Robotica (2024), 1–16
doi:10.1017/S0263574724001164

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Robot skill acquisition for precision assembly of flexible flat cable with force control

Xiaogang Song[1,2,3], Peng Xu[1,2,3], Wenfu Xu[1,2,3], Bing Li[1,2,3] and Lei Qin[4]

[1]State Key Laboratory of Robotics and System, Harbin Institute of Technology, Harbin, China
[2]Guangdong Key Laboratory of Intelligent Morphing Mechanisms and Adaptive Robotics, Harbin Institute of Technology, Shenzhen, China
[3]School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen, China
[4]Guangdong HUIBO Robot Technology Co., Ltd., Foshan, China
Corresponding author: Peng Xu; Email: xupeng919@hit.edu.cn

Received: 27 February 2024; **Revised:** 29 May 2024; **Accepted:** 6 June 2024

Keywords: Efficient robot skill acquisition; flexible flat cable (FFC) assembly; force control; reinforcement learning

Abstract

The flexible flat cable (FFC) assembly task is a prime challenge in electronic manufacturing. Its characteristics of being prone to deformation under external force, tiny assembly tolerance, and fragility impede the application of robotic assembly in this field. To achieve reliable and stable robotic automation assembly of FFC, an efficient assembly skill acquisition strategy is presented by combining a parallel robot skill learning algorithm with adaptive impedance control. The parallel robot skill learning algorithm is proposed to enhance the efficiency of FFC assembly skill acquisition, which reduces the risk of damaging FFC and tackles the uncertain influence resulting from deformation during the assembly process. Moreover, FFC assembly is also a complex contact-rich manipulation task. An adaptive impedance controller is designed to implement force tracking during the assembly process without precise environment information, and the stability is also analyzed based on the Lyapunov function. Experiments of FFC assembly are conducted to illustrate the efficiency of the proposed method. The experimental results demonstrate that the proposed method is robust and efficient.

1. Introduction

The demand for the capacity of electronic product manufacturing extends with the increase of consuming electronic products, such as mobile phones and laptops. The connector assembly is one of the most important and challenging stages in electronic manufacturing. Flexible flat cable (FFC) assembly is fairly commonly found in modern electronics for connecting two electronic components [1]. The difficulties of assembling FFC are the tiny assembly tolerance and uncertain disturbance resulting from the characteristic of being prone to deformation under external force. Currently, this task is still carried out manually. Robot assembly plays a significant role in automated production. Nevertheless, the adaptability of conventional robot assembly is poor in an uncertain environment since it is based on position control. The robot conducts the assembly task by tracking the desired trajectory. The tiny position deviation may result in unsuccessful assembly. Relatively few works have been conducted on FFC assembly. However, lots of research has been discussed to tackle similar assembly tasks.

The frequently used method is to model the contact state and analyze the geometric errors. Yao and Cheng [2] derived the geometrical compatibility condition to address the issue of force overshoot in the no-cylindrical part assembly tasks. Park et al. [3] analyzed the contact state between the peg and the hole. Then, unit motions based on the analysis are presented to perform the peg-in-hole assembly task. A novel modeing method based on a Non-Uniform Rational B-Splines surface was proposed by Zhang

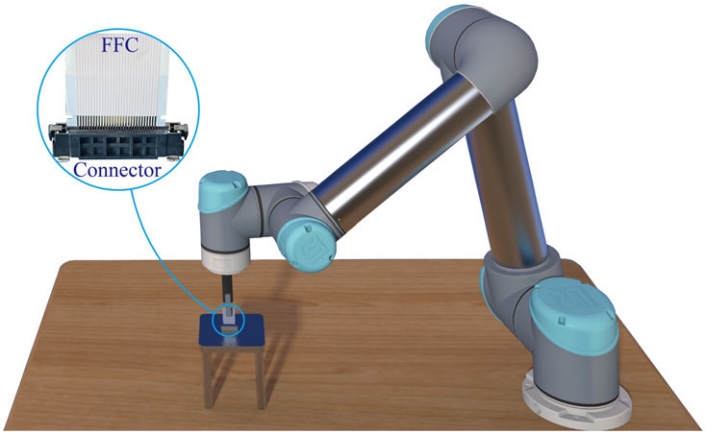

Figure 1. *FFC assembly through the robot.*

et al. [4] for precise assembly tasks. Tang et al. [5] built up a three-point contact model and estimated the pose misalignment between the peg and hole through force and geometric analysis.

However, it is difficult to analyze the contact state in a complex environment. A direct assembly strategy based on demonstration is commonly used. Duque et al. [6] generated assembly plans based on demonstration data captured by the Kinect motion sensor for assembling construction toys. To shorten the setup times of automated assembly tasks, Kramberger et al. [7] proposed a method for learning the constraints of desired tasks via demonstration and autonomous exploration. Su et al. [8] introduced a strategy to teach robots assembly skills by combining adaptive impedance control (AIC) with dynamic motion primitives. To solve the problem of peg-in-hole tasks, Abu-Dakka et al. [9] proposed an algorithm based on programming by demonstration. Roveda et al. [10] proposed an assembly method based on a Bayesian optimization algorithm, which can efficiently enable robots to perform assembly tasks through a few human's demonstrations while compensating for the task uncertainties.

As compared to rigid components, the assembly of FFC (Figure 1) entails more uncertain interference and presents greater challenges. FFC is a type of flexible electronics, so it is highly prone to deformation under external force. This causes the FFC to move out of sync with the robot, which is equivalent to the environment constantly changing. The assembly strategy according to the contact model is stable and efficient in the structural environment. Nevertheless, FFC assembly is a complex contact-rich manipulation task, and it is difficult to establish the contact model. The assembly method combining human's demonstrations data and iterative optimization algorithms demonstrates high learning efficiency and data utilization rates. However, the high-precision FFC assembly is fraught with uncertainties, primarily stemming from deformation and friction between the FFC and connector. It is difficult to get rich and enough demonstration data and take all uncertainties into account [11, 12]. Additionally, FFC is very fragile and easily damaged. Therefore, an advanced algorithm with remarkable adaptability is desirable to complete this task.

With the ultrafast development of machine learning technology, reinforcement learning (RL) is a direct optimization approach that enables robots to learn skills through trial-and-error without requiring prior knowledge. It shows impressive skill learning performance in the field of robotics. Roveda et al. [13, 14] combined RL and variable impedance control methods to recognize human intentions and ensure the robot moves safely in the intended motion direction, which significantly improves the performance of physical human–robot interaction. RL has also brought about new opportunities for robot assembly skill acquisition through trial-and-error. Ma et al. [12] presented an efficient robot assembly skill learning framework for precise assembly tasks by combining offline pretraining based on a few demonstrations from experts and online self-learning using RL. Inoue et al. [15] proposed a strategy to perform the high-precision peg-in-hole fitting task by training a recurrent neural network with RL.

