailing RL algorithms such as DQN [35,36]. Additionally, based on prior studies demonstrating the effectiveness of network structures in DQN-based methods, including DQN-LSTM [37], DQN-Transformer [38], DQN-ResNet [39] and DQN-BiLSTM [40], Soft Actor-Critic (SAC), and Proximal Policy Optimization (PPO), these approaches were selected for comparison. To ensure a fair comparison, all methods employed the action masking technique. All methods were tested on both the PC desktop (easy task) and the pump (complex task). To systematically evaluate the impact of task repetition and combination on cumulative physical exertion over different time scales,

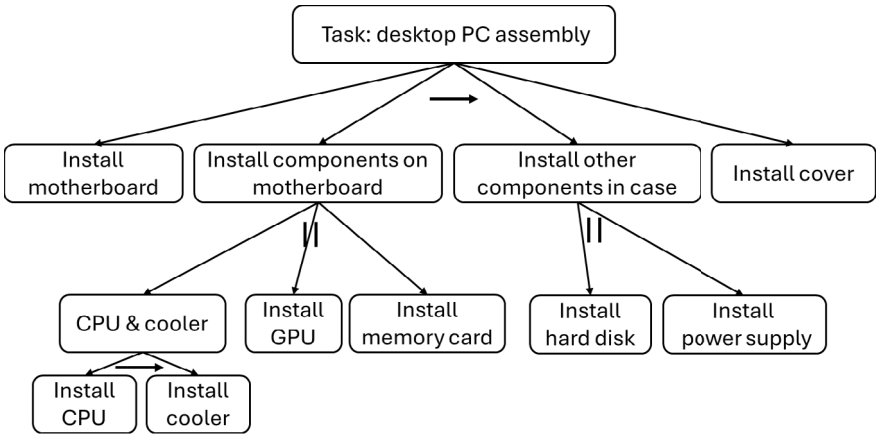

Fig. 4. An example of an And-Or graph. The graph illustrates an And-Or graph model for a desktop PC assembly task, consisting of And nodes and Or nodes, denoted by "→" and "‖" respectively. And nodes define sub-tasks that must be executed sequentially, while Or nodes define tasks that can be executed in parallel.

Table 2
The mean and standard deviation of 100 calculations of the trained policies of DuelingDQN-AM and the baseline across different product assembly rounds for PC desktops.

Methods	Round:1	Round:5	Round:10	Round:20	Round:50
DQN-BiLSTM	−57.26 ± 0.02	−285.83 ± 0.06	−570.91 ± 0.13	−1148.13 ± 0.11	−2855.71 ± 0.46
DQN-LSTM	−57.05 ± 0.07	−273.55 ± 0.26	−571.89 ± 0.7	−1143.46 ± 1.28	−2856.86 ± 0.9
DQN-ResNet	−28.71 ± 0.04	−139.13 ± 0.31	−279.05 ± 0.45	−558.83 ± 0.38	−1396.25 ± 2.03
DQN-Transform	−28.67 ± 0.03	−140.99 ± 0.11	−285.16 ± 0.13	−569.82 ± 0.38	−1436.89 ± 1.09
PPO	−108.44 ± 0.16	−295.01 ± 3.91	−1128.13 ± 1.96	−1155.28 ± 3.11	−5111.47 ± 9.86
SAC	−61.71 ± 0.27	−1865.93 ± 41.63	−25587.43 ± 210.44	−29167.83 ± 152.50	−6030.33 ± 109.97
DuelingDQN-AM	**−28.44 ± 0.03**	**−137.6 ± 0.06**	**−271.59 ± 0.28**	**−549.47 ± 0.32**	**−1359.98 ± 0.67**

Table 3
The mean and standard deviation of 100 calculations of the trained policies of DuelingDQN-AM and the baseline across different product assembly rounds for pump.

Methods	Round:1	Round:5	Round:10	Round:20	Round:50
DQN-BiLSTM	−184.39 ± 0.06	−921.82 ± 0.14	−1867.46 ± 0.62	−3722.62 ± 1.08	−8597.68 ± 3.11
DQN-LSTM	−178.3 ± 0.06	−931.24 ± 0.21	−1872.37 ± 0.1	−3735.82 ± 0.72	−8910.02 ± 2.16
DQN-ResNet	−61.18 ± 0.12	−303.27 ± 0.28	−600.74 ± 0.31	−1304.94 ± 3.42	−3026.87 ± 4.84
DQN-Transform	−60.07 ± 0.04	−305.51 ± 2.1	−618.29 ± 3.77	−1228.59 ± 3.89	−3082.81 ± 3.2
PPO	−186.74 ± 0.12	−944.64 ± 0.64	−2004.95 ± 4.27	−4489.00 ± 1.66	−9767.26 ± 19.91
SAC	−218.36 ± 0.38	−1121.17 ± 1.14	−2261.96 ± 0.72	−4562.64 ± 0.76	−11176.67 ± 3.70
DuelingDQN-AM	**−59.73 ± 0.04**	**−294.59 ± 0.16**	**−574.94 ± 0.19**	**−1174.6 ± 0.48**	**−2959.58 ± 1.61**

the performance of the algorithms was assessed across 1, 5, 10, 20, and 50 assembly cycles.

