J. Shanghai Jiao Tong Univ. (Sci.), 2022, 27(5): 602-613
https://doi.org/10.1007/s12204-022-2493-7

Generation Approach of Human-Robot Cooperative Assembly Strategy Based on Transfer Learning

LÜ Qibing (吕其兵), *LIU Tianyuan* (刘天元), *ZHANG Rong* (张 荣),
JIANG Yanan (江亚南), *XIAO Lei* (肖 雷), *BAO Jingsong** (鲍劲松)
(College of Mechanical Engineering, Donghua University, Shanghai 201620, China)

Abstract: In current small batch and customized production mode, the products change rapidly and the personal demand increases sharply. Human-robot cooperation combining the advantages of human and robot is an effective way to solve the complex assembly. However, the poor reusability of historical assembly knowledge reduces the adaptability of assembly system to different tasks. For cross-domain strategy transfer, we propose a human-robot cooperative assembly (HRCA) framework which consists of three main modules: expression of HRCA strategy, transferring of HRCA strategy, and adaptive planning of motion path. Based on the analysis of subject capability and component properties, the HRCA strategy suitable for specific tasks is designed. Then the reinforcement learning is established to optimize the parameters of target encoder for feature extraction. After classification and segmentation, the actor-critic model is built to realize the adaptive path planning with progressive neural network. Finally, the proposed framework is verified to adapt to the multi-variety environment, for example, power lithium batteries.

Key words: human-robot cooperation, strategy transfer, reinforcement learning, progressive neural network, power lithium battery
CLC number: TP 242.6 **Document code:** A

0 Introduction

In current small batch and customized production mode, the products, for example new-energy automobiles, have faster replacement speed and higher demand of customers. Due to the specific advantages of human and robot, the combination of human and robot can maximize the operation efficiency. In current era, human-robot cooperation (HRC) has good application effect in the scenes with high complexity or high flexibility, such as integral assembly of automobile.

The traditional HRC refers to that the robot assists human to complete the specific tasks by preprogramming. In this mode, the assembly system is sensitive to the environment, making it necessary to design corresponding control programs of robot for different assembly tasks. In flexible human-robot cooperative assembly (HRCA), the robot can adjust the behavior according to the change of environment and improve the robustness of assembly operation. Liu et al.[1] and Huang et al.[2] respectively analyzed the flexible HRCA process from the aspects of safety evaluation and intelligent human-robot interaction.

The expression of HRCA strategy has the characteristics of complexity, high abstraction and large time-variability. From the perspective of target domain samples, the samples or labels are few when new products appear, causing that the training process of model in target domain cannot be carried out. Besides, it is hard to apply the historical assembly knowledge into the coming assembly task due to the change of task scene. Therefore, the reuse of assembly knowledge is the key to improve the overall assembly responsiveness by comparing the distributions of the features between source domain and target domain.

The method of calculating the similarity between source domain and target domain is an important way to reuse assembly knowledge. Renu and Mocko[3] found that the retrieval of text-based assembly process and knowledge sharing could be realized by using the similarity algorithm of text. In the field of robotic control, transferring the weight parameters of intelligent model in multi-task environment is an approach to accelerating the training of model in new domain. This method can greatly reduce the training time of model

Received: 2021-09-04 **Accepted:** 2021-11-02
Foundation item: the National Key Research and Development Program of China (No. 2019YFB1706300), and the National Natural Science Foundation of China (No. 52075094)
***E-mail:** bao@dhu.edu.cn

and accumulate the skills learned by agent, further improving the responsiveness of system. In the process of skill learning, the reinforcement learning (RL) model is used as the agent for realizing the interaction with the environment.

In view of the above problems, this paper builds an encoder-decoder network to extract the assembly feature of part and design the specific cooperation strategies which will be stored in assembly strategy library. When the new assembly task appears, the target encoder is trained with the source encoder through RL training. Then the comprehensive evaluation process is conducted including parts, human and robot. By comparing the similarities of the assembly features between source domain and target domain, the HRC strategy is quickly retrieved. Due to the different assembly scenes, the trajectory of robot needs to be replanned according to the characteristics of current scene. In that, the actor-critic (AC) model is established with the progressive neural network (PNN) to realize the adaptive path planning of robot.

1 Related Work

1.1 HRCA

The mode of HRC fully combines the flexibility of human with the efficiency of robot. In the sharing environment of HRCA, cooperative security is the premise of efficient HRC. Table 1 shows the research status on the safety of HRC, which mainly concerns the system responsiveness and position accuracy prediction. Darvish et al.[4] proposed an HRC integrated system in order to improve the system's ability to deal with emergencies, while Abuduweili et al.[5] and Cheng et al.[6] proposed a trajectory prediction model to ensure the safety of robot's motion path. Liu and Wang[7] and Amorim et al.[8] studied the collision-free HRC method and took into account the safety and efficiency of cooperation based on accurate positioning towards human target position.

The application research of HRC in assembly field is shown in Table 2. Makris et al.[9] used augmented reality (AR) tools to improve the efficiency of workers in obtaining task information and decision-making, while Raatz et al.[10] used genetic algorithm to assign assembly tasks to human and robot based on analyzing human-robot ability. Besides, the recognition of human intention in assembly process is the key to plan HRC strategy. For example, Liu and Wang[11] and Berg et al.[12] improved the flexibility of assembly activities based on hidden Markov model.

Table 1 Research status on safety of HRC

Author	Time	Main method	Main conclusion
Darvish et al.[4]	2018	HRC integrated system	Improving worker's comfort and dealing with emergencies
Abuduweili et al.[5]	2019	Multi-task model	Reducing the error of trajectory prediction in cooperation
Cheng et al.[6]	2020	Plan identification and trajectory prediction module	Generating safe and effective motion of robot
Liu and Wang[7]	2021	Collision-free cooperation system	Considering the safety and efficiency of HRC
Amorim et al.[8]	2021	Online collision avoidance method	Accurately capturing the position of human body, strong robustness

Table 2 Research status of HRCA

Author	Time	Main method	Main conclusion
Makris et al.[9]	2016	AR integrated HRC	Improving decision-making efficiency
Raatz et al.[10]	2020	Genetic algorithm	Adaptive task allocation
Liu and Wang[11]	2017	Hidden Markov model	Improving the flexibility of assembly activities
Berg et al.[12]	2018	Hierarchical hidden Markov model	Intention understanding and safe path planning

In digital assembly environment, the real-time perceived environmental data optimizes the decision-making of HRC. The existing research covers sequence planning, task allocation, intention recognition, etc. However, the research on the expression and transfer of complex strategy information is rare.

1.2 Knowledge Reuse Based on Transfer Learning

Transfer learning is the transfer of general principles, methods, strategies and attitudes learned in one domain to another domain, which is widely used in the field of image processing, text, voice and others. For the reuse of product assembly knowledge, Li et al.[13] used the framework structure to represent the structure configuration tree of product and stored the design knowledge through the database for reusing the historical design knowledge. He et al.[14] proposed a novel triple deep workflow model for production decision support problem solving (P-DSPS) and found that the assembly knowledge in similar domains can be extracted or retrieved. The above methods have implemented the reuse of knowledge, but the transfer of robotic skills has not been fully studied.

In HRCA, the operation strategy required by the robot to solve the corresponding task mainly comes from two aspects: ① learning operation strategy through interaction with assembly environment; ② operation strategy transferred from other domains. Arana-Arexolaleiba et al.[15] proposed an operation system supporting RL, in which the robot learned the task trajectory from human experts and continuously improved its performance over time. Meanwhile, Raziei and Moghaddam[16] developed and tested a hyper-actor AC framework based on task modularization and transfer learning. This method can effectively migrate the strategies learned in historical tasks to new tasks which can expand the scope and flexibility of intelligent agent. Rodríguez et al.[17] matched the new task execution diagram to obtain similar assembly constraints, so as to realize the process of pattern recognition and classification based on storing semantic assembly constraints.

