

# A Novel Robust Imitation Learning Framework for Dual-Arm Object-Moving Tasks

Weiyong Wang, Chao Zeng *Member, IEEE*, Zhenyu Lu *Member, IEEE*, Chenguang Yang, *Fellow, IEEE*

**Abstract**—Drawing inspiration from the mechanism of human skill acquisition, imitation learning has demonstrated remarkable performance. Over recent years, model-based imitation learning combined with machine learning and control theory has been continuously developed and adapted to unstructured environments. However, most results for dual-arm tasks focus on relatively safe and stable environments, which still lack robustness to generalize skills. In this work, we propose a novel robust imitation learning framework for dual-arm object-moving tasks. During demonstration, we present a shared teleoperation strategy that actively assists the operator in remotely executing dual-arm tasks, aiming to reduce the operational difficulty and stress. During modeling and generalization, we propose a Coupled Linear Parameter Varying Dynamical System (CLPV-DS), which possesses the ability to protect and restore states against possible disturbances in the environment while maintaining good tracking accuracy and stability. To address the risk of box slipping caused by disturbances, we further introduce a mutual following strategy, enabling the arms to compliantly follow each other while maintaining appropriate contact force. Considering potential obstacles in a complex generalization environment, we introduce a reactive obstacle avoidance strategy in real time that ensures global asymptotic stability. Finally, we verified the effectiveness of the proposed framework through comprehensive testing in both 2D simulations and real-robot experiments.

**Index Terms**—Imitation learning, teleoperation, dynamic system, obstacle avoidance

## I. INTRODUCTION

AS new technologies advance, traditional industries have become fully automated and intelligent. The widespread adoption of robotic systems in intelligent manufacturing is attributed to their efficiency and high-precision operational capabilities. These systems allow humans to stay away from dangerous front-line production environments, freeing them from repetitive manual tasks. This grants individuals the time and energy to participate in higher-level decision-making processes. The study of these systems has been extensive over the

years. Currently, the application scenarios for robotic systems have naturally progressed from structured environments to complex unstructured environments. This evolution brings forth new demands for the development of task planners, such as robustness, collision avoidance, and human-robot or robot-robot cooperation [1]–[3].

Single-arm robots, with their simplicity, weak coupling, and limited constraints, stand as a classic operating paradigm. However, it has obvious limitations: a dedicated end-effector is necessary, operating dexterity lacks a human-like quality, and the load constraints are minimal. In contrast, a dual-arm robot possesses all the aforementioned advantages but introduces increased complexity in planning and control [4]. This arises from its high degree of hybridity, nonlinearity, and redundancy. Dual-arm tasks are typically accompanied by collaborative constraints, categorized into unity collaboration and task-based collaboration. The former entails both arms operating the same target object simultaneously, as seen in tasks like flipping boxes [5] or folding clothes [6]. The latter, on the other hand, entails the manipulation of different objects independently, such as shaft-hole assembly [7] or cooking [8]. Regardless of collaboration type, both arms are required to follow specific collaboration constraints to successfully execute the dual-arm task. Object-moving represents a classic dual-arm collaboration task and has wide application prospects in industry. For example, in logistics sorting, various boxes need transportation from one assembly line to another; in express delivery, boxes need to be moved from delivery trucks to shelves. It requires proper contact distance and contact force at the ends of both arms. Insufficient contact force risks slippage and damage to the box, while excessive force may result in over-squeezing and damage. It necessitates careful design and application of appropriate contact force to achieve smooth handling of objects.

Learning from demonstrations, commonly referred to as imitation learning, entails a robot rapidly acquiring new skills through human demonstration. Traditional skill learning requires the involvement of a large number of professionals and requires reprogramming whenever the environmental state changes slightly. During once imitation learning, the operator demonstrates specific task skills through kinesthetic, teleoperation or passive observation [9]–[11]. It is very intuitive and easy to implement, thereby lowering the programming bar for non-professionals and saving costs. Teleoperating systems typically comprise human operators, master devices, follower devices, operating environments and communication channels. It allows the operator to remotely operate the follower arm to complete specific tasks through the master device, thus

This work was supported in part by National Nature Science Foundation of China (NSFC) under Grant U20A20200 and Grant 62003096, and Major Research Grant No. 92148204, in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2023B1515120019, in part by Industrial Key Technologies R & D Program of Foshan under Grant 2020001006308 and Grant 2020001006496.

Weiyong Wang is with College of Automation Science and Engineering, South China University of Technology, Guangzhou, China. Email: 202120117998@mail.scut.edu.cn.