Xu et al. [16] introduced a model-driven deep RL algorithm to complete multiple peg-in-hole tasks. Luo et al. [17] combined RL and force controller to solve the problem of a gear assembly task. In previous works, there are few studies involving electronic product assembly tasks [18–20], especially the FFC. Data inefficiency restricts the application of RL in real complex systems [21]. Large-scale trial-and-error brings the risk of damage to the FFC. Improving learning efficiency and reducing trial-and-error are the problems to be solved. In addition, the assembly of FFC is a complex contact-rich manipulation task, and small position deviation can result in a large contact force. During the assembly process, a force controller with excellent robustness and adaptability should be taken into account to deal with the uncertain disturbance. Impedance control (IC) is one of the popular compliant control methods proposed by Hogan [22], and many efforts have been made to further improve the force-tracking performance within this control frame. Roveda et al. [23, 24] combined fuzzy learning and iterative learning algorithms to enhance the force-tracking capability of robots in unknown environments, demonstrating strong robustness. The adaptive control algorithm is also applied to minimize the force-tracking error under uncertain disturbances [25–29]. The purpose of this paper is to design a force controller that is easy to implement and highly robust in order to solve the force control issues in the FFC assembly process. To address the two main issues mentioned above, the main contributions are summarized as follows:

1. An efficient parallel assembly skill learning algorithm is proposed based on RL. The simulation system and real robot system share the learning experiences in parallel. The environment information obtained from the real robot system is utilized for the training optimal policy, providing efficient guidance to perform FFC assembly tasks in a real robot system. The optimal policy trained by the simulation system is also refined by updating real physical information simultaneously.
2. An AIC algorithm is presented to track the desired force in the unstructured environment during the assembly process, and the stability is also analyzed based on the Lyapunov function.
3. The proposed parallel learning algorithm and AIC are combined to enable the robot to acquire the skill of FFC assembly efficiently.

The rest of the paper is organized as follows. In Section 2, an efficient parallel assembly skill learning algorithm is proposed. Section 3 presents the AIC algorithm. In Section 4, FFC assembly experiments are conducted to demonstrate the performance of the proposed assembly skill acquisition strategy. Finally, a conclusion is drawn in Section 5.

2. Efficient parallel assembly skill learning algorithm

The thickness of FFC is only 0.3 mm, which makes it highly susceptible to deformation under external force. Additionally, FFC is very fragile. These characteristics require assembly skill learning algorithms to have efficient learning capabilities to cope with the uncertainty caused by deformation and reduce the risk of damaging FFC. Sim-to-real is a fast approach for training robots to acquire skills [30, 31]. Inspired by this, we present an efficient parallel assembly skill learning algorithm to accelerate the training process for high-precision FFC assembly tasks based on RL. The main idea of the algorithm is to share information, which includes learning experience, goal state, and reward information, between the simulation system and real robot system in parallel. The simulation system efficiently trains optimal policy based on the successful and failed assembly pose data collected from the real robot system. The agent receives rewards or penalties when the robot reaches the corresponding poses. The optimal policy trained in the simulation system is used to guide the real robot system in completing the assembly task. Meanwhile, the real robot system rectifies the optimal policy trained by the simulation system. The simulation system in this paper is built using the V-rep. However, it does not take the interaction model into consideration. Due to the complex contact involved in high-precision FFC assembly, the impact of friction on the FFC assembly is significant. In addition, FFC is made of flexible material which

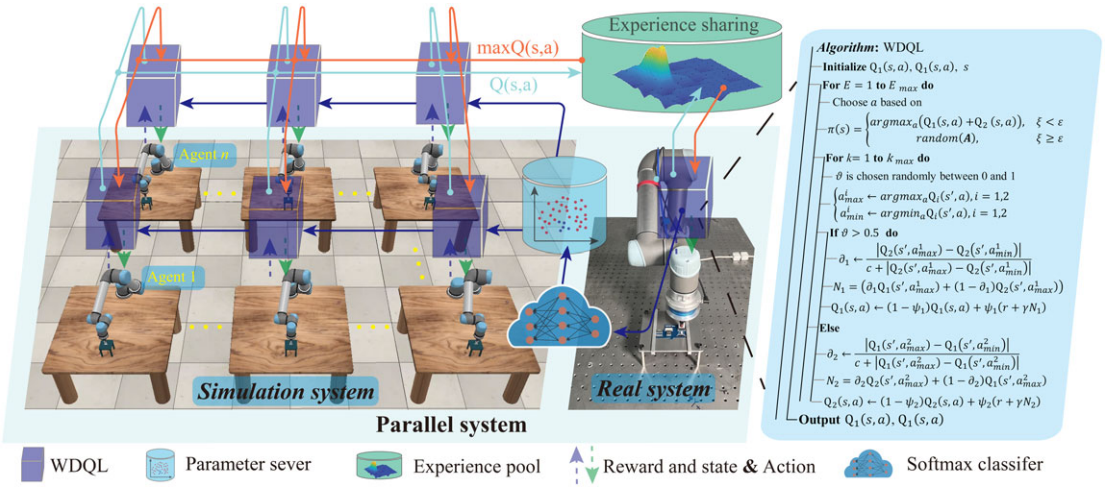

Figure 2. *The framework of the parallel assembly skill learning algorithm.*

easily deforms under external force. Therefore, it is extremely challenging to establish a simulation environment consistent with the real robot system. The FFC assembly skill learning algorithm is shown in Figure 2 and Algorithm 1. The parameter server is used to store the evaluation model trained by the Softmax classifier using physical environment information from the real robot system. The parameter server is also applied for the simulation system to train optimal policy, and it will be rectified consecutively by the real robot system. The experience pool consisting of two shared Q-tables is applied to save the learning experience of all agents so that the agent learns the optimal experiences from other agents to accelerate the training process. All agents have individual Q-function and consecutively update the shared Q-table. The experience means the records from state space to action space.

The weighted double Q-learning algorithm (WDQL) is utilized for each agent to acquire assembly skills, which dilutes the influence of the overestimation in Q-learning (QL) [32]. The assembly process can be described as the Markov decision process, which consists of agents, environment, action space $\mathcal{A}$, state space $\mathcal{S}$, and reward strategies. At each step t, the agent chooses action $a \in \mathcal{A}$, interacting with the environment, and then gets the reward r_t. The state space S is defined as:

$$\mathcal{S} = \begin{bmatrix} f_x & f_y & f_z & \tau_z & x & y & z & \theta_z \end{bmatrix} \tag{1}$$

where f_x, f_y, f_z, and τ_z denote the contact force and moment from the force sensor along x-, y-, and z-axes. x, y, z, and θ_z are the position and angle of the end-effector of the manipulator along x-, y-, and z-axes.

According to the actual assembly process of the FFC, the action space consisting of six components (Figure 3) can be defined as:

$$\mathcal{A} = \begin{bmatrix} \Delta x & -\Delta x & \Delta y & -\Delta y & \Delta\theta & -\Delta\theta \end{bmatrix} \tag{2}$$

where Δx, Δy, and $\Delta\theta$ denote the step lengths for translation and rotation adjustment along x-, y-, and z-axes.

The assembly skill acquisition process starts with a random exploration. The performance of exploration is improved by maximizing the cumulative reward:

$$\Re_k = r_k + \gamma r_{k+1} + \gamma^2 r_{k+2} \cdots + \gamma^{n-k} r_n = r_k + \gamma \Re_{k+1} \tag{3}$$

where γ is the discount rate, r is the current reword, and k stands for the step number.