The core of knowledge or skill transfer lies in the similarity score between domains, but the current stage of similarity calculation is not used in the migration process of HRC strategy. Therefore, the cross-domain transfer of HRCA strategies is a key research issue which can be realized by the comparison of assembly features between source domain and target domain.

2 Adaptive HRCA Approach

Figure 1 shows the proposed HRCA framework which realizes the reuse of assembly knowledge from source domain to target domain. The framework consists of three main modules: ① Expression of HRCA strategy; ② Transferring of HRCA strategy; ③ Adaptive motion planning of HRCA.

For a specific assembly task, it can be divided into many subtasks according to the constituent elements. Using the feature extraction model, the main assembly features of parts can be obtained, such as the threaded hole and threaded axis. The mapping rule between assembly features and assembly strategies is constructed by evaluating the attributes and defining corresponding HRCA strategy. In that, the HRCA strategy is

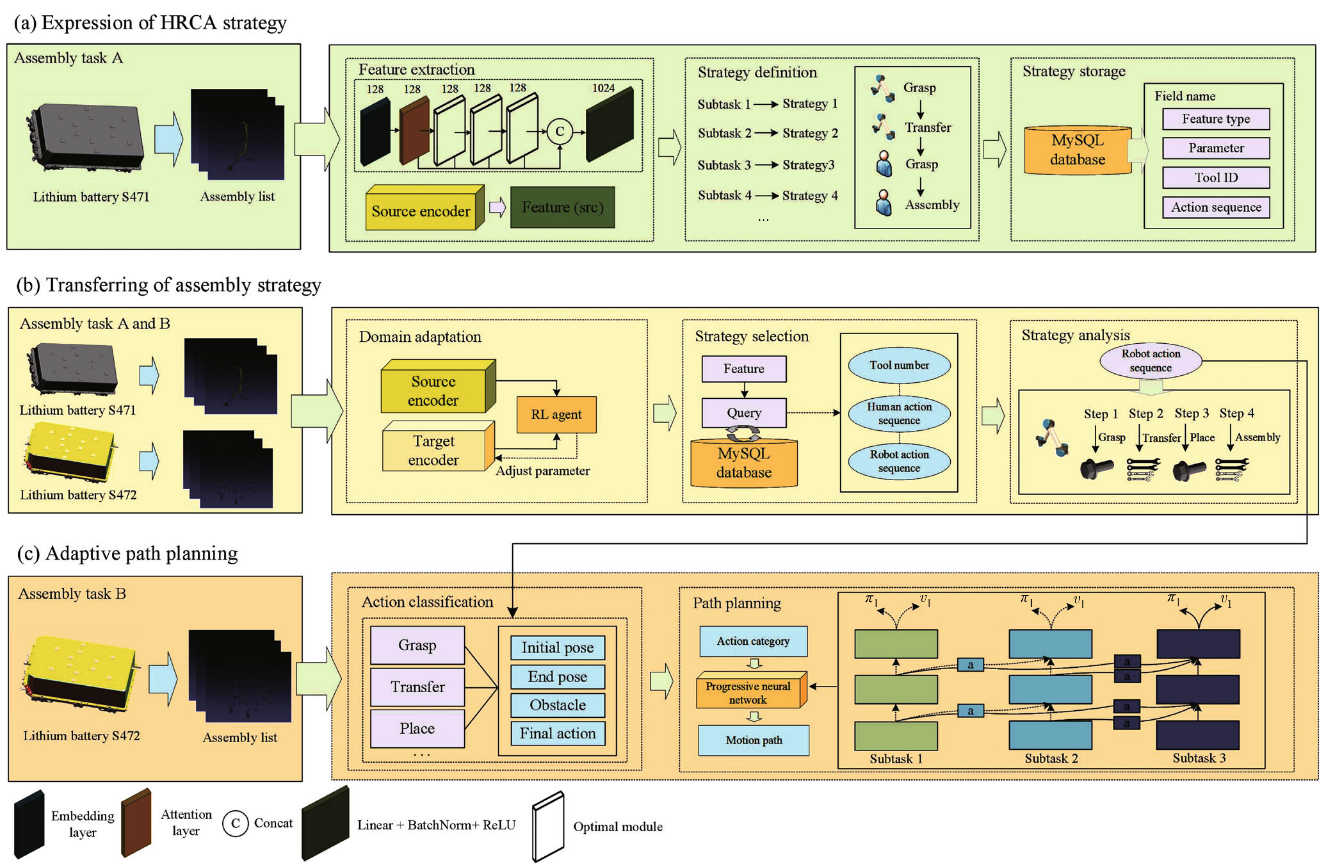
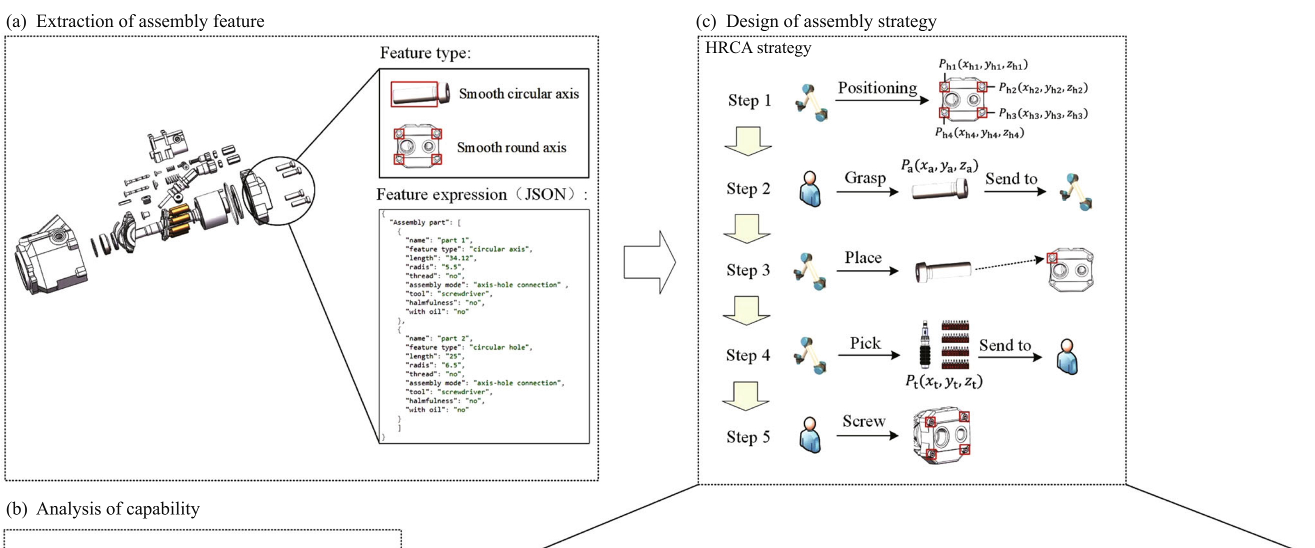


Fig. 1 HRCA framework with strategy transfer

designed according to the feature characteristics, such as feature type and parameters. The combination of feature type, feature parameters, assembly tool and action sequence are stored in MySQL database. For achieving the recognition of assembly features in target domain, the RL model is established to dynamically adjust the parameters of target encoder. When the feature distribution in source domain and the feature distribution in target domain are close, the classification and segmentation decoder in source domain can be used to process the assembly features in target domain. Through searching in the defined database, the assembly tool and human-robot action sequence corresponding to current task can be obtained. After action classification, the actions which are executed by robot mainly contain four elements: initial pose, end pose, obstacle and final action. Corresponding to specific action category, AC model is adopted in order to plan the safe trajectory of robot. For shortening the training time of intelligent agent in similar subtasks, the PNN is introduced in this architecture to reuse the pre-training model in historical subtasks and realize the continuous learning of agent.