Chao Zeng is with the Department of Informatics, Universität Hamburg. Email: chaozeng@ieee.org.

Zhenyu Lu is with the Faculty of Environment and Technology, Bristol Robotics Lab, University of the West of England, BS16 1QY Bristol, U.K. Email: zhenyu.lu@uwe.ac.uk.

Chenguang Yang is with Department of Computer Science, University of Liverpool, Liverpool, L69 3BX, UK. Email: cyang@ieee.org

avoiding the need to directly manipulate the end of the robotic arm and improving operational flexibility. The control signal from the master device and the haptic feedback signal from the follower device are streamed through a bidirectional communication channel. However, for dual-arm collaboration tasks, simultaneously controlling both arms separately through two master devices will cause a huge mental burden on the operator and affect the smooth execution of the task. In shared-control teleoperation, the robotic arm assumes partial autonomy to assist the operator in completing tasks and simplifying teleoperation [12]–[14]. [15] extracts task-level behavior patterns on the master device and uses them to infer various effective shared control strategies to assist the operator in remote operation. [16] assists the operator in cutting by imposing different kinematic constraints on the robot arm dynamics. These methods assist operators in performing specific task skills by designing appropriate autonomous assistance strategies.

In imitation learning, a basic behavior is often referred to as a motion primitive, serving as the smallest unit for modeling and generalization. It usually presents as a sensor-detected motion sequence directed toward a specific goal. Numerous efforts have been made to model demonstration data using Dynamical Systems (DS) while ensuring stability. DS is generally categorized into time-varying and time-invariant forms, where the former explicitly evolves with time or phase, and the latter is state-driven. The pioneering work by [17] introduced a time-varying Dynamical Movement Primitives (DMPs) with global asymptotic stability, which regarded the motion sequence as a second-order spring-damping system and a superposition of a nonlinear function. The former maintains the convergence of the model, and the latter fits the complex shape of the skill. However, it cannot simultaneously encode high-dimensional motion sequences, and the fitting effect gradually weakens over time. It can extend a lot of flexible scalability by combining control theory, such as obstacle avoidance, two-arm cooperation, upper and lower limit constraints and other improvements [18]–[25]. These time-varying DS exhibit flexibility in adjusting starting and ending points, but lack spatio-temporal robustness due to their strong time dependence, leading to unexpected behavior during disturbances. As a stochastic modeling method, [26] proposed Gaussian Mixture Model (GMM) to depict the probability distribution of motion sequences. The conditional posterior distribution is obtained through Gaussian Mixture Regression (GMR). [27] proposed the Stable Estimator of Dynamical Systems (SEDS) to convert the learned GMM into a Linear Parameter Varying Dynamical System (LPV-DS) with global asymptotic stability. [28] further introduces a physically consistent non-parametric Bayesian prior for the mixing coefficient of SEDS, allowing it to autonomously infer the appropriate number of clusters from the motion sequences. [29] use neural networks to approximate Lyapunov functions consistent with motion sequences, thereby improving generalization accuracy. [30] introduced contraction theory into SEDS, ensuring stability by decreasing the distance between trajectories according to a given matrix. [31] introduce diffeomorphisms to transform a stable well-understood system into a complex nonlinear

system. Currently, these methods still sacrifice part of the system's complexity in exchange for stability.

In this paper, we propose a robust imitation learning framework for dual-arm object-moving tasks to overcome the limitations of traditional single-arm skill learning. Our work has the following several contributions and innovations:

- We propose a teleoperation strategy to assist in moving the object. With the concept of shared control, it provides an intuitive and user-friendly remote operation demonstration experience by endowing the robot with autonomous assistance capabilities.
- We propose a Coupled LPV-DS (CLPV-DS) with two separative states. It effectively identifies disturbances and has complete disturbance recovery capabilities while maintaining global asymptotic stability. In order to prevent the box from slipping during disturbances, we further implement a mutual following strategy with compliant control to consistently maintain object clamping.
- For obstacles appearing in new scenes, we propose a reactive obstacle avoidance strategy to avoid unexpected collisions in real time. The control strategy is designed by simulating the Lorentz force, allowing flexibly change the direction of the artificial magnetic field to adjust the obstacle avoidance effect. Additionally, the strategy is proven to maintain global asymptotic stability through the application of Lyapunov stability theory.
- To verify the effectiveness of our proposed imitation learning framework, we conducted a 2D simulation and real robot scenarios to intuitively demonstrate the convergence, fitting accuracy, anti-disturbance capabilities, as well as obstacle avoidance performance.