In the skill learning process, the purpose is to complete the task as quickly as possible. The reward strategy is applied to evaluate the performance of the action a at state $s \in S$ in the real robot and simulation system. The parameter server is utilized to evaluate whether the assembly task is completed

Algorithm 1 Efficient parallel assembly skill learning algorithm

Initialize: the parameters of parallel assembly skill learning: E_{max}, k_{max}, $Q_1(s,a)$, $Q_2(s,a)$, ψ_1, ψ_2, Δx, Δy, $\Delta\theta$, the successful assembly number of real robot system $N_{rel}=0$, the number of agents in the simulation system n, and the safe boundary of the contact force f_b.
Initialize: the parameters of the AIC (see Section 3): f_d, m, b, $\varphi(0)$, $\lambda_p(0)$, $\lambda_d(0)$, μ_0, ϱ_0, μ_1, ϱ_1, μ_2, and ϱ_2.
Start the thread of the real robot system.
Pick up FFC1 and regulate contact force using the proposed AIC.
Begin to learn and start the thread of Softemax classifier.
for episode $=1$ **to** E_{max} **do**
  if $N_{rel}=1$ **then**
    Start the thread of the simulation system.
  end if
 # Thread of the simulation system
  while thread of the real robot system is alive **then**
    for $k=1$ **to** k_{max} **do**
      Choose action a based on $\pi(s)$.
      Take action a and get S_{t+1}.
      Get r based on parameter server and $k \geq k_{max}$.
      Update the experience pool based on Eq. (6).
    end for
  end while
 # Thread of the real system
  for $k=1$ **to** k_{max} **do**
    Choose action a based on $\pi(s)$.
    Take action a and get S_{t+1}.
    Get r based on reward strategy Eq. (4)
    Update the experience pool based on Eq. (6).
    Update the parameter server based on Softmax classifier
    if successful assembly **then**
      $N_{rel}=N_{rel}+1$
    end if
  end for
end for

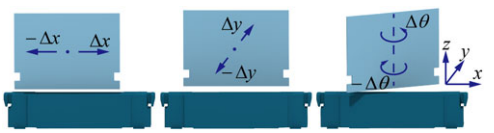

Figure 3. *Actions for FFC assembly.*

successfully. According to the strategy [16] and production specifications [38], the reward strategy is designed as follows:

$$r=\begin{cases} -0.5 & \text{if } k \geq k_{max} \\ 1-\frac{k}{k_{max}}, & \text{if successful assembly} \\ -1, & \text{if danger occurs} \end{cases} \tag{4}$$

where k_{max} denotes the maximum step in an episode. If the robot fails to assemble after attempting k ($\geq k_{max}$ steps, it receives a reward $r = -0.5$. The danger occurs means that the absolute value of the contact force exceeds the safe boundary $f_b = [f_x^b, f_y^b, f_z^b, \tau_z^b]$, it receives a reward $r = -1$. f_x^b, f_y^b, f_z^b, and τ_z^b denote the safe boundary of contact force and moment along the x-, y-, and z-axes, respectively. When one of the three situations in Eq. (4) occurs, the robot will return to the initial position and start the next episode.

In order to obtain the optimal assembly strategy, the WDQL is applied to select the best possible action. The action selection policy is defined as follows:

$$\pi(s) = \begin{cases} \text{argmax}_a(Q_1(s,a) + Q_2(s,a)), & \xi < \varepsilon \\ \text{random}(\mathcal{A}), & \xi \geq \varepsilon \end{cases} \tag{5}$$

where ξ is generated between 0 and 1 randomly and $\varepsilon \in (0, 1)$ is the greedy parameter. $Q_1(s,a)$ and $Q_2(s,a)$ are the Q-functions which recursively update by the Bellman function randomly

$$\begin{cases} Q_1(s,a) \leftarrow (1-\psi_1)\, Q_1(s,a) + \psi_1 \left(r + \gamma N_1\right), \text{if } \vartheta > 0.5 \\ \quad N_1 = \left(\partial_1 Q_1\left(s', a_{max}^1\right) + (1-\partial_1)\, Q_2\left(s', a_{max}^1\right)\right) \\ Q_2(s,a) \leftarrow (1-\psi_2)\, Q_2(s,a) + \psi_2 \left(r + \gamma N_2\right), \text{if } \vartheta < 0.5 \\ \quad N_2 = \partial_2 Q_2\left(s', a_{max}^2\right) + (1-\partial_2)\, Q_1\left(s', a_{max}^2\right) \end{cases} \tag{6}$$

where ψ_1 and ψ_2 are the learning rate and ϑ is generated between 0 and 1 randomly. In the WDQL, the Q-value is recorded in the double Q-tables. ∂_1 and ∂_2 are the weighted parameters given by:

$$\begin{cases} \partial_1 \leftarrow \dfrac{\left|Q_2\left(s', a_{max}^1\right) - Q_2\left(s', a_{min}^1\right)\right|}{c + \left|Q_2\left(s', a_{max}^1\right) - Q_2\left(s', a_{min}^1\right)\right|} \\ \partial_2 \leftarrow \dfrac{\left|Q_1\left(s', a_{max}^2\right) - Q_1\left(s', a_{min}^2\right)\right|}{c + \left|Q_1\left(s', a_{max}^2\right) - Q_1\left(s', a_{min}^2\right)\right|} \\ a_{max}^i \leftarrow \text{argmax}_a Q_i(s', a), i = 1, 2 \\ a_{min}^i \leftarrow \text{argmin}_a Q_i(s', a), i = 1, 2 \end{cases} \tag{7}$$

where c is a constant.

To preliminarily validate the feasibility of the proposed method, a simulation of rectangular peg-in-hole assembly is implemented using Python. The successful assembly position and orientation are set to [4 4 5], and a search is performed within the given range to simulate the process of peg-in-hole assembly. The search range is set to $x \epsilon [-8, 8]$, $y \epsilon [-8, 8]$ and $\theta \epsilon [-6, 6]$. The comparative analysis of the performance of commonly used QL, deep Q-learning (DQL), actor-critic (AC), policy gradient (PG), proximal policy optimization (PPO), and the proposed method is implemented. The parameters of the skill learning algorithm are set to $\psi_1, \psi_2 = 0.01$, $\varepsilon = 0.85$, $\Delta x = 1mm$, $\Delta y = 1mm$, $\Delta\theta = 1deg$, and $\gamma = 0.85$. The learning rate of the neural network in DQL, AC, PG, and PPO is set to 0.01. The simulation results are shown in Figure 4. With the utilization of QL, DQL, AC, PG, and PPO, the learning step has no notable drop trend. When using the proposed method, the learning efficiency improves as the number of agents n increases. However, in real physical experiments, it is impractical to increase the number of agents of the simulation system indefinitely, as this would affect the real-time performance of the real robot system and the force-tracking performance of the force controller. Therefore, a trade-off must be made between force-tracking performance and the number of agents in a real assembly process.

3. Adaptive impedance control

The assembly of the FFC is a complex contact-rich manipulation task. AIC is used to ensure stable contact force between the FFC and connector during the learning assembly process. Meanwhile, the contact force information is also utilized as a crucial element for assembly learning. The combination of the parallel assembly skill learning algorithm and AIC in this paper refers to using the action sequence

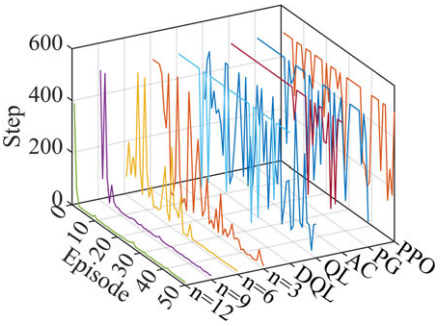

Figure 4. *Learning efficiency of the different methods in simulation.*

output by the learning strategy as the desired trajectory input to the AIC. The AIC is responsible for maintaining the desired contact force along the input trajectory. Throughout the process, the parallel assembly skill learning algorithm does not adjust the parameters of AIC. In this section, the proposed AIC is derived, and the stability is analyzed based on the Lyapunov function. The general form of the Cartesian IC model composed of both the translational and rotational parts can be expressed as follows:

$$\boldsymbol{M}\left(\ddot{\boldsymbol{X}}_c-\ddot{\boldsymbol{X}}_d\right)+\boldsymbol{B}\left(\dot{\boldsymbol{X}}_c-\dot{\boldsymbol{X}}_d\right)+\boldsymbol{K}\left(\boldsymbol{X}_c-\boldsymbol{X}_d\right)=\boldsymbol{F}_m-\boldsymbol{F}_d \tag{8}$$

where $\boldsymbol{M}\epsilon R^{6\times6}$ is the mass matrix, $\boldsymbol{B}\epsilon R^{6\times6}$ is the damping matrix, and $\boldsymbol{K}\epsilon R^{6\times6}$ is the stiffness matrix. $\boldsymbol{F}_d\epsilon R^6$ denotes the prescribed force, and $\boldsymbol{F}_m\epsilon R^6$ is the measured force between the end-effector and the environment. $\boldsymbol{X}_c\epsilon R^6$ is the command trajectory of the end-effector sent to the robot, and $\boldsymbol{X}_d\epsilon R^6$ is the desired trajectory of the end-effector. $\boldsymbol{M}$, $\boldsymbol{B}$, and $\boldsymbol{K}$ are all diagonal matrices [33]. Each direction can be independently controlled [34–36].