2.1 Expression of HRCA Strategy

The HRCA strategy reflects the action relationship among assembly parts, assembly objects and assembly tools. The process of obtaining the assembly strategy can be divided into three steps: ① obtaining assembly feature; ② ability analysis of workers and robots; ③ design of HRCA strategy. As shown in Fig. 2, there will be multiple effective HRC action sequences for a specific assembly task. In that, the analysis of ability towards workers and robots determines the design scope of HRC strategy.



Strategy 1	Action sequence	Grasp part → Send part → Place part → Pick tool → Send tool → Screw
	Motion list	$\mathrm{Path}_{\mathrm{grasp}} \rightarrow \mathrm{Path}_{\mathrm{send\text{-}r}} \rightarrow \mathrm{Path}_{\mathrm{place}} \rightarrow \mathrm{Path}_{\mathrm{pick}} \rightarrow \mathrm{Path}_{\mathrm{send\text{-}h}} \rightarrow \mathrm{Path}_{\mathrm{screw}}$
Strategy 2	Action sequence	Grasp part → Send part → Place part → Pick tool → Send tool → Screw
	Motion list	$\mathrm{Path}_{\mathrm{grasp}} \rightarrow \mathrm{Path}_{\mathrm{send\text{-}h}} \rightarrow \mathrm{Path}_{\mathrm{place}} \rightarrow \mathrm{Path}_{\mathrm{pick}} \rightarrow \mathrm{Path}_{\mathrm{send\text{-}h}} \rightarrow \mathrm{Path}_{\mathrm{screw}}$
Strategy 3	Action sequence	Grasp part → Send part → Place part → Pick tool → Send tool → Screw
	Motion list	$\mathrm{Path}_{\mathrm{grasp}} \rightarrow \mathrm{Path}_{\mathrm{send\text{-}h}} \rightarrow \mathrm{Path}_{\mathrm{place}} \rightarrow \mathrm{Path}_{\mathrm{pick}} \rightarrow \mathrm{Path}_{\mathrm{send\text{-}h}} \rightarrow \mathrm{Path}_{\mathrm{screw}}$
Strategy 4	Action sequence	Grasp part → Send part → Place part → Pick tool → Send tool → Screw
	Motion list	$\mathrm{Path}_{\mathrm{grasp}} \rightarrow \mathrm{Path}_{\mathrm{send\text{-}r}} \rightarrow \mathrm{Path}_{\mathrm{place}} \rightarrow \mathrm{Path}_{\mathrm{pick}} \rightarrow \mathrm{Path}_{\mathrm{send\text{-}r}} \rightarrow \mathrm{Path}_{\mathrm{screw}}$

Fig. 2 Expression of HRCA strategy

(1) Extraction of assembly feature. The assembly properties of two adjacent components are composed of feature type, feature parameters, assembly tool and assembly mode. As shown in Fig. 3, the encoder-decoder network[18] is used to extract the assembly feature which contains multi-attention layers. The encoder network first converts the point cloud into a high-dimensional vector to describe the semantic association through the embedding layer. Then the feature extraction is realized by connecting several attention layers. The decoder is adopted to process the obtained feature vectors and output feature type and point set after segmentation. In feature extraction, the attention module is replaced by self-attention module for obtaining the relationship between points with surrounding points. For the point set, the dimensions of features can be obtained by calculating the coordinates of feature points, for example the length and width of circular hole.

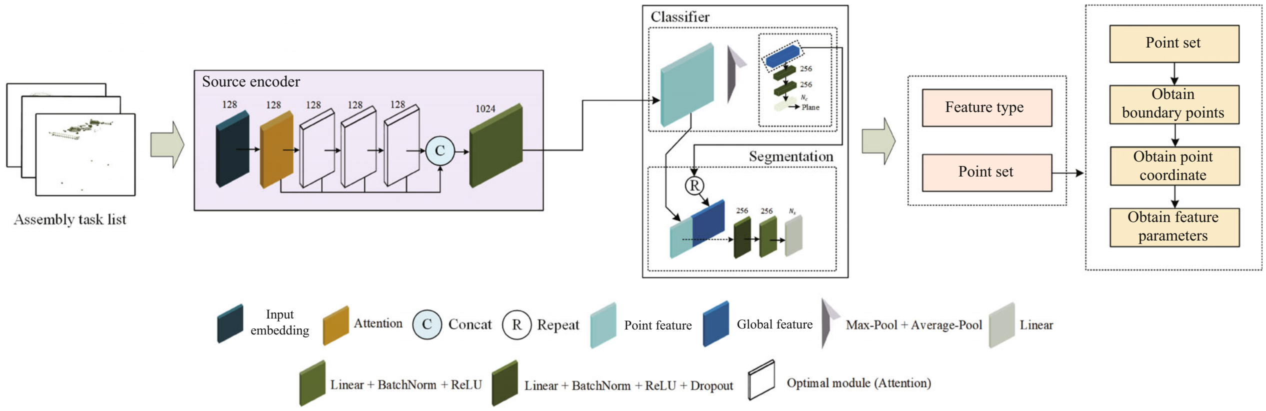


Fig. 3 Encoder-decoder network[18] for assembly feature processing

The assembly features can be roughly divided into the following types: hole, axis, slot, cone, buckle and others. In that, the connection between axis and hole is the main assembly approach. According to the thread feature on the interface of axis or hole, the assembly features can be further divided as threaded axis, smooth axis, threaded hole and smooth hole. In Fig. 4, the defined assembly mode has 8 types and the defined assembly tool has 4 types. According to the category and parameters of assembly feature, the assembly method and tools used for current assembly can be got by the reference manual. Besides, the oiliness and harmfulness of part will be recorded in the Json file.