## II. METHODOLOGY

### A. System description

The block diagram illustrating our proposed robust imitation learning framework for dual-arm object-moving tasks is shown in Fig. 1. Compared with traditional models that explore the balance between stability and fitting accuracy, we pay more attention to resist disturbances and avoid obstacles in the environment. The entire system reveals the complete learning process, including three parts: demonstration, robust generalization, and reactive strategy.

**Demonstration** It allows for the transfer of skills by directly operating a robot to complete a task. In our framework, the operator remotely controls the movement of the follower through the master, and records the motion trajectory of the robot as an instance of the skill. Based on the characteristics of dual-arm tasks, the shared teleoperation system actively assists the follower arm in task completion, enhancing operation intuitiveness and convenience.

**Robust generalization** We encode the dual-arm skills through CLPV-DS. Different from the general modeling methods, we introduce a new follower state. When a disturbance occurs, the separation of the two DS states enables the arms to promptly protect and restore the state before the disturbance. Disturbances applied to a single arm are easily detected through the state difference of the CLPV-DS. Additionally,

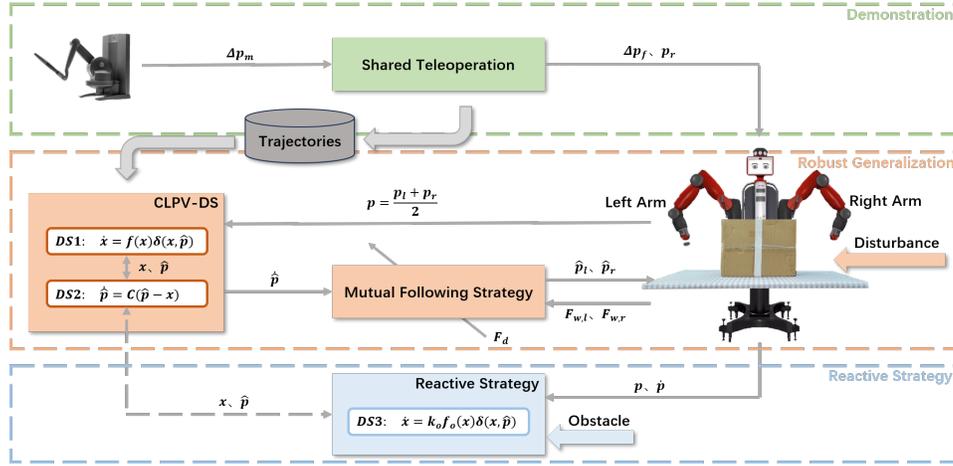


Fig. 1: The proposed imitation framework for dual-arm object-moving tasks includes three parts, i.e., demonstration (green), robust generalization (red) and reactive strategy (blue). The demonstration part allows for an intuitive and user-friendly skill acquisition. The robust generalization part encodes and generalizes dual-arm skills to new situations against uncertain disturbances. The reactive strategy part enables the robot to avoid emerging obstacles during generalization.

we incorporate a mutual following strategy with compliant control, ensuring two arms smoothly clamping the box to prevent slipping or over-squeezing.

**Reactive strategy** In an unstructured environment, traditional path planning methods cannot respond to new and unforeseen emergencies. We combine a reactive obstacle avoidance strategy with motion planning to provide real-time obstacle avoidance responses with lightweight calculations during the skill generalization process, thereby ensuring the smooth execution of tasks. Importantly, the introduction of the reactive obstacle avoidance strategy does not compromise the global asymptotic stability of CLPV-DS.

### B. Dual-arm teleoperation

In the teleoperation system, the operator manipulates the master device, and its terminal pose is recorded, stored, and converted into control commands. These commands are then transmitted to the remote follower arm through appropriate teleoperation strategies, enabling remote and flexible control. Given that the master and the follower often differ in structure, with non-corresponding joints and positions, achieving user-friendly one-to-one mapping in the task space interaction proves challenging. For some master-side operations, such as a 180° rotation, it is inconvenient for the operator. Therefore, it is necessary to consider an intuitive and easy-to-operate dual-arm teleoperation strategy.