Without loss of generality, a one-dimensional case is studied. Eq. (8) can be rewritten as:

$$m\Delta\ddot{x}+b\Delta\dot{x}+k\Delta x=f_m-f_d \tag{9}$$

where $\Delta x=x_c-x_d$. By applying the Laplace transform to Eq. (9), it can be expressed as:

$$\left(ms^2+bs+k\right)\Delta x=f_m-f_d=\Delta f \tag{10}$$

Assuming that the environment can be represented by a linear spring model with stiffness k_e, it has $f_m=k_e(x_e-x_c)$, where x_e is the environment location. Then, we can get

$$\Delta f=\frac{ms^2+bs+k}{ms^2+bs+k+k_e}\left(f_d+k_e\left(x_e-x_d\right)\right) \tag{11}$$

According to Eq. (11), the steady-state force-tracking error can be written as:

$$\Delta f_{ss}=\frac{k}{k+k_e}\left(k_e\left(x_e-x_d\right)-f_d\right) \tag{12}$$

If $\Delta f_{ss}=0$, either $k=0$ or k_e and x_e are known exactly. Typically, the environment is complicated. It is difficult to get information on x_e and k_e; hence, the target IC can be modified as:

$$m\ddot{x}_c+b\left(\dot{x}_c-\dot{x}_d\right)=f_m-f_d \tag{13}$$

Generally, the force-tracking performance of traditional IC is poor in complex environments. To overcome this issue, an AIC algorithm based on the Lyapunov function is presented to further improve the force-tracking capability against uncertain disturbance. It can be expressed as:

$$m\ddot{x}_c\left(t\right)+b\left(\dot{x}_c\left(t\right)-\dot{\hat{x}}_d\left(t\right)\right)=\Delta f\left(t\right) \tag{14}$$

$$\dot{\hat{x}}_d\left(t\right)=\varphi\left(t\right)+\lambda_p\left(t\right)\Delta f\left(t\right)+\lambda_d\left(t\right)\Delta\dot{f}\left(t\right) \tag{15}$$

where $\lambda_p(t)$ and $\lambda_d(t)$ are the adaptive gains, and $\varphi(t)$ is the auxiliary compensate function.

Substituting Eq. (15) into Eq. (14) and with the modification, then Eq. (14) becomes

$$m\Delta\ddot{f} + (b + \lambda_d)\Delta\dot{f} + \left(k_e + \lambda_p\right)\Delta f = -bk_e\varphi \tag{16}$$

Let us define force-tracking error $\boldsymbol{E} = [\ \Delta f \quad \Delta\dot{f}\ ]^T$, then Eq. (16) can be represented by:

$$\dot{\boldsymbol{E}} = \begin{bmatrix} 0 & 1 \\ \dfrac{-k_e + \lambda_d}{m} & \dfrac{-b + \lambda_p}{m} \end{bmatrix} \boldsymbol{E} + \begin{bmatrix} 0 \\ \dfrac{-bk_e\varphi}{m} \end{bmatrix} \tag{17}$$

The prescribed performance of Δf can be described by the reference model:

$$\dot{\boldsymbol{E}}_r = \boldsymbol{R}_r \tag{18}$$

where $R = \begin{bmatrix} 0 & 1 \\ -\omega^2 & -2\eta\omega \end{bmatrix}$, $\boldsymbol{E}_r = [\ e_r\ \dot{e}_r\ ]^T$, ω is the natural frequency, and η denotes damping ratio. the reference model is stable, that is, $\boldsymbol{E}_r \equiv 0$.

Defining $\boldsymbol{E}_m = \boldsymbol{E}_r - \boldsymbol{E}$, and substituting Eq. (17) and Eq. (18) into it, the error differential equation can be obtained:

$$\dot{\boldsymbol{E}}_m = \begin{bmatrix} 0 & 1 \\ -\omega^2 & -2\eta\omega \end{bmatrix} \boldsymbol{E}_m + \begin{bmatrix} 0 \\ \dfrac{bk_e\varphi}{m} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ \dfrac{k_e + \lambda_d}{m} - \omega^2 & \dfrac{b + \lambda_p}{m} - 2\eta\omega \end{bmatrix} \boldsymbol{E} \tag{19}$$

A Lyapunov function is defined as:

$$\begin{aligned} V = {} & \boldsymbol{E}_m^T \boldsymbol{P} \boldsymbol{E}_m + \beta_0 \left(\frac{bk_e\varphi}{m} - \varphi \right)^2 \\ & + \beta_1 \left(\frac{k_e + \lambda_p}{m} - \omega^2 - \lambda_p \right)^2 \\ & + \beta_2 \left(\frac{b + \lambda_d}{m} - 2\eta\omega - \lambda_d \right)^2 \end{aligned} \tag{20}$$

where $\beta_i (i = 0, 1, 2)$ is the positive parameter and φ, λ_d, and λ_p are the function of time. $\boldsymbol{P} = \begin{bmatrix} p_1 & p_2 \\ p_2 & p_3 \end{bmatrix}$ is the symmetric positive-definite constant matrix.

Differentiating Eq. (20), we can get

$$\begin{aligned} \dot{V} = {} & -\boldsymbol{E}_m^T \boldsymbol{Q} \boldsymbol{E}_m + 2\frac{bk_e\varphi}{m} \left(\beta_0 \left(\frac{bk_e\dot{\varphi}}{m} - \dot{\varphi} \right) - \delta \right) \\ & - 2\beta_0\varphi \left(\frac{bk_e\dot{\varphi}}{m} - \dot{\varphi} \right) \\ & + 2 \left(\frac{k_e + \lambda_p}{m} - \omega^2 \right) \left(\beta_1 \left(\frac{\dot{\lambda}_p}{m} - \dot{\lambda}_p \right) - \delta\Delta f \right) \\ & - 2\beta_1\lambda_p \left(\frac{\dot{\lambda}_p}{m} - \dot{\lambda}_p \right) \\ & + 2 \left(\frac{k_e + \lambda_d}{m} - 2\eta\omega \right) \left(\beta_2 \left(\frac{\dot{\lambda}_d}{m} - \dot{\lambda}_d \right) - \delta\Delta\dot{f} \right) \\ & - 2\beta_1\lambda_d \left(\frac{\dot{\lambda}_d}{m} - \dot{\lambda}_d \right) \end{aligned} \tag{21}$$

where $\delta = p_2\Delta f + p_3\Delta\dot{f}$ and $\boldsymbol{Q} = -\boldsymbol{P}\boldsymbol{R} - \boldsymbol{D}^T\boldsymbol{P}$ [37].