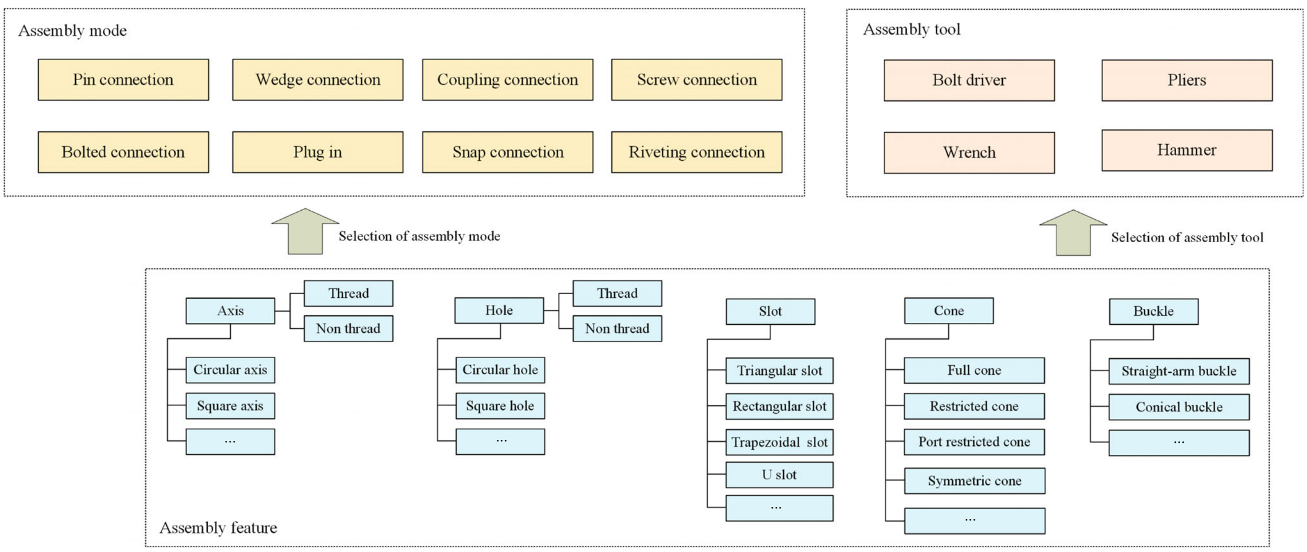


Fig. 4 Relationship among assembly feature, assembly mode and assembly tool

(2) Analysis of capability. HRCA requires that human and robot need to undertake part of work in collaborative space. The assignment of tasks is based on the ability analysis of human and robot. The indicators of capability analysis for human and robot are different which are designed by manual in advance. For human, the evaluation dimensions mainly include: ① grasping range of human (R_h); ② grasping load of human (L_h); ③ flexibility of human ($F_\mathrm{h} = 1$). For robot, the evaluation items are expressed as the following items: ① grasping range of robot (R_r); ② grasping load of robot (L_r); ③ flexibility of robot ($F_\mathrm{r} = 0$). To measure the ability of human and robot in handling current assembly task, the evaluation is executed from the distance between assembly part and objects (d_po), the weight of assembly part (W_p), and the flexibility of assembly part (F_p). Finally, the feasibility of human and robot in handling current assembly task will be obtained:

$$\mathrm{Feasible}_\mathrm{h} = \begin{cases} \text{true}, & \text{if } d_\mathrm{po} < R_\mathrm{h} \text{ and } W_\mathrm{p} < L_\mathrm{h} \\ \text{false}, & \text{else} \end{cases}, \quad (1)$$

$$\mathrm{Feasible}_\mathrm{r} = \begin{cases} \text{true}, & \text{if } d_\mathrm{po} < R_\mathrm{r} \text{ and } W_\mathrm{p} < L_\mathrm{r} \\ & \text{and } F_\mathrm{p} = 0 \\ \text{false}, & \text{else} \end{cases}. \quad (2)$$

When both human and robot can handle the current task, the robot will give priority to this task.

(3) Design of assembly strategy. After analyzing the ability of human and robot, assembly tasks can be further assigned to human and robot. The assembly strategy mainly contains four parts: operation subjects, executed action, executed trajectory and executed sequence. Figure 5 shows the human-focused and robot-focused assembly strategies. When Feasible_h is true, the category of current assembly subtask can be defined as human-focused assembly task. When Feasible_h and Feasible_r are both true, the assembly task is preferentially determined as robot-focused assembly task. In human-focused assembly task, human operator will execute the final assembly process. As shown in Fig. 5, the main cooperative strategy can be selected from the provided three forms. Due to the differences on assembly scenes, the moving trajectory needs to be planned according to the specific assembly environment.

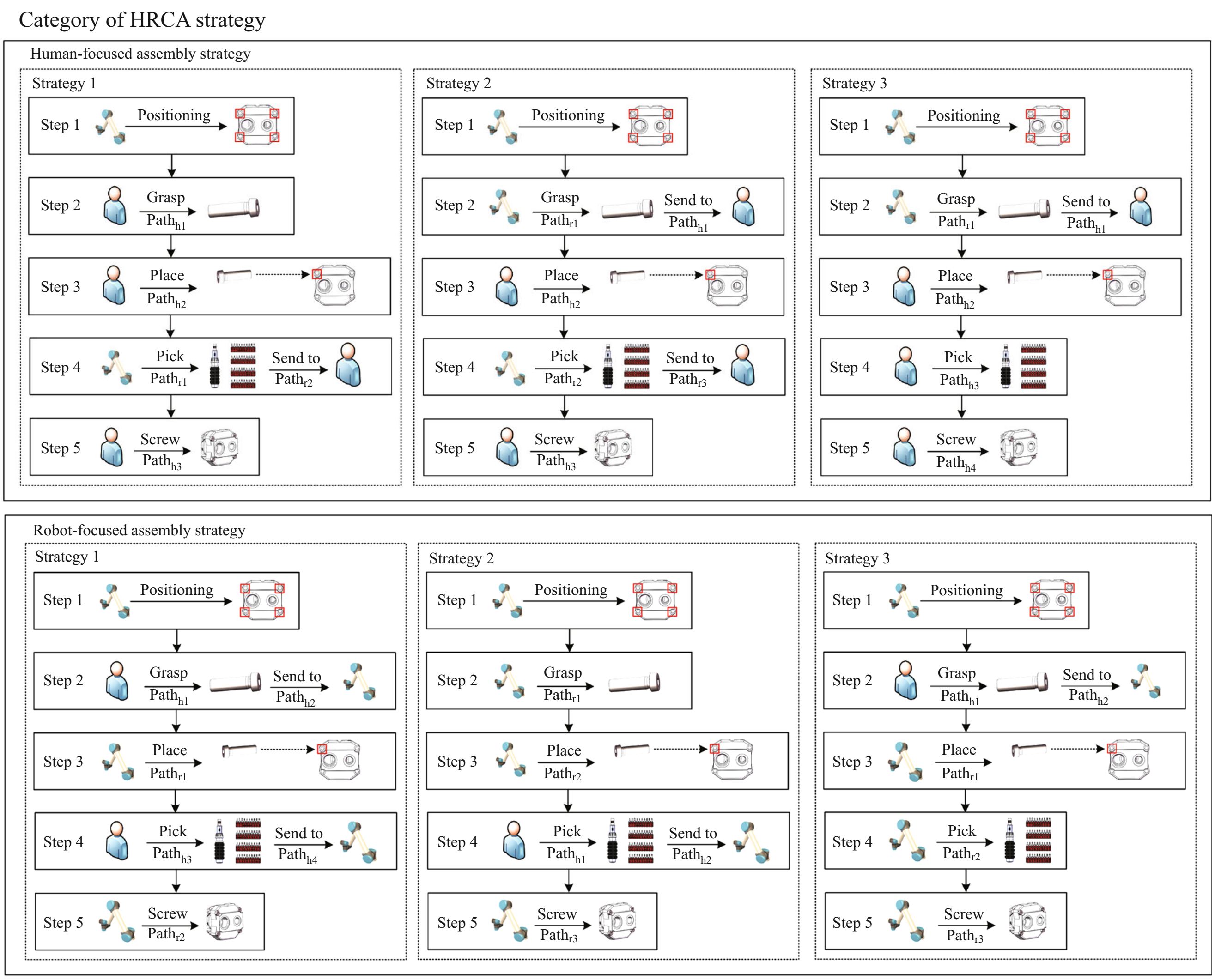


Fig. 5 Human-focused and robot-focused assembly strategies

2.2 Transfer of HRCA Strategy

Because different types of products have differences in component structure, component size and assembly sequence, the assembly strategy needs to be redesigned according to the specific assembly task. The trained encoder-decoder network in source domain cannot be directly used to identify the assembly features of parts in target domain. At the same time, the characteristics of small sample and few labels in target domain also lead to the failure of normal training in target domain. Based on copying the encoder network in source domain, the parameters of network should be further adjusted for processing the assembly features in target domain. As shown in Fig. 6, RL is established to dynamically adjust the weight parameters of encoder. In that, deep Q-learning (DQN) model is the agent which