The evaluation metrics focused on the following:

1. Rewards, which refer to the accumulated reward achieved by the trained policies;
2. The physical exertion index ξ, to evaluate the level of fatigue alleviation experienced by workers under each method;
3. The number of invalid actions;
4. The number of successfully assembled products.

An ablation study was conducted to validate the importance of action masking in HRCA tasks. Specifically, we evaluated the performance of DuelingDQN in assembling a single unit of the PC desktop and the pump, with and without the action masking technique. An additional ablation study was conducted to validate the effectiveness of the advantage function, a key component of Dueling DQN, by comparing its performance with that of DQN. The network of Dueling DQN and DQN have the same structure. Both of them use the action masking techniques.

5.1.3. Results and discussion

Overall performance. The training process of the selected RL-based methods for handling physical exertion alleviation tasks of varying assembly difficulty is presented in Fig. 5, using different colours to distinguish between them. DQN-LSTM and DQN-BiLSTM did not converge for both products with varying numbers of assembly components, because LSTM and BiLSTM did not provide benefits for solving the physical exertion alleviation problem. LSTM and BiLSTM networks are better suited for capturing temporal dependencies in sequential data. However, our problem primarily focuses on state–action pair evaluation, which does not rely on long-term temporal dependencies but instead depends on the current fatigue state and the potential physical exertion induced by the task. This makes FN a more suitable approach for this problem.

Methods such as DQN-ResNet, and DQN-Transformer achieved results comparable to our proposed approach. The performance of PPO and SAC was suboptimal, as the rewards obtained were lower than those achieved by the DQN-based methods. The primary reason lies in the nature of our problem, where both the action space and observation space are discrete, and the rewards are sparse. These factors make DQN-based approaches more suitable, whereas PPO and SAC are less effective in this context.

Among all methods, DuelingDQN-AM demonstrated the best performance. Its advantages were particularly evident in its rapid convergence: regardless of product complexity or the number of assembly cycles n, the method consistently reached convergence the fastest. This is critical for RL tasks, as faster convergence significantly reduces training time. Additionally, we conducted 100 calculations using the trained policies to evaluate each policy. The results are presented in Table 2 (PC

Table 4
The experimental results for 100 calculations of trained DuelingDQN-AM policy include the following metrics. Rounds: The required number of executions of the task. Reward: The mean of the accumulated rewards. Invalid Action: The total number of invalid actions recorded. Completion: The total cycles of products successfully assembled. Steps: The total number of steps required to complete the required rounds.

Product	Methods	Rounds	Reward	Invalid action	Completion	Steps	Physical exertion index
PC desktop	DuelingDQN-AM	1	−28.44	0	1	6.1	1.62
		5	−137.6	0	5	34.3	7.67
		10	−271.59	0	10	75.2	14.82
		20	−549.47	0	20	134.1	30.32
		50	−1359.98	0	50	369.1	74.54
Pump	DuelingDQN-AM	1	−59.73	0	1	20	2.49
		5	−294.59	0	5	106.1	11.92
		10	−574.94	0	10	209.9	23.25
		20	−1174.6	0	20	405.1	48.48
		50	−2959.58	0	50	1057.1	120.12

Table 5
The ablation study of action masking techniques.

Product	Methods	Rounds	Action masking	Reward	Invalid action
PC desktop	DuelingDQN	1	Y	−28.44	0
PC desktop	DuelingDQN	1	N	−29.60	4.18
Pump	DuelingDQN	1	Y	−59.73	0
Pump	DuelingDQN	1	N	−128.00	4.88

desktop) and Table 3 (pump), with the best outcomes highlighted in bold. As shown, DuelingDQN-AM achieved the best performance across all rounds for both products. Our proposed method demonstrates both optimality and robustness, excelling in terms of convergence speed and reward attainment.

To further analyse the performance of DuelingDQN-AM, we use other metrics, as presented in Table 4. For both the PC desktop and pump, the number of steps and the physical exertion index increased approximately linearly with the number of assembly cycles. Besides, all required rounds of assembly were completed in each group. The average steps and physical exertion index per product were approximately 6.5 and 1.6 for the PC desktop, and 20 and 2.5 for the pump, respectively. This indicates that the policy's performance did not degrade with an increasing number of products, demonstrating the method's effectiveness in handling multiple assembly cycles.