For the purpose that $\boldsymbol{E}_m$ tends to zero asymptotically, $\dot{V}$ should be negative-definite. To achieve this, we set

$$\begin{aligned}
\delta &= \beta_0 \left(\frac{bk_e\dot{\varphi}}{m} - \dot{\varphi} \right), \varphi = \beta_0 \delta \\
\delta \Delta f &= \beta_1 \left(\frac{\dot{\lambda}_p}{m} - \dot{\lambda}_p \right), \lambda_p = \beta_1 \delta \Delta f \\
\delta \Delta \dot{f} &= \beta_2 \left(\frac{\dot{\lambda}_d}{m} - \dot{\lambda}_d \right), \lambda_d = \beta_1 \delta \Delta \dot{f}
\end{aligned} \tag{22}$$

Substituting Eq. (22) to Eq. (21) and simplifying, Eq. (21) can be rewritten as:

$$\dot{V} = -\boldsymbol{E}_m^T \boldsymbol{Q} \boldsymbol{E}_m - 2\beta_0 \delta^2 - 2\beta_1 \delta^2 \Delta f^2 - 2\beta_2 \delta^2 \Delta \dot{f}^2 \tag{23}$$

As $\dot{V}$ is negative-definite, it implies $\boldsymbol{E}_m \to 0$ and Eq. (17) is asymptotically stable. From Eq. (22), the adaptive control law can be obtained as:

$$\begin{aligned}
\dot{\varphi} &= \frac{m}{bk_e} \left(\frac{\delta}{\beta_0} + \beta_0 \dot{\delta} \right) \\
\dot{\lambda}_p &= \frac{m}{\beta_1} \delta \Delta f + m\beta_1 \frac{d}{dt} \left(\delta \Delta f \right) \\
\dot{\lambda}_d &= \frac{m}{\beta_2} \delta \Delta \dot{f} + m\beta_2 \frac{d}{dt} \left(\delta \Delta \dot{f} \right)
\end{aligned} \tag{24}$$

In order to make the control laws independent of m, b, and k_e, we set

$$\begin{aligned}
\beta_0 &= \frac{m}{bk_e\mu_0}, \qquad \beta_0 = \frac{bk_e\varrho_0}{m} \\
\beta_1 &= \frac{\mu_1}{m}, \qquad \beta_1 = \frac{\varrho_1}{m} \\
\beta_2 &= \frac{\mu_2}{m}, \qquad \beta_2 = \frac{\varrho_2}{m}
\end{aligned} \tag{25}$$

Substituting Eq. (25) into Eq. (24) and then integrating Eq. (24), the adaptive control law can be expressed as:

$$\begin{aligned}
\varphi &= \varphi\,(0) + \mu_0 \int_0^t \delta\,(t)\,dt + \varrho_0 \delta\,(t) \\
\lambda_p &= \lambda_p\,(0) + \mu_1 \int_0^t \delta\,(t)\,\Delta f\,(t)\,dt + \varrho_1 \delta\,(t)\,\Delta f\,(t) \\
\lambda_d &= \lambda_d\,(0) + \mu_2 \int_0^t \delta\,(t)\,\Delta \dot{f}\,(t)\,dt + \varrho_2 \delta\,(t)\,\Delta \dot{f}\,(t)
\end{aligned} \tag{26}$$

To verify the performance of the proposed method, the force-tracking experiments on the curve surface are conducted (Figure 5) using IC, the commonly used AIC [25–29], and the proposed AIC. The parameters of the proposed method are set to $f_d = 10N$, $m = 1$, $b = 200$, $\varphi(0) = 0.09$, $\lambda_p(0) = 0.1$, $\lambda_d(0) = 0.4$, $\mu_0 = 1.5$, $\varrho_0 = 0.3$, $\mu_1 = 0.4$, $\varrho_1 = 0.1$, $\mu_2 = 0.1$, and $\varrho_2 = 0.1$. The initial value of the x_c is set to 0.193 m. The robot moves from free space to the contact space. The experiments are executed five times using each aforementioned approach. Fig 6(a) and (b) illustrate the position and force-tracking performance. The solid line and the shaded area represent the mean and standard deviation after five groups of experiments. It demonstrates the proposed AIC has better force-tracking capability without a large force overshoot compared with IC and commonly used AIC.

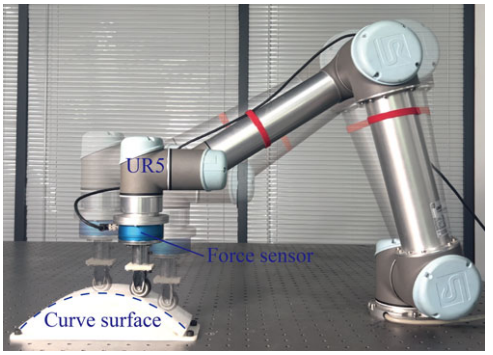

Figure 5. *Force-tracking experiment platform.*

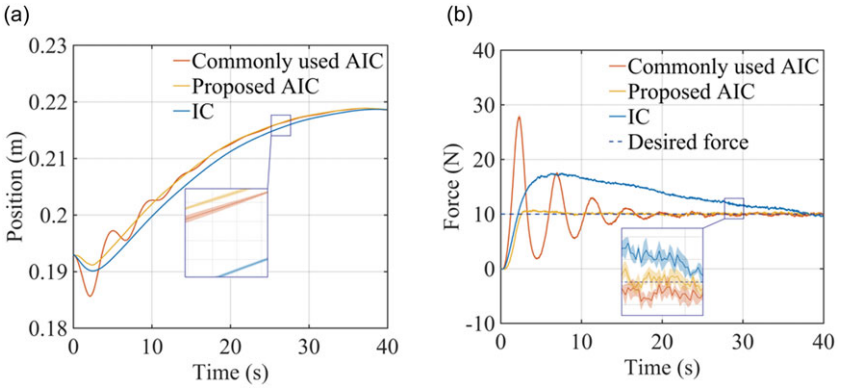

Figure 6. *Force-tracking experiments on a curved surface. (a) Position tracking. (b) Force tracking.*

4. Experiments

To evaluate the performance of the proposed method, the FFC assembly experiments are conducted in this section. First, the experiment platform is detailed presented. Second, FFC assembly experiments, which consist of three phases: picking up and stable contacting, model training, and further verification, are conducted. In the model training experiments, the robot picks up an FFC and trains for 48 episodes. In further verification experiments, the robot picks up another FFC and implements the assembly task two times according to the model trained in the model train experiments to further verify the robustness of the proposed assembly skill learning method. Third, the comparative analysis of the performance of the QL, DQL, AC, PG, PPO, and the proposed method in FFC assembly tasks is implemented.

4.1. Experiment setup

An overview of the assembly experiment platform is shown in Figure 7. This platform consists of a 6-DOF robot (Universal robot, maximum load 5 kg, communication frequency 125 Hz, repeated positioning accuracy: 0.03 mm), a six-axis force/torque sensor (NRS-6050-D80, maximum force/torque ± 500 N/ ±10 Nm, sampling rate 1000Hz, force/torque resolution 0.015 N/0.312 × 10-3 Nm) mounted on the end-effector of the robot, air pump (maximum pressure 8 kpa), a pneumatic gripper, and a computer (NUC Intel Core i7, Ubuntu 18.04, Python 3.10) communicating with robot via TCP protocol.

The FFC and connector produced by I-PEX Co., Ltd. are used to verify the effectiveness of the proposed assembly skill acquisition strategy. The length and width of the FFC are 18.5 mm and 0.3 mm, respectively. However, the length and width of the slot of the connector are 18.56 mm and 0.37 mm, respectively. The clearances between the FFC and connector are approximately 0.06 mm and 0.07 mm.