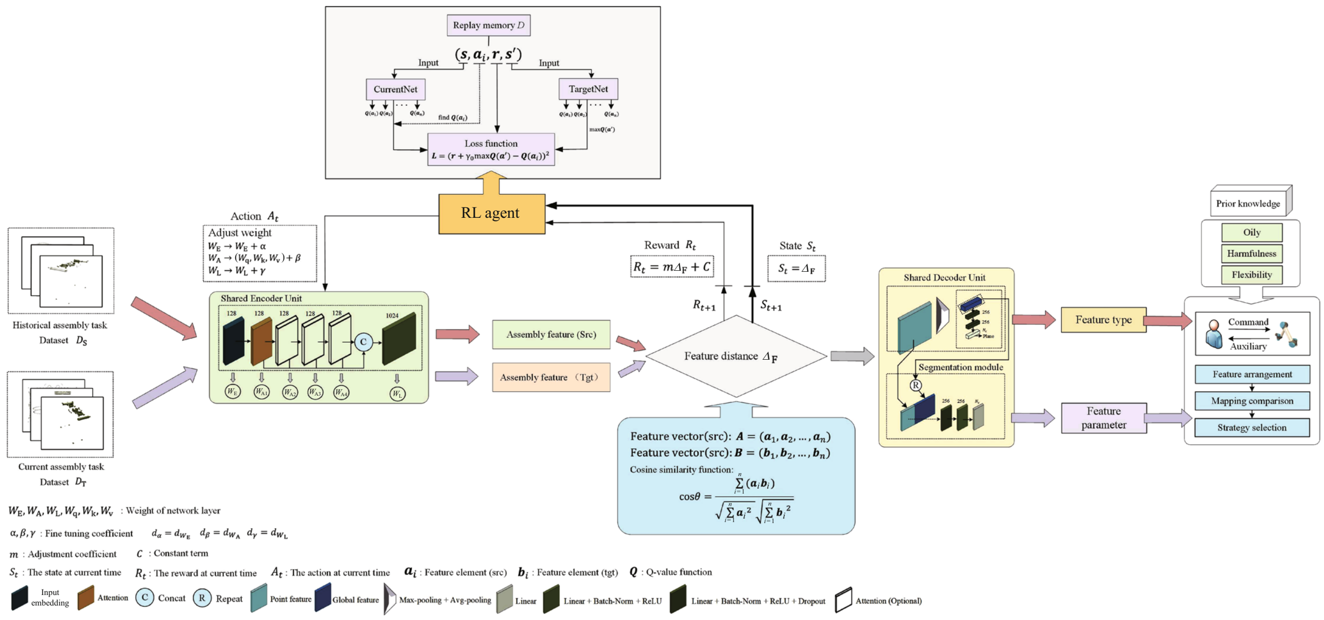


Fig. 6 Cross-domain recognition of assembly feature based on RL

takes actions and changes the state of encoder network. So the action of agent is defined:

$$\boldsymbol{A} = (a_1, a_2, a_3), \tag{3}$$

$$a_1 = W_{\mathrm{E}} + \alpha, \quad a_2 = W_{\mathrm{A}} + \beta, \quad a_3 = W_{\mathrm{L}} + \gamma, \tag{4}$$

$$\boldsymbol{W}_{\mathrm{A}} = (W_{\mathrm{q}}, W_{\mathrm{k}}, W_{\mathrm{v}}), \tag{5}$$

where, W_{E}, W_{A} and W_{L} respectively stand for the weights of embedding layer, attention layer and linear layer; the dimension $d_\alpha = d_{W_{\mathrm{E}}}$, $d_{\beta_1} = d_{W_{\mathrm{q}}}$, $d_{\beta_2} = d_{W_{\mathrm{k}}}$, $d_{\beta_3} = d_{W_{\mathrm{v}}}$, $d_\gamma = d_{W_{\mathrm{L}}}$; α, β, γ respectively represent the offsets of the weights towards the embedding layer, attention layer and linear layer; $W_{\mathrm{q}}, W_{\mathrm{k}}, W_{\mathrm{v}}$ are the weights of query, key and value, respectively. When adjusting the action, the variation range of α, β, γ is expressed as adding or subtracting one from each element of matrix.

The state of environment is related to the feature distance ($\varDelta_{\mathrm{F}}$) between source domain (D_{S}) and target domain (D_{T}). Meanwhile, the assembly features extracted by the encoder is described as an array vector which consists of feature elements ($\boldsymbol{a}_i$). For measuring the similarity of assembly features in different domains, the cosine similarity evaluation function is adopted as the measurement matrix:

$$S_t = \varDelta_{\mathrm{F}}, \tag{6}$$

$$\varDelta_{\mathrm{F}} = \cos\theta = \frac{\sum_{i=1}^{n} (\boldsymbol{a}_i \boldsymbol{b}_i)}{\sqrt{\sum_{i=1}^{n} \boldsymbol{a}_i^2} \sqrt{\sum_{i=1}^{n} \boldsymbol{b}_i^2}}, \tag{7}$$

where, the variation range of $\cos\theta$ is $[-1, 1]$ and the similarity reaches the maximum if the value is equal to 1; $\boldsymbol{b}_i$ is the assembly feature elements in target domain.

When the state of environment changes from S_t to S_{t+1}, a reward (R_{t+1}) will be generated to evaluate the effect of action taken by the RL agent. The reward function is designed:

$$R_t = m\varDelta_{\mathrm{F}} + C, \tag{8}$$

where, m is the equilibrium coefficient (set by experiment); C is the constant coefficient for initialing the reward value.

In Fig. 6, the adopted DQN model includes two layers: main network and target network. Through the mechanism of memory replay, these two networks will be trained by the loss function:

$$\boldsymbol{L} = (\boldsymbol{r} + \gamma_0 \max \boldsymbol{Q}(\boldsymbol{a}_i') - \boldsymbol{Q}(\boldsymbol{a}_i))^2, \tag{9}$$

where, γ_0 stands for the discounting factor of RL; $\boldsymbol{r}$ is the reward value of R_t; $\boldsymbol{Q}$ is the Q function; $\boldsymbol{a}_i'$ presents the next action taken by the RL agent.

After RL training stage, the extracted feature of assembly tasks in target domain can be automatically recognized by classification and segmentation module from source domain. Finally, the elements of assembly feature will be output, including feature type and parameters. Based on the assembly features, the corresponding assembly mode and tools are deduced. Then the HRC action sequence including operation subject, operation action, operation path and execution sequence is obtained by comparing the attributes of the assembly parts in two domains. In that, the spatial distance between assembly part and operators is input by visual computing and the weight of part is priori knowledge.

2.3 Adaptive Motion Planning of HRCA

For the specific assembly task, the HRCA strategy can be expressed as a set of action sequence. Each motion in this sequence contains its initial pose, end pose, running path and final action. Unlike source domain (D_S), the path corresponding to the same type of motion is not same in target domain (D_T). For example, grasping screwdriver and grasping pliers are different actions under the same type of motion, which requires the agent to retrain the model. For enabling the agent to quickly plan the path in target domain, transferring the empirical parameters of model from source domain to target domain can play a key role, as shown in Fig. 7.