Ablation study. The Table 4 shows that the number of invalid actions was consistently 0. In contrast, DuelingDQN without the action masking technique performed poorly, as shown in Table 5. Even within the assembly of a single unit, multiple invalid actions were observed (PC desktop: 4.18, pump: 4.88). This is unacceptable for collaborative robots in assembly lines, as invalid actions can result in defective products or even compromise safety. The comparison shows the importance of action masking techniques in this problem to ensure the validity of the policy action.

From Table 6, it can be observed that the Dueling DQN-AM method consistently outperforms DQN-AM across all assembly rounds and product types. For example, in the Desktop product assembly assembly round 50, DuelingDQN-AM achieves -1359.98 ± 0.67 compared to DQN-AM's -1405.71 ± 3.11. This indicates that the Dueling architecture contributes to the overall performance, leading to a more stable improvement in the reward.

Efficiency. We conducted 100 inferences using a well-trained model, with an average inference time of 0.00068 s for the desktop PC task and 0.00063 s for the pump task. The inference times are similar, as the model parameter sizes are 53,505 (desktop PC) and 55,809 (pump), respectively. These results demonstrate that the model can make decisions in near real-time.

5.2. Experiment 2

5.2.1. Experimental setup

In Experiment 2, a proof-of-concept PC desktop assembly experiment was conducted in a lab environment. This experiment received review and approval from the School of Engineering Research Ethics Committee (reference: 2023-PGR-YY-R1). The PC desktop used in the study, depicted in Fig. 3(b), consists of 9 components. The weights of these components were customised, ranging from 0.03 to 3 kg, with an average of 1.3 kg, to simulate the physical exertion experienced by assembly workers. The assembly requirement is shown in Fig. 4. The selected muscles are consistent with experiment 1. Besides, nine IMUs are strategically attached to specific body parts of each participant: the chest, scapula, upper arm, forearm, and hand. The working frequency of IMUs is 40 Hz. The hardware and software used in the experiment are listed in Table 7.

The experiment took place in a human–robot cell, which includes the parts to be assembled and the assembly areas, as shown in Fig. 6. A staff member, positioned near the assembly area, would disassemble the PC desktop once the assembly was completed, simulating a production line. The human and robot assembled the parts in turn. There is a keyboard for participants to inform the robot of the completion of the task.

The experiment included an experimental group and a control group. Participants were asked to assemble the PC desktop five times in each group. A total of 8 participants (6 male and 2 female, ages 23–35, height 160–185 cm.) were recruited to participate in both groups, and they were blind to the group assignment. Participants completed a questionnaire to subjectively evaluate their physical exertion in each group. The two questions in the questionnaire are both the Modified Borg Scale [41], with a score ranging from 0 to 10, to assess the physical exertion required to complete each group of experiments. We also evaluate the physical exertion of participants using our proposed methods. The effectiveness of the physical exertion alleviation method is validated by comparing the results obtained from our method with participants' subjective feedback.

In the experimental group, the robot operates using a policy trained with the DuelingDQN-AM method, incorporating empirical fatigue data. This policy prioritises tasks that are most fatiguing for humans, referred to as the **RL strategy**. In the control group, the robot operates using a random policy that selects executable tasks at random during the robot's turn, referred to as the **random strategy**. No specific assembly order is imposed on the human operator in either group, allowing them to choose their tasks based on personal preference. The experimental process for participants is listed in Table 8.

In the experiment, the hardware used included a KUKA iiwa LBR robot, a Robotiq 3f gripper, a Realsense D435 camera for object tracking, and an RTX 3080 GPU for algorithm training. Xsens awinda IMU was used for action tracking in the experiment. The screen was used to display the robot's operational status. The whole system was developed under the robot operation system (ROS). The human musculoskeletal model was visualised in OpenSim 4.4.

5.2.2. Result and discussion

In our work, eight participants completed the experiments, and Fig. 7 shows the process of the experiment. We conduct a comparative analysis and discussion between the results obtained from the questionnaire and those derived from our proposed method.