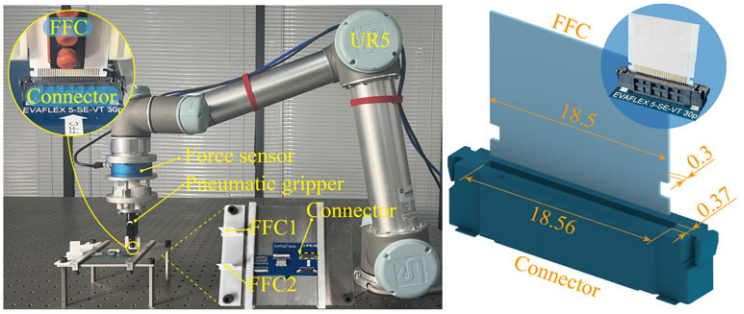

Figure 7. *Assembly experiment platform.*

Figure 8. *The procedure of the FFC assembly experiment.*

Before the assembly task is conducted, the connector is fixed on the fixture, and two FFCs are also placed on the fixture. The position for gripping the FFC is determined by demonstration.

4.2. Experiment implementation

The FFC assembly experiments consist of three phases. The procedure of the FFC assembly experiment is shown in Figure 8, and the details of the experiments can be found in the supplemental video.

1. *Picking up and stable contacting*: The robot picks up the FFC1 and approaches the connector. Then, the proposed AIC is applied to bring FFC1 and the connector into stable contact. The initial position is arbitrarily set manually with a large positional error. The parameters of AIC are set the same as the parameters of the force-tracking experiment in Section 3. According to the production specifications [38] provided by I-PEX Co., Ltd., the maximum mating force of FFC (Product model: EVAFLEX 5-SE-VT 30p) is 18N. So we referred to this parameter and chose a smaller desired contact force $f_d = 5N$ along the z-axis.
2. *Model training*: The parallel assembly skill learning algorithm and proposed AIC are combined to enable the robot to have the capability of acquiring FFC assembly skills through training. The training episode E is 48. One episode is terminated when the FFC reaches the goal depth successfully or the contact force-moment exceeds the maximum safe boundary f_b or the learning steps are above the maximum learning step $k_{max} = 500$. Then, the robot moves to the initial position and executes the next learning episode. The parameters of the skill learning algorithm are set to $\psi_1, \psi_2 = 0.01$, $\varepsilon = 0.85$, $\Delta x = 0.3mm$, $\Delta y = 0.3mm$, $\Delta\theta = 1deg$, and $\gamma = 0.9$. The number of agents in the simulation system is set to 6. $f_b = [1.5, 1.5, 10, 1]$ is set according to the accumulated experience gained through a large number of experiments.
3. *Further verification*: After 48 episodes, the model is stored, and the robot places the FFC1 back in its initial placement position and then picks up FFC2 as a new task for assembly. In order to ensure the successful gripping of the FFC, there is a clearance of approximately 0.8 mm between the fixture and the FFC. However, it results in variations in the position of the FFC within the fixture when gripping the FFC each time. The purpose of this section is to further validate the robustness of the proposed algorithm with the aforementioned uncertainties.

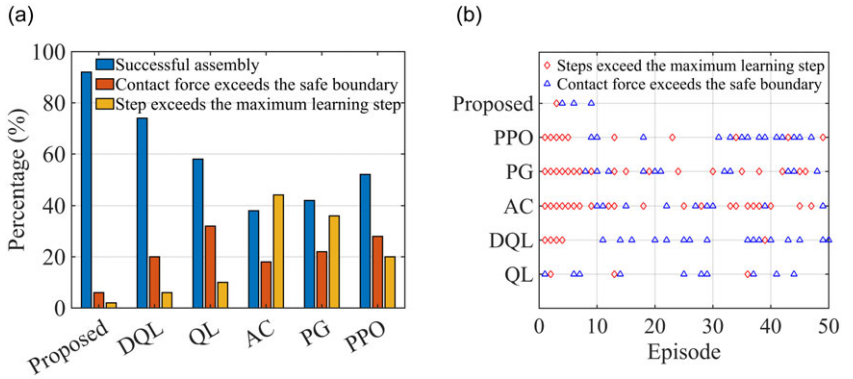

Figure 9. *Performance of different methods in FFC assembly experiments. (a) Successful and failed rate. (b) Distribution of unsuccessful assembly.*

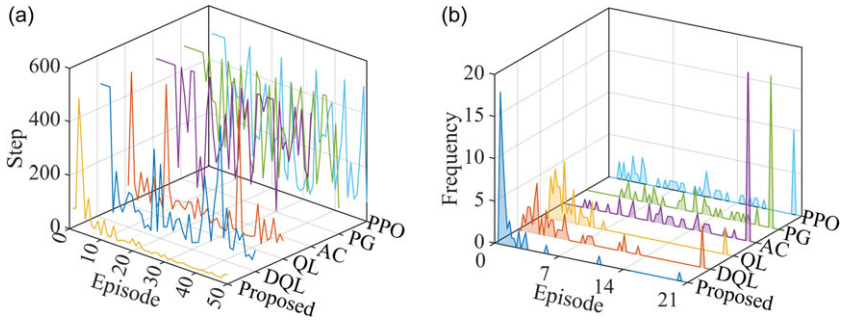

Figure 10. *Learning efficiency of different methods in FFC assembly experiments. (a) Learning step. (b) Execution time.*

4.3. Experiment results

To comprehensively evaluate the performance of the proposed method, we apply the proposed method, QL, DQL, AC, PG, and PPO, respectively, to conduct FFC assembly experiments and analyze their performance from multiple perspectives.

Figure 9 presents the comparisons of the FFC assembly performance of QL, DQL, AC, PG, and PPO and the proposed method. From Figure 9(a), it is clear that the proposed method has a higher success rate of 92% exceeding 58%, 74%, 38%, 42%, and 52% of DQL, QL, AC, PG, and PPO, respectively. The rates of failed assembly caused by contact force exceeding safe boundary and the learning steps exceeding the maximum step are 6% and 2%, which are lower than 20%, 6% of QL, 32%, 10% of DQL, 18%, 44% of AC, 22%, 36% of PG, and 28%, and 20% of PPO. Moreover, Figure 9(b) shows that the cases of failed assembly only occur in the early stages of training when using the proposed method. With the utilization of the QL, DQL, AC, PG, and PPO, these cases occur throughout the entire learning process.