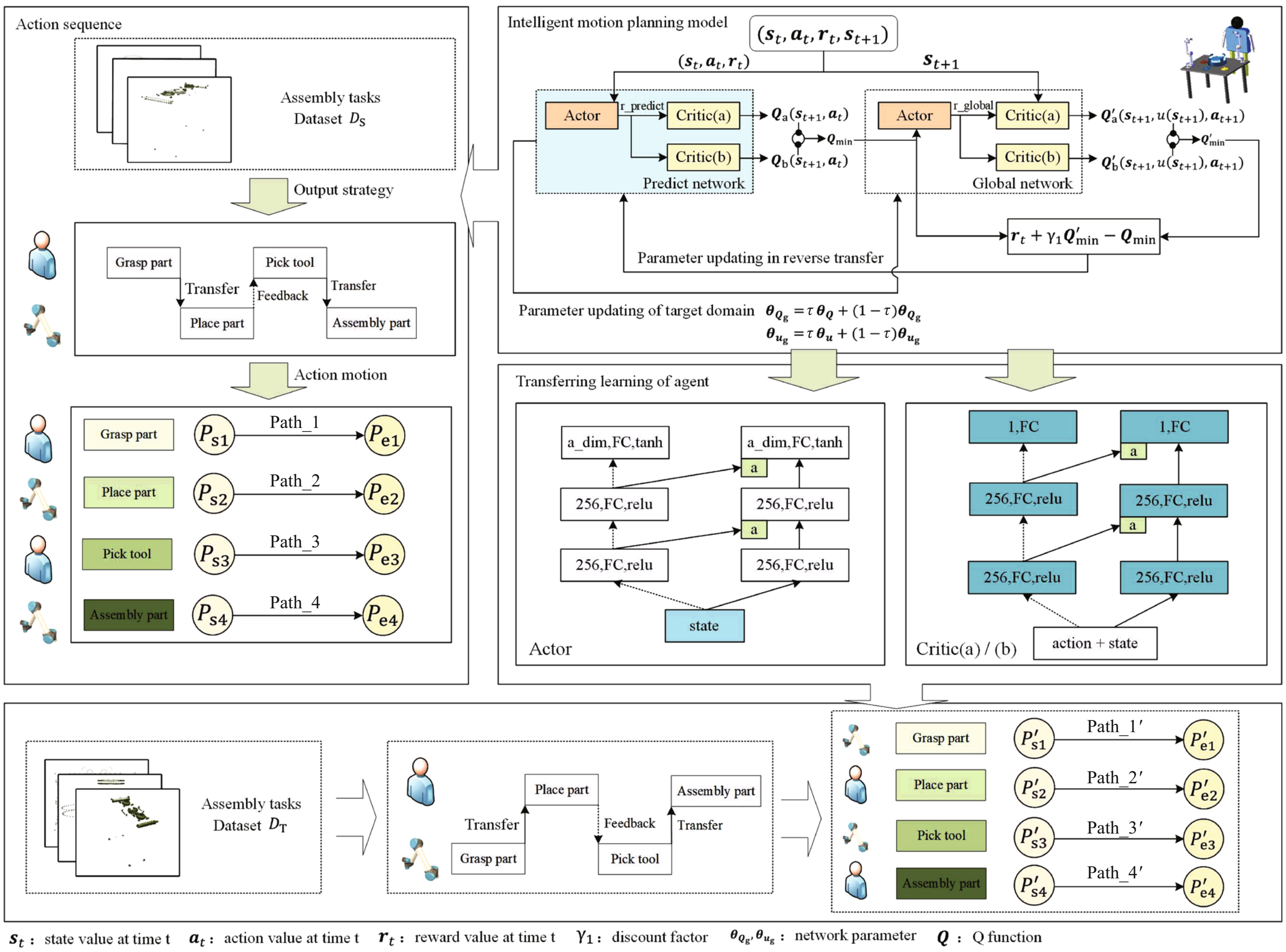


Fig. 7 Adaptive motion planning based on PNN

Because the AC framework has unique advantages in controlling the continuous motion of robot, a dual AC agent is built in order to adapt to planning the safe path. As seen from Fig. 7, the network consists of predict network and global network, in which one actor subnetwork and two parallel critic subnetworks are contained. The actor generates action according to the current state of environment and inputs it to critic to output Q value. In the loop, the time difference (TD) error is used to update the parameters of network. In Fig. 7, "256, FC, relu" respectively indicate the network with number of neurons, type of network layer and type of activation function. The dotted line stands for that the weight remains unchanged during transmission.

In training stage, the state of environment is related to the following three elements: the distance between robot and target object (D), the distance between end effector and obstacle (D_o), and safety of robot motion (S_r). Therefore, the state of environment can be set:

$$\boldsymbol{s}_t = (D, D_\mathrm{o}, S_\mathrm{r}), \tag{10}$$

$$D = \sqrt{(x_\mathrm{r} - x_\mathrm{o})^2 + (y_\mathrm{r} - y_\mathrm{o})^2 + (z_\mathrm{r} - z_\mathrm{o})^2}, \tag{11}$$

$$\left.\begin{array}{l} D_\mathrm{o} = \mathrm{list}(\sqrt{(x_\mathrm{r} - x_i)^2 + (y_\mathrm{r} - y_i)^2 + (z_\mathrm{r} - z_i)^2}) \\ i \in [0, I], \quad I = \mathrm{num(obstacle)} \end{array}\right\}, \tag{12}$$

$$S_\mathrm{r} = \begin{cases} 1, & \text{if } 0 \in D_\mathrm{o} \\ -1, & \text{else} \end{cases}, \tag{13}$$

where, $(x_\mathrm{r}, y_\mathrm{r}, z_\mathrm{r})$ stands for the position of end effector; $(x_\mathrm{o}, y_\mathrm{o}, z_\mathrm{o})$ is the position of target object; (x_i, y_i, z_i) represents the position of obstacles.

The action of RL is the angle value (θ_i) of six joints:

$$\boldsymbol{a}_t = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_6). \tag{14}$$

Then the reward function is set to drive the agent to learn the skill of planning safe path:

$$r_t = 300 - D + S_\mathrm{r} \sum_{i=1}^{n} 8D_{\mathrm{o}_i}. \tag{15}$$

In Fig. 7, the transferring learning of agent can be realized by establishing the lateral connection of network between source domain and target domain. In that, both actor and critic networks have built the lateral connection for improving the learning effect. For multiple assembly tasks, the transfer of empirical data can be achieved by connecting the former networks and target network. The neuron output of layer i can be described as $\boldsymbol{h}_i^k$ and it can be expressed with lateral connection:

$$\boldsymbol{h}_i^k = f(\boldsymbol{W}_i^k \boldsymbol{h}_{i-1}^k + \sum_{j<k} \boldsymbol{M}_i^{j:k} \boldsymbol{h}_{i-1}^j), \tag{16}$$

where, $\boldsymbol{W}_i^k$ stands for the weight matrix of neurons at layer i of network k; $\boldsymbol{M}_i^{j:k}$ is the lateral connection from layer $i-1$ of network j to layer i of network k; $f(\cdot)$ is the activation function corresponding to layer i.

Based on the linear transfer of parameters, the construction of non-linear lateral connection is conductive to expand the transferring range. In that, the output composition vector of all layer $i-1$ at network $k-1$ can be assumed as $\boldsymbol{h}_{i-1}^{:k} = (\boldsymbol{h}_{i-1}^1, \boldsymbol{h}_{i-1}^2, \cdots, \boldsymbol{h}_{i-1}^{k-1})$ with dimension of $\boldsymbol{n}_{i-1}^{:k}$. Therefore, a single-layer neural network can be used to replace the linear lateral connection. The output of layer i at network k is expressed:

$$\boldsymbol{h}_i^k = f(\boldsymbol{W}_i^k \boldsymbol{h}_{i-1}^k + \boldsymbol{M}_i^k \sigma(\boldsymbol{V}_i^k \boldsymbol{\alpha}_{i-1}^{:k} \boldsymbol{h}_{i-1}^{:k})), \tag{17}$$

where, $\boldsymbol{V}_i^k$ is the projection matrix, making the dimension of $\boldsymbol{h}_{i-1}^{:k}$ compressed from $\boldsymbol{n}_{i-1}^{:k}$ to $\boldsymbol{n}_{i-1}^{k}$; $\boldsymbol{\alpha}_{i-1}^{:k}$ is the adjustment coefficient, which is used to adjust the input into the network with single hidden layer; $\sigma(\cdot)$ is the activation function.

Under the flexibility, the motion of human can be fine-tuned by getting the motion with RL agent. After long-term training, the robot can continuously grow the capability of adaptive path planning in different assembly scenes.