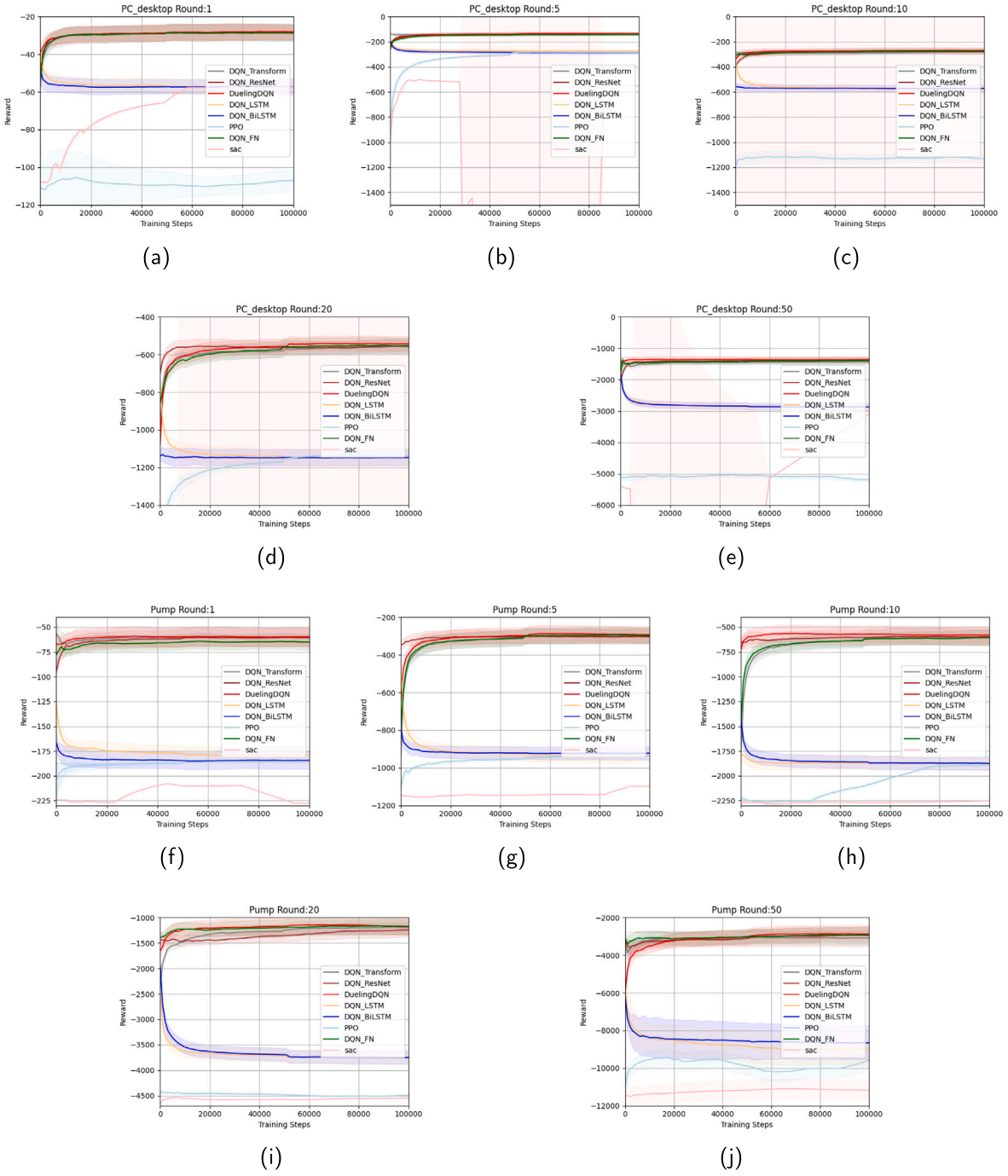

Fig. 5. The variation in the reward throughout the training process (0–100,000 steps, 0–100 epochs) is compared across the selected methods for both (a)–(e) desktop assembly tasks and (f)–(j) pump assembly tasks, evaluated over n=1, 5, 10, 20, and 50 execution rounds. Most methods converge within 100,000 steps, so only the results for 0–100,000 steps are presented.

Table 6
The ablation study of the Dueling architecture.

Methods	Product	Round:1	Round:5	Round:10	Round:20	Round:50
DQN-AM	Desktop	-28.88 ± 0.08	-143.52 ± 0.37	-278.22 ± 0.35	-562.95 ± 2.53	-1405.71 ± 3.11
DuelingDQN-AM	Desktop	-28.44 ± 0.03	-137.6 ± 0.06	-271.59 ± 0.28	-549.47 ± 0.32	-1359.98 ± 0.67
DQN-AM	Pump	-65.13 ± 0.06	-302.81 ± 1.55	-621.34 ± 3.37	-1184.13 ± 0.74	-2978.11 ± 4.09
DuelingDQN-AM	Pump	-59.73 ± 0.04	-294.59 ± 0.16	-574.94 ± 0.19	-1174.6 ± 0.48	-2959.58 ± 1.61

Table 7
Hardware and software used in the experiment.