To enhance usability, we employ a relative position matching between the end of the master and the follower, allowing the operator to decompose this large-range movement into multiple small-range movements to improve flexibility. The relative end positions of the master and follower are respectively defined as

$$\begin{aligned} \Delta p_m &= p_{m,t} - p_{m,t_0}, \\ \Delta p_f &= p_{f,t} - p_{f,t_0}, \end{aligned} \quad (1)$$

where  $p_{m(f),t}, p_{m(f),t_0} \in \mathbb{R}^3$  are the absolute end position of the master (follower) at time  $t$  and  $t_0$  respectively. The follower's relative position is designed as

$$\Delta p_f = \alpha R_{mf} \Delta p_m, \quad (2)$$

where  $\alpha$  is the scale factor, and  $R_{mf} \in \mathbb{R}^{3 \times 3}$  is the rotation matrix from the master's base coordinate system to the follower's.

To alleviate operational difficulty and mental burden on the operator, we design a simpler and more convenient single-master and dual-follower teleoperation strategy based on the constraints of the object-moving task. Since clamping the object typically requires force in one direction, for simplicity, we assume that the robot only needs to clamp along the y-axis of the robot's base coordinate system. Specifically, the left arm is used as the sole follower arm, while the right arm tracks the right side of the box accordingly to ensure that the box is securely grasped and transported

$$p_r = p_l - [0, L, 0]^T = p_f - [0, L, 0]^T, \quad (3)$$

where  $p_{l(r)} \in \mathbb{R}^3$  is the absolute end position of the left (right) arm,  $L$  is the length of the box on the y-axis of the robot's base coordinate system.

During demonstration, the operator completes the teaching process through our single-master and dual-follower teleoperation system, recording and storing the motion trajectories of both arms. We further select the midpoint of the absolute end position of both arms as the demonstration data

$$p = \frac{p_l + p_r}{2} \quad (4)$$

It is worth noting that if we choose to record and train the absolute end positions of both arms separately, it involves substantial redundant degrees of freedom. Moreover, the distance constraint between the arms cannot be guaranteed due to training errors, causing the box to slip and the task to fail. In order to fully ensure task constraints and computational

efficiency, our solution is to refine and learn the midpoint, then restore the two arms' end position based on the object parameters during generalization, thereby reestablishing the task constraints.

### C. Coupled LPV-DS

Since skills can be modeled as mapping functions, the goal of skill learning is to transform limited motion skill observations into a concise set of function parameters. Autonomous DS is a function from the state to its first-order differential, pursuing global asymptotic convergence and spatio-temporal robustness. In this paper, we use GMM-based LPV-DS with quadratic Lyapunov (QLF) function constraints for the first DS encoding. Similar to SEDS [27], GMM learning needs to be introduced before learning LPV-DS

$$p(x) = \sum_{k=1}^K \pi_k \mathcal{N}(x | \mu_k, \Sigma_k), \quad (5)$$

where  $x$  and  $\dot{x}$  are position and velocity,  $K$  is the number of the total mixture components. The component prior  $\pi_k$  (satisfying the constraint  $\sum_{k=1}^K \pi_k = 1$ ), the mean  $\mu_k$  and the variance  $\Sigma_k$  collectively characterize the features of a GMM. Traditional GMM learning requires manual selection of a wide range of mixing component numbers  $K$ , with the optimal  $K$  being selected after conducting multiple sets of learning error tests. To address this problem, [32] further introduces the physically consistent non-parametric Bayesian prior to the GMM learning process. This integration automatically infers the number of mixing components  $K$  based on the data characteristics, avoiding blind manual selection and laborious testing. In Dirichlet process sampling through the Chinese Restaurant Process (CPR), physical consistency provides a locality and directionality between sampled clusters, making the spatial distribution of the clusters more reasonable.

A nonlinear LPV-DS can be expressed as a mixture of multiple linear DSs

$$\dot{x} = \sum_{k=1}^K \gamma_k(x) (A_k x + b_k) \stackrel{def}{=} f(x), \quad (6)$$

where  $\{A_k, b_k\}$  is the parameter set of the  $k$ -th linear DS. Its mixing coefficients are defined by the previously learned GMM and the current state, that is  $\gamma_k(x) = \frac{\pi_k \mathcal{N}(x | \mu_k, \Sigma_k)}{p(x)}$ .

Assuming that we obtain  $N$  demonstration trajectories  $\{\{p_{d,t}^{(n)}, \dot{p}_{d,t}^{(n)}\}_{t=1}^{T_n}\}_{n=1}^N$  from the teleoperation process, the parameters of LPV-DS can be solved through convex quadratic optimization on minimizing the speed error

$$\min_{\theta} J(\theta) = \sum_{n=1}^N \sum_{t=1}^{T_n} \|f(x_{d,t}^{(i)}) - \dot{x}_{d,t}^{(i)}\| \quad (7)$$

To satisfy global asymptotic stability, each linear system component needs to satisfy a conservative QLF constraint

$$\begin{cases} (A_k)^T + A_k = Q_k, Q_k = (Q_k)^T \prec 0 \\ b_k = -A_k x_g \end{cases}, k = 1, \dots, K, \quad (8)$$

where  $Q_k$  expands the search domain of parameter space, and  $x_g$  is the global convergence point. These conditions are not

difficult to derive from the Lyapunov function  $V(x) = (x - x_g)^T(x - x_g)$ .