3 Experiment

3.1 Experimental Setup

For verifying the effectiveness of the proposed framework, the assembly experiment of power lithium battery is carried out. The main components of power lithium battery include bottom shell, sealing ring, battery pack, main cable, top cover, fuse, protective film and so on. The battery pack part is composed of a battery module, a heat dissipation guard plate, a fireproof guard plate and a metal partition plate. In this experiment, two types of lithium batteries are used, containing S471 standard C box and S472 standard G box. As shown in Fig. 8, the structure and assembly sequence of these two lithium batteries are presented, which shows that the composition of them is similar. Multi-variety lithium batteries with different appearance dimensions and voltage are in demand for coping with the differences in on-board structure and power supply voltage

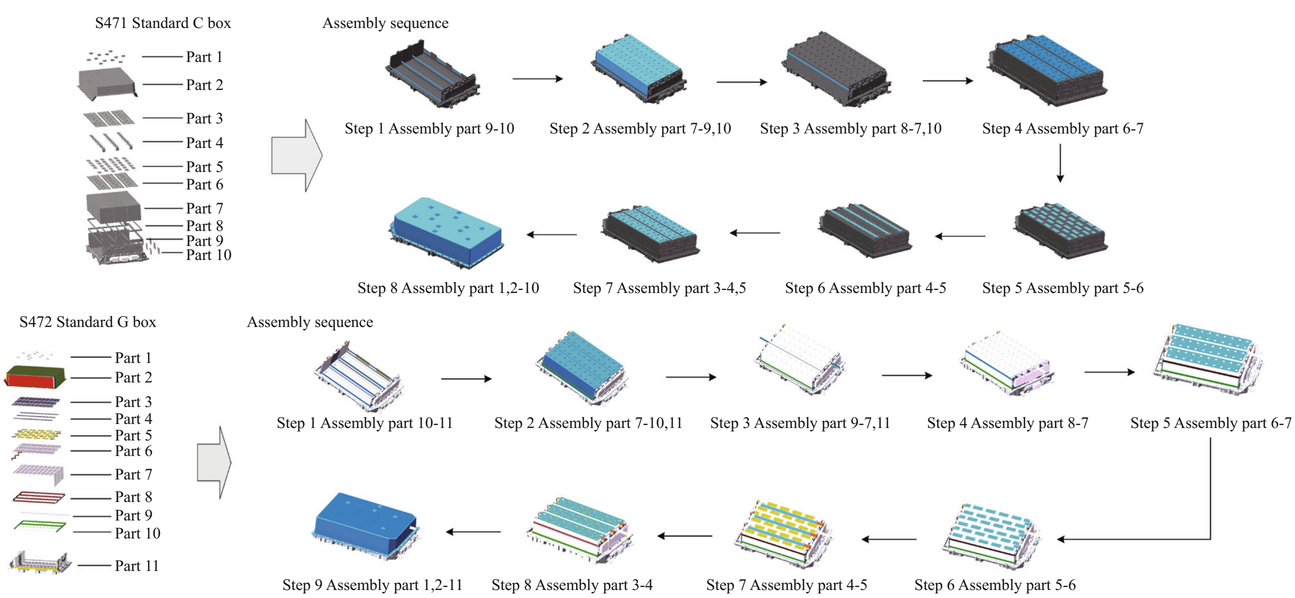


Fig. 8 Assembly sequence of different power lithium batteries

of different new-energy vehicles. S471 standard C box and S472 standard G box are two different categories under the same general category. Therefore, the assembly experiment with S471 standard C box and S472 standard G box can be regarded as the multi-variety assembly. For improving the efficiency of HRCA, the assembly strategy towards HRC is an effective way which can deal with flexible and rigid assembly tasks at the same time. The HRC strategy can be represented as human and robot complete the specific assembly task through the certain execution sequence.

The assembly task towards S471 standard C box is set as the source domain task while the assembly task towards S472 standard G box is set as the target domain task. In experiment, the assembly task in source domain is represented as assembling 100 S471 standard C boxes while the assembly task in target domain is assembling 100 S472 standard G boxes. The list of software and hardware used is shown as follows: GTX1080Ti graphic card with 8 GB memory, Pytorch framework, MySQL database and others.

The comparative experiment is conducted with two aspects. ① In cross-domain recognition of assembly feature, the domain adaption methods based on common transfer learning with domain adaptive neural network (DANN) and deep adaptation network (DAN), and RL with DQN are compared; ② In adaptive path planning, the performance of intelligent agent with or without PNN module is compared.

The datasets composed of the assembly parts from S471 standard C box and S472 standard G box are separately used for feature extraction in source domain and target domain. In the dataset, the number of samples corresponding to each part is set as 1 000 through data enhancement. The number of points in single sample is maintained at 10 000 by down sampling. Therefore, the dataset in source domain contains 10 types of part while the dataset in target domain contains 11 types of part.

The settings of hyper-parameters towards model training during assembly feature recognition and adaptive path planning are shown in Table 3. In cross-domain recognition of assembly feature, the round of model training is all set to 250 episodes or epochs with a learning rate of 10^{-3}. In adaptive path planning, the AC framework is adopted to train the agent with PNN. In agent training of target domain, the episode is set to 280 with PNN while it is 500 without PNN.

Table 3 Settings of hyper-parameters in model training

Stage	Model	Episode/epoch	Learning rate	Optimizer
Assembly feature recognition	DQN	250	10^{-3}	Adam
	DANN	250	10^{-3}	Momentum (0.9)
	DAN	250	10^{-3}	Momentum (0.9)
Adaptive path planning	With PNN	280	10^{-3}	Momentum (10^{-4})
	Without PNN	500	10^{-3}	Adam

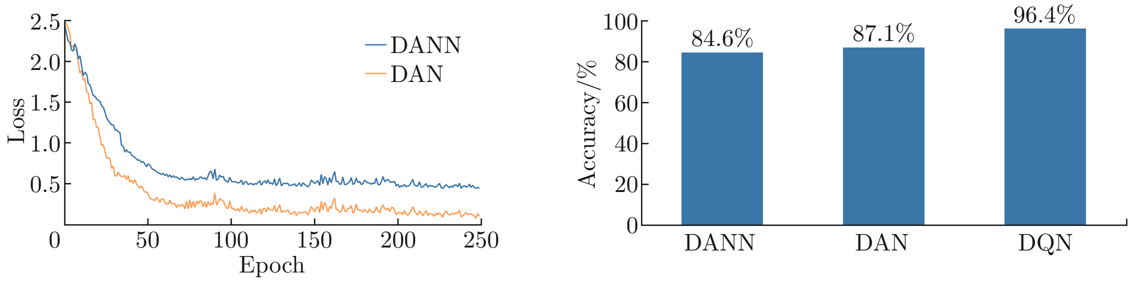
3.2 Experimental Result

3.2.1 *Cross-Domain Recognition of Assembly Feature*

Figure 9 represents the performance among different domain adaption models, including RL model, DANN and DAN. After training, the source encoder and decoder has realized the recognition and segmentation of assembly feature towards S471 standard C box. Then the weight of target decoder is dynamically adjusted by RL model or domain adaptive network. As shown in Fig. 9, when using DAN model, the classification loss converges faster and the final loss is smaller than the case using DANN model. From the perspective of classification accuracy, the classification accuracy is 0.871 with DAN model, which is higher than 0.846 with DANN model. Compared with that of former models, the classification accuracy is highest (0.964) with DQN model. The reason for this phenomenon is that DAN model adopts multi-layer adaption and multi-kernel MMD methods, which improves the accuracy of model. Besides, RL method continuously optimizes the encoder parameters in target domain, which can make the performance achieves the best.