Hardware	Software
Xsens IMU sensors	Xsens MT manager (Human movement data collection)
Kuka iiwa LBR robot	OpenSim (IK-BiLSTM-AM training data preparation)
Robotiq 3-Finger Robot Gripper	Pytorch (Dueling DQN modelling)
	Robot Operating System (Robot control)

Table 8
The experiment procedures for participants.

Procedure	Details
1	The participant sits quietly for five minutes to ensure they are well-rested.
2	The robot then randomly selects either the RL or random model and assembles the PC desktop 5 times with the participant.
3	After completing the previous phase, the participant takes a ten-minute break.
4	The robot then switches to the alternative mode and assembles the PC desktop 5 times with the participant.
5	The participant fills out a questionnaire. The experiment ends.

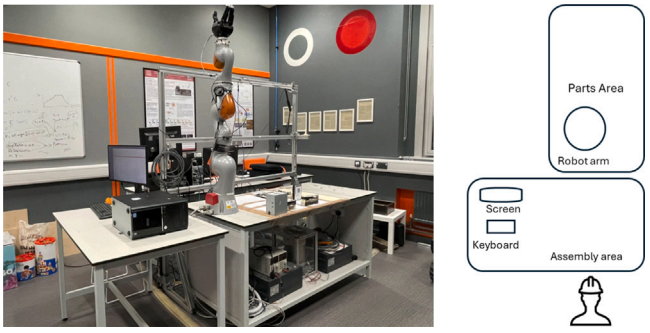

Fig. 6. The photo and layout diagram of the laboratory assembly line for Experiment 2.

1. Questionnaire: Participants provided subjective evaluations of their physical exertion using a modified Borg scale, with the results shown in Fig. 8(a). From the figure, it is evident that physical exertion under the RL strategy (mean: 2.31) is significantly lower than under the random strategy (mean: 4.56).
2. Our proposed method: The second part of the results is based on objective physical exertion analysis derived from the proposed method, as shown in Fig. 8(b). Similar conclusions can be drawn: physical exertion under the RL strategy (mean: 0.32) is 15.63% lower than that under the random strategy (mean: 0.37).

To further verify whether the RL strategy can alleviate human physical exertion, we performed a Mann–Whitney U test on the experimental results. The experiment hypothesises that the physical exertion under the RL strategy is significantly lower than that under the random strategy. Our results yielded P-values of $0.0456 < 0.05$ (Physical Exertion Index) and $0.0123 < 0.05$ (Borg Scale). The experimental results support the hypothesis, indicating that the fatigue under the RL strategy is significantly lower than that under the random strategy. These findings demonstrate that the proposed RL strategy effectively reduces physical exertion for workers performing the same tasks, as confirmed by both subjective reports and objective analysis. The authors believe that over longer work cycles, the reduction in physical exertion would become even more pronounced.

6. Conclusions

This study explored the application of RL methods to address the problem of physical exertion alleviation in HRCA. Based on the multi-agent modelling of this problem, we proposed a DuelingDQN approach combined with action masking to filter out invalid actions. We designed and conducted simulation experiments, demonstrating the advantages of DuelingDQN-AM over other methods in terms of convergence speed and stability across multiple cycles and products of varying complexity. Additionally, we conducted real-world experiments, where both Borg scale reports and physical exertion index confirmed that the RL strategy mitigate physical exertion by 15.63% compared to the random strategy.

As an exploratory study, the RL-based task planner proposed in this paper focuses primarily on mitigating human physical exertion. While this approach effectively addresses the physical exertion alleviation problem, it may not fully satisfy the multifaceted requirements of real-world HRC tasks. To overcome this limitation, future research will focus on developing a multi-objective task planner that incorporates additional factors such as time efficiency and safety into collaborative task planning. Furthermore, the RL model will need to be refined to operate within a multi-objective framework.

Additionally, the validation of this study is based on a proof-of-concept experiment conducted in a controlled laboratory setting. For broader applicability, validation in more complex, factory-like environments is essential. Future work should therefore also aim to evaluate the extended task planner on more complex tasks to assess its scalability and suitability for real-world industrial applications.

The muscle force estimation method does not account for contact forces during assembly, which may introduce inaccuracies in the estimated physical exertion. In future work, we will explore methods to address this issue.

CRediT authorship contribution statement

Yingchao You: Writing – original draft, Validation, Software, Methodology, Formal analysis, Conceptualization. **Ze Ji:** Writing – review & editing, Supervision, Resources, Methodology, Formal analysis, Conceptualization.

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.

Acknowledgements

Yingchao You thanks the Chinese Scholarship Council for providing the living stipend for his PhD programme (No. 202006020046).

Data availability

Data will be made available on request.