It can also be observed in Figure 10(a) that the assembly steps rapidly converge to about 20 steps after training 15 episodes using the proposed method, while the assembly steps of the QL and DQL show a slight drop trend and fluctuate greatly. With the utilization of AC, PG, and PPO, there is no noticeable decrease. The main reason is that the thickness of FFC is only 0.3 mm, which makes it extremely susceptible to deformation under external force. This can result in a lack of synchronization between FFC and the robot end-effector, which is equivalent to the environment constantly changing. The assembly

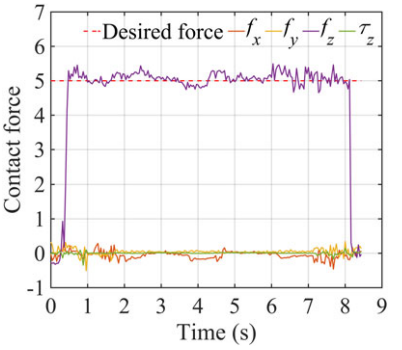

Figure 11. *Contact force between FFC and connector.*

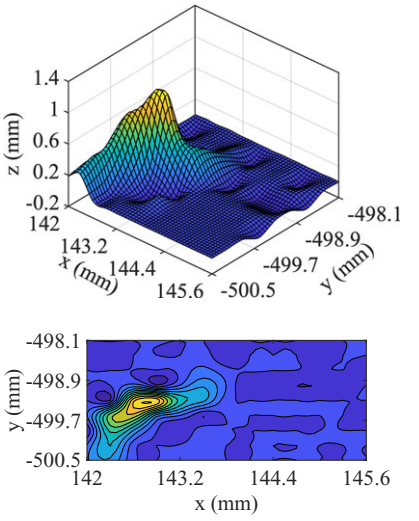

Figure 12. *Optimal policy after 50 episodes.*

experience that the agent has learned may become outdated when the environment changes, so the agent needs to adapt its model to get with the new environment quickly. However, the data inefficiency of the QL, DQL, AC, PG, and PPO impedes the learning efficiency of FFC assembly. The last two episodes of the proposed method require about 30 steps to perform the FFC2 assembly task, indicating that the proposed method is robust against the uncertain disturbance. The main reason why the number of steps for assembling FFC2 is slightly higher than FFC1 is due to the position differences when gripping FFC2 and the repeated positioning errors of the robot.

Figure 10(b) shows the execution time distribution of different methods in FFC assembly experiments. The execution time of other methods spreads more widely and is moved further right than the execution time of the proposed method, which indicates the high efficiency of the proposed method. Figure 11 shows the contact force of one episode between the FFC and connector during the assembly process. The abrupt damping of external force along the z-axis indicates the critical phase in which the position of the connector is found successfully. Then, the robot moves downwards until it reaches the goal depth or the contact force exceeds the safe boundary. Figure 12 is obtained by smoothly fitting the maximum Q-values under different states. In order to depict in a three-dimensional space, we reduced one dimension concerning the rotational state along the z-axis. It demonstrates the changing trend of the assembly policy. The agent moves towards the area with bright colors.

5. Conclusion and future work

For challenging FFC assembly tasks, an efficient assembly skill acquisition strategy is presented by combining a parallel assembly skill learning algorithm with AIC. The force-tracking experiments on the curve surface show that the proposed AIC has a better performance compared with the IC and the commonly used AIC. Thus, it solves the complex contact issues during the assembly process. The experiments of FFC assembly illustrate that the proposed skill acquisition strategy enables the robots to perform the FFC assembly task through fewer steps after training. It is robust against the disturbance of uncertain factors and has a more efficient assembly skill learning efficiency compared with other commonly used methods.

The current work still has some limitations. First, position for gripping the FFC is determined by demonstration. This will increase the workload of workers in practical applications. Second, the proposed parallel assembly skill learning algorithm requires tuning a significant number of parameters. Our future work will focus on combining the proposed method with the vision algorithm to achieve autonomous object recognition, grasping, and skill learning of FFC assembly tasks. Additionally, inspired by reference [10], we will combine the Bayesian optimization algorithm with the assembly skill learning algorithm proposed in this paper to achieve autonomous parameter tuning and further enhance learning efficiency.

Author contributions. **Xiaogang Song**: Writhing – original draft, Methodology, and Conclusion. **Peng Xu**: Supervision, Writing – review and editing, and Project administration. **Wenfu Xu**: Supervision and Writing – review and editing. **Bing Li**: Supervision, Writing – review and editing, and Project administration. **Lei Qin**: Supervision and Writing – review and editing.

Financial support. This work was supported in part by the National Natural Science Foundation of China under Grant U22A20176 and Grant 52305016, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2022B1515120078, in part by the Science and Technology Innovation Committee of Shenzhen under Grant JSGGZD20220822095401004, and in part by the Shenzhen Peacock Innovation Team Project under Grant KQTD20210811090146075.

Competing interests. The authors declare no conflict of interest.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0263574724001164.

References

[1] J. Chapman, G. Gorjup, A. Dwivedi, S. Matsunaga, T. Mariyama, B. MacDonald and M. Liarokapis, "A locally-adaptive, parallel-jaw gripper with clamping and rolling capable, soft fingertips for fine manipulation of flexible flat cables," **In:** *2021 IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China (2021) pp. 6941–6947. doi: 10.1109/ICRA48506.2021.9561970.

[2] Y. L. Yao and W. Y. Cheng, "Model-based motion planning for robotic assembly of non-cylindrical parts," *Int. J. Adv. Manuf. Technol.* **15**(9), 683–691 (1999). doi: 10.1007/s001700050119.

[3] H. Park, J. Park, D. Lee, J. Park, M. Baeg and J. Bae, "Compliance-based robotic peg-in-hole assembly strategy without force feedback," *IEEE Trans. Ind. Electron.* **6**(28), 6299–6309 (2017). doi: 10.1109/TIE.2017.2682002.

[4] Z. Zhang, Z. Zhang, X. Jin and Q. Zhang, "A novel modelling method of geometric errors for precision assembly," *Int. J. Adv. Manuf. Technol.* **94**(1-4), 1139–1160 (2018). doi: 10.1007/s00170-017-0936-3.

[5] T. Tang, H. Lin, Y. Zhao, W. Chen and M. Tomizuka, "Autonomous alignment of peg and hole by force/torque measurement for robotic assembly," **In:** *2016 EEE International Conference on Automation Science and Engineering (CASE)*, Fort Worth, TX, USA (2016) pp. 162–167. doi: 10.1109/COASE.2016.7743375.

[6] D. A. Duque, F. A. Prieto and J. G. Hoyos, "Trajectory generation for robotic assembly operations using learning by demonstration," *Robot. Comput.-Integr. Manuf.* **57**, 292–302 (2019). doi: 10.1016/j.rcim.2018.12.007.

[7] A. Kramberger, R. Piltaver, B. Nemec, M. Gams and A. Ude, "Learning of assembly constraints by demonstration and active exploration," *Ind. Robot.* **43**(5), 524–534 (2016). doi: 10.1108/IR-02-2016-0058.

[8] J. Su, Y. Meng, L. Wang and X. Yang, "Learning to assemble noncylindrical parts using trajectory learning and force tracking," *IEEE-ASME Trans. Mechatron.* **27**(5), 3132–3143 (2022). doi: 10.1109/TMECH.2021.3110825.

[9] F. J. Abu-Dakka, B. Nemec, A. Kramberger, A. G. Buch, N. Krüger and A. Ude, "Solving peg-in-hole tasks by human demonstration and exception strategies," *Ind. Robot.* **41**(6), 575–584 (2014). doi: 10.1108/IR-07-2014-0363.

[10] L. Roveda, M. Magni, M. Cantoni, D. Piga and G. Bucca, "Human-robot collaboration in sensorless assembly task learning enhanced by uncertainties adaptation via bayesian optimization," *Robot. Auton. Syst.* **136**, 103711 (2021). doi: 10.1016/j.robot.2020.103711.

[11] Z. Hou, J. Fei, Y. Deng and J. Xu, "Data-efficient hierarchical reinforcement learning for robotic assembly control applications," *IEEE Trans. Ind. Electron.* **68**(11), 11565–11575 (2021). doi: 10.1109/TIE.2020.3038072.

[12] Y. Ma, Y. Xie, W. Zhu and S. Liu, "An efficient robot precision assembly skill learning framework based on several demonstrations," *IEEE Trans. Autom. Sci. Eng.* **20**(1), 124–136 (2023). doi: 10.1109/TASE.2022.3144282.

[13] L. Roveda, J. Maskani and P. Franceschi, "Model-based reinforcement learning variable impedance control for human-robot collaboration," *J. Intell. Robot. Syst.* **100**(2), 417–433 (2020). doi: 10.1007/s10846-020-01183-3.