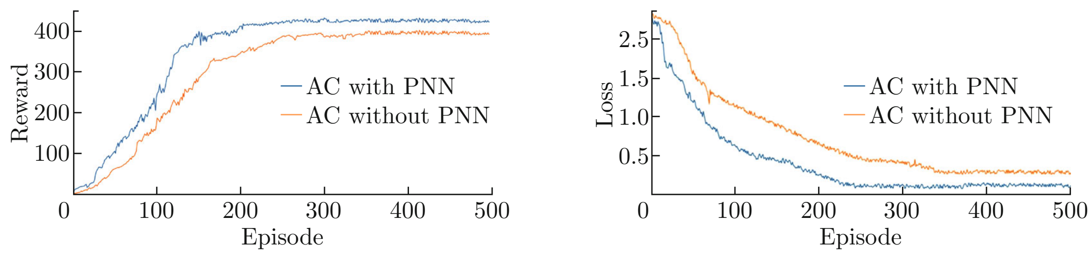
3.2.2 *Adaptive Path Planning*

In experiment, four different subtasks are extracted to test the performance of different adaptive path planning methods, including manual method, AC method and AC-PNN method. As shown in Fig. 10, grasping, transferring, placing and assembling single battery cell are four subtasks used to test the performance of different methods. Each subtask in target domain is all tested 500 times and the average value is taken as the result. In model training as Fig. 10(a), the reward reaches its maximum value in 250 episodes with AC-PNN model while it is 350 episodes with AC model. Besides, the maximum reward value of the former is higher than that of the latter. From Fig. 10(b), the loss curve of AC-PNN model converges earlier than AC model and the loss value of the former is smaller than

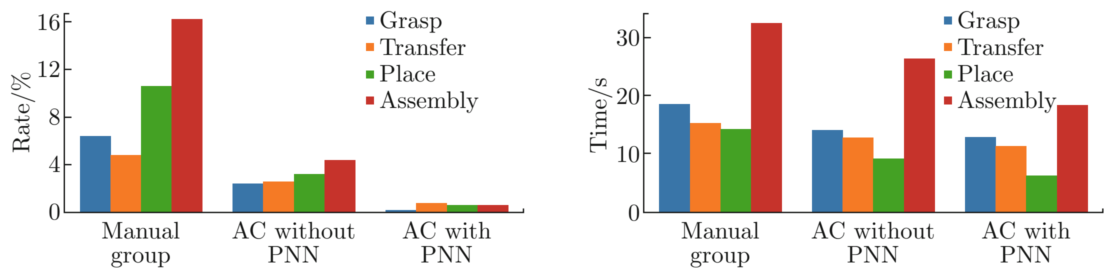


(a) Loss curve in domain adaption (b) Classification accuracy in target domain

Fig. 9 Performance comparison among different models



(a) Reward in RL training (b) Loss in RL training



(c) Average collision rate in four subtasks (d) Average running time in four subtasks

Fig. 10 Performance comparison among different methods

that of the latter. The reason for this phenomenon is that PNN transfers the empirical parameters learned from similar tasks to the target model, which speeds up the training process. Figures 10(c) and 10(d) respectively present the average collision rate and running time of path planning under three different methods. In four subtasks, the path generated by manual design will make the robot consume longer running time and produce higher collision rate. Using AC model to plan the motion path can shorten the running time of robot and maintain the collision rate within 5%. In combination with PNN, the collision rate and running time of path will be limited into a smaller range than common AC model. This phenomenon is caused by the reason that PNN can enable the target model to learn new skills based on the historical skills.

4 Conclusion

The proposed HRCA system is researched from the following aspects: expression of collaboration strategy, transferring of HRCA strategy and adaptive motion planning of HRCA. The relevant conclusions are shown as follows.

(1) The proposed HRCA framework realizes the recognition and classification of cross-domain assembly features. On this basis, the historical assembly experience is adopted to guide the design of human-robot action sequence and planning of path towards new tasks.

(2) Compared with traditional domain adaption method, the RL method can more effectively improve the classification accuracy of assembly features in target domain which reaches 96.4%.

(3) Compared with manual approach, the AC model can reduce the collision rate and running time of motion path in target domain. Besides, the PNN framework will connect target tasks with historical tasks, which greatly shortens the training cycle of agent.

References

[1] LIU Z H, WANG X R, CAI Y J, et al. Dynamic risk assessment and active response strategy for industrial human-robot collaboration [J]. *Computers & Industrial Engineering*, 2020, **141**: 106302.

[2] HUANG J, PHAM D T, LI R Y, et al. An experimental human-robot collaborative disassembly cell [J]. *Computers & Industrial Engineering*, 2021, **155**: 107189.

[3] RENU R S, MOCKO G. Computing similarity of text-based assembly processes for knowledge retrieval and reuse [J]. *Journal of Manufacturing Systems*, 2016, **39**: 101-110.

[4] DARVISH K, WANDERLINGH F, BRUNO B, et al. Flexible human-robot cooperation models for assisted shop-floor tasks [J]. *Mechatronics*, 2018, **51**: 97-114.

[5] ABUDUWEILI A, LI S Y, LIU C L. Adaptable human intention and trajectory prediction for human-robot collaboration [DB/OL]. (2019-09-11). https://arxiv.org/abs/1909.05089.

[6] CHENG Y J, SUN L T, LIU C L, et al. Towards better human robot collaboration with robust plan recognition and trajectory prediction [DB/OL].(2020-02-17). https://arxiv.org/abs/1903.02199.

[7] LIU H Y, WANG L H. Collision-free human-robot collaboration based on context awareness [J]. *Robotics and Computer-Integrated Manufacturing*, 2021, **67**: 101997.

[8] AMORIM A, GUIMARES D, MENDONA T, et al. Robust human position estimation in cooperative robotic cells [J]. *Robotics and Computer-Integrated Manufacturing*, 2021, **67**: 102035.

[9] MAKRIS S, KARAGIANNIS P, KOUKAS S, et al. Augmented reality system for operator support in human-robot collaborative assembly [J]. *CIRP Annals*, 2016, **65**(1): 61-64.

[10] RAATZ A, BLANKEMEYER S, RECKER T, et al. Task scheduling method for HRC workplaces based on capabilities and execution time assumptions for robots [J]. *CIRP Annals*, 2020, **69**(1): 13-16.

[11] LIU H Y, WANG L H. Human motion prediction for human-robot collaboration [J]. *Journal of Manufacturing Systems*, 2017, **44**: 287-294.

[12] BERG J, RECKORDT T, RICHTER C, et al. Action recognition in assembly for human-robot-cooperation using hidden Markov models [J]. *Procedia CIRP*, 2018, **76**: 205-210.

[13] LI Z, JIN X L, JIA H Y, et al. The knowledge representation and reuse in product design [J]. *Journal of Shanghai Jiao Tong University*, 2006, **40**(7): 1183-1186 (in Chinese).

[14] HE L L, GUO W, JIANG P Y. A decision-making model for knowledge collaboration and reuse through scientific workflow [J]. *Advanced Engineering Informatics*, 2021, **49**: 101345.

[15] ARANA-AREXOLALEIBA N, URRESTILLA-ANGUIOZAR N, CHRYSOSTOMOU D, et al. Transferring human manipulation knowledge to industrial robots using reinforcement learning [J]. *Procedia Manufacturing*, 2019, **38**: 1508-1515.

[16] RAZIEI Z, MOGHADDAM M. Adaptable automation with modular deep reinforcement learning and policy transfer [J]. *Engineering Applications of Artificial Intelligence*, 2021, **103**: 104296.

[17] RODRÍGUEZ I, NOTTENSTEINER K, LEIDNER D, et al. Pattern recognition for knowledge transfer in robotic assembly sequence planning [J]. *IEEE Robotics and Automation Letters*, 2020, **5**(2): 3666-3673.

[18] GUO M H, CAI J X, LIU Z N, et al. PCT: Point cloud transformer [J]. *Computational Visual Media*, 2021, **7**(2): 187-199.