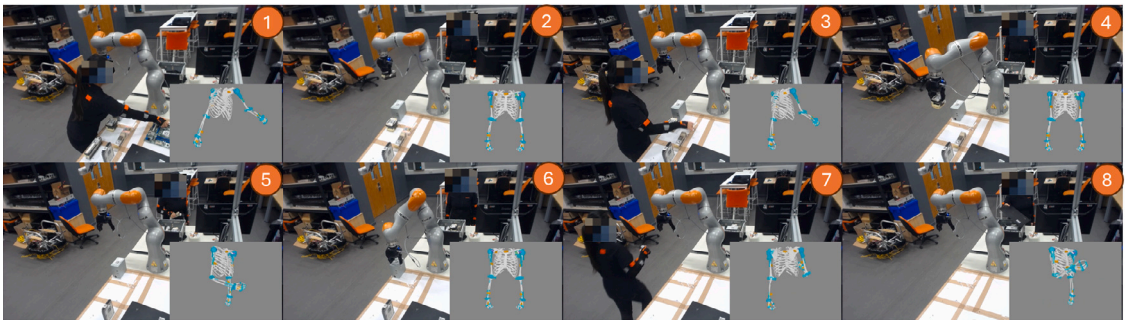

Fig. 7. The images from Experiment 2 illustrate an HRCA task under the RL strategy as follows. H: Motherboard, R: GPU, H: CPU, R: Cooler, H: Memory card, R: Power supply, H: Hard disk, R: idle H: Cover. Here, H represents tasks performed by the human operator, and R represents tasks performed by the robot. It is important to note that we assume that the robot is unable to assemble the motherboard and cover in our case due to its gripper capability. The human operator is capable of assembling all components.

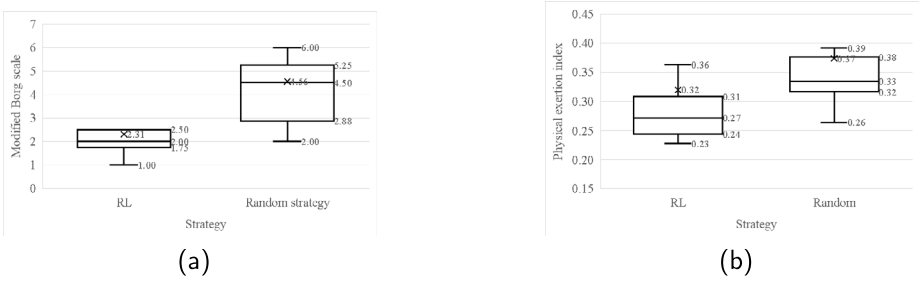

Fig. 8. Figure (a) shows a box plot of the results of the Modified Borg scale for 8 participants under the RL and random strategies. Figure (b) displays a box plot of the physical exertion index for 8 participants under the RL and random strategies.

References

[1] X. He, B. Xiao, J. Wu, C. Chen, W. Li, M. Yan, Prevalence of work-related musculoskeletal disorders among workers in the automobile manufacturing industry in China: a systematic review and meta-analysis, BMC Public Health 23 (1) (2023) 2042, Publisher: Springer.

[2] M.-L. Lee, S. Behdad, X. Liang, M. Zheng, Task allocation and planning for product disassembly with human–robot collaboration, Robot. Comput.-Integr. Manuf. 76 (2022) 102306, http://dx.doi.org/10.1016/j.rcim.2021.102306, URL: https://www.sciencedirect.com/science/article/pii/S0736584521001861.

[3] A.A. Malik, A. Bilberg, Complexity-based task allocation in human-robot collaborative assembly, Ind. Robot.: Int. J. Robot. Res. Appl. 46 (4) (2019) 471–480, Publisher: Emerald Publishing Limited.

[4] C. Zhang, W. Song, Z. Cao, J. Zhang, P.S. Tan, X. Chi, Learning to dispatch for job shop scheduling via deep reinforcement learning, Adv. Neural Inf. Process. Syst. 33 (2020) 1621–1632.

[5] Y.Y. Liau, K. Ryu, Genetic algorithm-based task allocation in multiple modes of human–robot collaboration systems with two cobots, Int. J. Adv. Manuf. Technol. 119 (11) (2022) 7291–7309, Publisher: Springer.

[6] F. Fusaro, E. Lamon, E.D. Momi, A. Ajoudani, An integrated dynamic method for allocating roles and planning tasks for mixed human-robot teams, in: 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN), 2021, pp. 534–539, http://dx.doi.org/10.1109/RO-MAN50785.2021.9515500.

[7] E. Merlo, E. Lamon, F. Fusaro, M. Lorenzini, A. Carfì, F. Mastrogiovanni, A. Ajoudani, An ergonomic role allocation framework for dynamic human–robot collaborative tasks, J. Manuf. Syst. 67 (2023) 111–121, http://dx.doi.org/10.1016/j.jmsy.2022.12.011, URL: https://www.sciencedirect.com/science/article/pii/S027861252200231X.