[14] L. Roveda, A. Testa, A. A. Shahid, F. Braghin and D. Piga, "Q-Learning-based model predictive variable impedance control for physical human-robot collaboration," *Artif. Intell.* **312**, 103771 (2022). doi: 10.1016/j.artint.2022.103771.

[15] T. Inoue, G. De Magistris, A. Munawar, T. Yokoya and R. Tachibana, "Deep reinforcement learning for high precision assembly tasks," **In:** *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Vancouver, BC, Canada (2017) pp. 819–825. doi: 10.1109/IROS.2017.8202244.

[16] J. Xu, Z. Hou, W. Wang, B. Xu, K. Zhang and K. Chen, "Feedback deep deterministic policy gradient with fuzzy reward for robotic multiple peg-in-hole assembly tasks," *IEEE Trans. Ind. Inform.* **15**(3), 1658–1667 (2019). doi: 10.1109/TII.2018.2868859.

[17] J. Luo, E. Solowjow, C. Wen, J. A. Ojea, A. M.Agogino, A.Tamar and P.Abbeel, "Reinforcement learning on variable impedance controller for high-precision robotic assembly," **In:** *2019 International Conference on Robotics and Automation (ICRA)*, Montreal, QC, Canada (2019) pp. 3080–3087. doi: 10.1109/ICRA.2019.8793506.

[18] Y.Shi, Z.Chen, H.Liu, S.Riedel, C.Gao, Q.Feng, J.Deng and J.Zhang, "Proactive action visual residual reinforcement learning for contact-rich tasks using a torque-controlled robot," **In:** *2021 IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China (2021) pp. 765–771. doi: 10.1109/ICRA48506.2021.9561162.

[19] F. Li, Q. Jiang, S. Zhang, M. Wei and R. Song, "Robot skill acquisition in assembly process using deep reinforcement learning," *Neurocomputing.* **345**, 92–102 (2019). doi: 10.1016/j.neucom.2019.01.087.

[20] G. Schoettler, A. Nair, J. Luo, S. Bahl, J. A. Ojea, E. Solowjow and S. Levine, "Deep reinforcement learning for industrial insertion tasks with visual inputs and natural rewards," **In:** *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Las Vegas, NV, USA (2020) pp. 5548–5555. doi: 10.1109/IROS45743.2020.9341714.

[21] T. Liu, B. Tian, Y. Ai, L. Li,D. Cao and F. -Y.Wang, "Parallel reinforcement learning: A framework and case study, IEEE-CAA," *J. Automatica Sin.* **5**(4), 827–835 (2018). doi: 10.1109/JAS.2018.7511144.

[22] N. Hogan, "Impedance control: An approach to manipulator," *ASME J. Dyna. Syst. Measure. Control.* **107**(1), 1–24 (1985). doi: https://doi.org/10.1115/1.3140702.

[23] L. Roveda, D. Riva, G. Bucca and D. Piga, "Sensorless optimal switching impact/force controller," *IEEE Access.* **9**, 58167–158184 (2021). doi: 10.1109/ACCESS.2021.3131390.

[24] L. Roveda, G. Pallucca, N. Pedrocchi, F. Braghin and L. M. Tosatti, "Iterative learning procedure with reinforcement for high-accuracy force tracking in robotized tasks," *IEEE Trans. Ind. Inform.* **14**(4), 1753–1763 (2018). doi: 10.1109/TII.2017.2748236.

[25] P. Wang, D. Zhang and B. Lu, "Collision detection and force control based on the impedance approach and dynamic modelling," *Ind. Robot.* **47**(6), 813–824 (2020). doi: 10.1108/IR-08-2019-0163.

[26] Z. Chen, Q. Guo, Y. Shi and Y. Yan, "Distributed cooperative control from position motion to interaction synchronization," **In:** *2022 American Control Conference (ACC)*, Atlanta, GA, USA (2022) pp. 3844–3849. doi: 10.23919/ACC53348.2022.9867144.

[27] X. Shu, F. Ni, K. Min, Y. Liu and H. Liu, "An adaptive force control architecture with fast-response and robustness in uncertain environment," In: An adaptive force control architecture with fast-response and robustness in uncertain environment," **In:** *2021 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Sanya, China (2021) pp. 1040–1045. doi: 10.1109/ROBIO54168.2021.9739648.

[28] J. Duan, Y. Gan, M. Chen and X. Dai, "Symmetrical adaptive variable admittance control for position/force tracking of dual-arm cooperative manipulators with unknown trajectory deviations," *Robot. Comput.-Integr. Manuf.* **57**, 357–369 (2019). doi: doi.org/10.1016/j.rcim.2018.12.012.

[29] K. Xu, S. Wang, B.Yue, J. Wang, H. Peng, D. Liu, Z. Chen and M. Shi, "Adaptive impedance control with variable target stiffness for wheel-legged robot on complex unknown terrain," *Mechatronics.* **69**,102388 (2020). doi: 10.1016/j.mechatronics.2020.102388.

[30] Y. Narang, B. Sundaralingam, M. Macklin, A. Mousavian and D. Fox, "Sim-to-real for robotic tactile sensing via physics-based simulation and learned latent projections," **In:** *2021 IEEE International Conference on Robotics and Automation (ICRA)*, Xi'an, China (2021) pp. 6444–6451. doi: 10.1109/ICRA48506.2021.9561969.

[31] A. A. Shahid, Y. Narang, V. Petrone, E. Ferrentino, A. Handa, D. Fox, M. Pavone and L. Roveda, "Scaling population-based reinforcement learning with GPU accelerated simulation (2024). arXiv preprint arXiv:2404.03336v3.

[32] Z. Zhang, Z. Pan and M. J. Kochenderfer, "Weighted double Q-learning," **In:** *26th International Joint Conference on Artificial Intelligence (IJCAI'17)*, AAAI Press (2017) pp. 3455–3461.

[33] F. Ferraguti, C. T. Landi, L. Sabattini, M. Bonfè, C. Fantuzzi and C. Secchi, "A variable admittance control strategy for stable physical human-robot interaction," *Int. J. Robot. Res.* **38**(6), 747–765 (2019). doi: 10.1177/0278364919840415.

[34] J. Duan, Y. Gan, M. Chen and X. Dai, "Adaptive variable impedance control for dynamic contact force tracking in uncertain environment," *Robot. Auton. Syst.* **102**, 54–65 (2018). doi: 10.1016/j.robot.2018.01.009.

[35] S. Jung, T. C. Hsia and R. G. Bonitz, "Force tracking impedance control of robot manipulators under unknown environment," *IEEE Trans. Control Syst. Technol.* **12**(3), 474–483 (2004). doi: 10.1109/TCST.2004.824320.
[36] H. Seraji and R. Colbaugh, "Force tracking in impedance control," *Int. J. Robot. Res.* **16**(1), 97–117 (1997). doi: 10.1177/027836499701600107.
[37] H. Seraji, "Decentralized adaptive control of manipulators: Theory, simulation, and experimentation," *IEEE Trans. Robot. Autom.* **5**(2), 183–201 (1989). doi: 10.1109/70.88039.
[38] I-PEX Co., Ltd. (2022). EVAFLEX 5-SE-VT product specifications, pp. 4. Available at: https://www.i-pex.com/sites/ https://www.i-pex.com/sites/.

Cite this article: X. Song, P. Xu, W. Xu, B. Li and L. Qin, "Robot skill acquisition for precision assembly of flexible flat cable with force control", Robotica. https://doi.org/10.1017/S0263574724001164