[8] W. Zheng, D. Wang, F. Song, OpenGraphGym: A parallel reinforcement learning framework for graph optimization problems, in: International Conference on Computational Science, Springer, 2020, pp. 439–452.

[9] P.T. Kyaw, A. Paing, T.T. Thu, R.E. Mohan, A.V. Le, P. Veerajagadheswar, Coverage path planning for decomposition reconfigurable grid-maps using deep reinforcement learning based travelling salesman problem, IEEE Access 8 (2020) 225945–225956, Publisher: IEEE.

[10] A.D. Workneh, M. El Mouhtadi, A. El Hilali Alaoui, Deep reinforcement learning for adaptive flexible job shop scheduling: coping with variability and uncertainty, Smart Sci. 12 (2) (2024) 387–405, Publisher: Taylor & Francis.

[11] N. Williams, The borg rating of perceived exertion (RPE) scale, Occup. Med. 67 (5) (2017) 404–405, Publisher: Oxford University Press UK.

[12] L. McAtamney, N. Corlett, Rapid upper limb assessment (RULA), in: Handbook of Human Factors and Ergonomics Methods, CRC Press, 2004, pp. 86–96.

[13] B. Rekiek, P. De Lit, A. Delchambre, Hybrid assembly line design and user's preferences, Int. J. Prod. Res. 40 (5) (2002) 1095–1111, Publisher: Taylor & Francis.

[14] Y. You, Y. Liu, Z. Ji, Human digital twin for real-time physical fatigue estimation in human-robot collaboration, in: 2024 IEEE International Conference on Industrial Technology (ICIT), 2024, pp. 1–6, http://dx.doi.org/10.1109/ICIT58233.2024.10541029.

[15] C. Petzoldt, M. Harms, M. Freitag, Review of task allocation for human-robot collaboration in assembly, Int. J. Comput. Integr. Manuf. 36 (11) (2023) 1675–1715, Publisher: Taylor & Francis.

[16] 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, http://dx.doi.org/10.1016/j.procir.2019.04.127, URL: https://www.sciencedirect.com/science/article/pii/S2212827119307413.

[17] M. Zhang, C. Li, Y. Shang, Z. Liu, Cycle time and human fatigue minimization for human-robot collaborative assembly cell, IEEE Robot. Autom. Lett. 7 (3) (2022) 6147–6154, http://dx.doi.org/10.1109/LRA.2022.3149058.

[18] B. Yao, X. Li, Z. Ji, K. Xiao, W. Xu, Task reallocation of human-robot collaborative production workshop based on a dynamic human fatigue model, Comput. Ind. Eng. (2023) 109855, Publisher: Elsevier.

[19] L. Peternel, C. Fang, N. Tsagarakis, A. Ajoudani, Online human muscle force estimation for fatigue management in human-robot co-manipulation, in: 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2018, pp. 1340–1346, http://dx.doi.org/10.1109/IROS.2018.8593705.

[20] C. Messeri, A. Bicchi, A.M. Zanchettin, P. Rocco, A dynamic task allocation strategy to mitigate the human physical fatigue in collaborative robotics, IEEE Robot. Autom. Lett. 7 (2) (2022) 2178–2185, http://dx.doi.org/10.1109/LRA.2022.3143520.

[21] K. Zhu, T. Zhang, Deep reinforcement learning based mobile robot navigation: A review, Tsinghua Sci. Technol. 26 (5) (2021) 674–691, Publisher: TUP.

[22] R. Jangir, G. Alenya, C. Torras, Dynamic cloth manipulation with deep reinforcement learning, in: 2020 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2020, pp. 4630–4636.

[23] C.F. Biemer, S. Cooper, Level assembly as a markov decision process, 2023, arXiv preprint arXiv:2304.13922.

[24] I. Onaji, D. Tiwari, P. Soulatiantork, B. Song, A. Tiwari, Digital twin in manufacturing: conceptual framework and case studies, Int. J. Comput. Integr. Manuf. 35 (8) (2022) 831–858, Publisher: Taylor & Francis.

[25] E.M. Argyle, A. Marinescu, M.L. Wilson, G. Lawson, S. Sharples, Physiological indicators of task demand, fatigue, and cognition in future digital manufacturing environments, Int. J. Hum.-Comput. Stud. 145 (2021) 102522, Publisher: Elsevier.

[26] J. Wang, M. Pang, P. Yu, B. Tang, K. Xiang, Z. Ju, Effect of muscle fatigue on surface electromyography-based hand grasp force estimation, Appl. Bionics Biomech. 2021 (2021) Publisher: Hindawi.
[27] C.L. Dembia, N.A. Bianco, A. Falisse, J.L. Hicks, S.L. Delp, Opensim moco: Musculoskeletal optimal control, PLoS Comput. Biol. 16 (12) (2020) e1008493, Publisher: Public Library of Science San Francisco, CA USA.
[28] H. Aftabi, R. Nasiri, M.N. Ahmadabadi, Simulation-based biomechanical assessment of unpowered exoskeletons for running, Sci. Rep. 11 (1) (2021) 1–12, Publisher: Springer.
[29] D.C. McFarland, E.M. McCain, M.N. Poppo, K.R. Saul, Spatial dependency of glenohumeral joint stability during dynamic unimanual and bimanual pushing and pulling, J. Biomech. Eng. 141 (5) (2019) 051006, Publisher: American Society of Mechanical Engineers.
[30] L. Peternel, N. Tsagarakis, D. Caldwell, A. Ajoudani, Robot adaptation to human physical fatigue in human–robot co-manipulation, Auton. Robots 42 (2018) 1011–1021, Publisher: Springer.
[31] M. Sewak, Deep Q network (DQN), double DQN, and dueling DQN, in: M. Sewak (Ed.), Deep Reinforcement Learning: Frontiers of Artificial Intelligence, Springer, Singapore, 2019, pp. 95–108, http://dx.doi.org/10.1007/978-981-13-8285-7_8.
[32] Y. Cheng, L. Sun, M. Tomizuka, Human-aware robot task planning based on a hierarchical task model, IEEE Robot. Autom. Lett. 6 (2) (2021) 1136–1143, Publisher: IEEE.
[33] M. Cifrek, V. Medved, S. Tonković, S. Ostojić, Surface EMG based muscle fatigue evaluation in biomechanics, Clin. Biomech. 24 (4) (2009) 327–340, Publisher: Elsevier.
[34] T.J. Armstrong, P. Buckle, L.J. Fine, M. Hagberg, B. Jonsson, A. Kilbom, I.A. Kuorinka, B.A. Silverstein, G. Sjogaard, E.R. Viikari-Juntura, A conceptual model for work-related neck and upper-limb musculoskeletal disorders, Scand. J. Work. Environ. Heal. (1993) 73–84, Publisher: JSTOR.
[35] B. Peng, Q. Sun, S.E. Li, D. Kum, Y. Yin, J. Wei, T. Gu, End-to-end autonomous driving through dueling double deep Q-network, Automot. Innov. 4 (3) (2021) 328–337, http://dx.doi.org/10.1007/s42154-021-00151-3.
[36] H.v. Hasselt, A. Guez, D. Silver, Deep reinforcement learning with double Q-learning, Proc. the AAAI Conf. Artif. Intell. 30 (1) (2016) http://dx.doi.org/10.1609/aaai.v30i1.10295, URL: https://ojs.aaai.org/index.php/AAAI/article/view/10295.
[37] Y. Lin, M. Wang, X. Zhou, G. Ding, S. Mao, Dynamic spectrum interaction of UAV flight formation communication with priority: A deep reinforcement learning approach, IEEE Trans. Cogn. Commun. Netw. 6 (3) (2020) 892–903, http://dx.doi.org/10.1109/TCCN.2020.2973376, URL: https://ieeexplore.ieee.org/abstract/document/8995473/authors#authors.
[38] F. Ding, Y. Yuan, L. Lv, R. Zhang, W. Zhou, Transformer-enhanced DQN approach for energy and cost-efficient large-scale dynamic workflow scheduling in heterogeneous environment, IEEE Internet Things J. (2024) 1, http://dx.doi.org/10.1109/JIOT.2024.3442997, URL: https://ieeexplore.ieee.org/abstract/document/10634877.
[39] Y. Wang, X. Zhou, H. Zhou, L. Chen, Z. Zheng, Q. Zeng, S. Cai, Q. Wang, Transmission network dynamic planning based on a double deep-q network with deep ResNet, IEEE Access 9 (2021) 76921–76937, http://dx.doi.org/10.1109/ACCESS.2021.3083266, URL: https://ieeexplore.ieee.org/abstract/document/9439918.
[40] Y. Huang, X. Wan, L. Zhang, X. Lu, A novel deep reinforcement learning framework with bilstm-attention networks for algorithmic trading, Expert Syst. Appl. 240 (2024) 122581, http://dx.doi.org/10.1016/j.eswa.2023.122581, URL: https://www.sciencedirect.com/science/article/pii/S095741742303083X.
[41] R.C. Wilson, P. Jones, A comparison of the visual analogue scale and modified borg scale for the measurement of dyspnoea during exercise, Clin. Sci. 76 (3) (1989) 277–282, Publisher: Portland Press Ltd..