Due to the simple Lyapunov function formula, the above optimization of LPV-DS still encounters the dilemma of accuracy and stability. Although the generalization accuracy in the demonstration range is high, it is still not ideal in other ranges. To address the potential impact of possible disturbances that may cause the motion to deviate from the demonstration range, we consider a strategy for complete trajectory recovery. We introduce a second DS

$$\dot{\hat{p}} = C(x - \hat{p}), \quad (9)$$

where  $C$  is a scaling factor to adjust the speed at which state  $\hat{p}$  converges to state  $x$ . It is adaptively determined by

$$C = \dot{\hat{p}}_{max} (1 - e^{-k_c \|\hat{p} - x\|}), \quad (10)$$

where  $\dot{\hat{p}}_{max}$  is the maximum speed allowed to execute,  $k_c$  is a hyperparameter used to adjust the convergence speed. (9) depends on the state  $x$  evolved by (6), thus coupling two different DS. We further modify (6) to

$$\dot{x} = f(x) \delta(x, \hat{p}) \quad (11)$$

$$\delta(x, \hat{p}) = \begin{cases} 1, & x = \hat{p} \\ 0, & x \neq \hat{p} \end{cases} \quad (12)$$

Here state  $x$  represents the desired generalized motion trajectory, and state  $\hat{p}$  implements a virtual following planning. When  $\hat{p} = x$ ,  $x$  starts to evolve along the desired trajectory, and then (9) will make  $\hat{p}$  converge to the new state  $x$ . This coupling method serves to protect the task execution even when a disturbance occurs. In the event of a disturbance causing  $\hat{p}$  to deviate,  $x$  remains unchanged, and  $\hat{p}$  will return to the original deviation position. It is worth mentioning that CLPV-DS still has the ability to generalize skills based on different initial positions; when suffering from disturbances, CLPV-DS will improve generalization accuracy. The global asymptotic stability of the final DS will be given in subsection II-E with Lyapunov analysis.

### D. Generalization with mutual following strategy

For the generalization of dual-arm tasks, a core concern lies in fully accounting for the collaborative relationship between the two arms, that is, maintaining task constraints between the arms. For the task of object-moving, the ends of the arms need to be kept at a suitable distance. Additionally, the box must be grasped stably to avoid sliding or falling off, and must not be damaged due to excessive force. During generalization, we preserve the stability and spatio-temporal robustness provided by the CLPV-DS modeling, and then recover the motion trajectories of the arms according to the constraints.

Since the process of object-moving only involves position constraints along a single degree of freedom, we assume that the two opposing clamping surfaces of the box are located on both sides of the  $y$ -axis in the robot base coordinate system. The generalized positions of both arms are generalized as

$$\begin{aligned} \hat{p}_l &= \hat{p} + [0, \frac{L}{2}, 0]^T \\ \hat{p}_r &= \hat{p} - [0, \frac{L}{2}, 0]^T \end{aligned} \quad (13)$$

Unfortunately, this midpoint-based generalization is not fully robust to disturbances for the dual-arm collaboration task. When a disturbance is applied to a single arm, the task constraints will be directly destroyed. In fact, when the disturbance occurs, the error in estimating the center of the box through (4) will increase, and the desired positions of the arms can no longer be restored through (13). Since the CLPV-DS separates the states of the two DS, the disturbance can be detected by the two-state deviation. At this point, we adopt the mutual following strategy for both arms

$$\hat{p}_l = \begin{cases} p_r + [0, L, 0]^T, & |x - p| > \delta \\ \hat{p} + [0, \frac{L}{2}, 0]^T, & |x - p| \leq \delta \end{cases} \quad (14)$$

$$\hat{p}_r = \begin{cases} p_l - [0, L, 0]^T, & |x - p| > \delta \\ \hat{p} - [0, \frac{L}{2}, 0]^T, & |x - p| \leq \delta \end{cases},$$

where  $\delta$  is the deviation threshold,  $p_l(r)$  is the actual end positions of the left (right) arm.

In order to reduce the hysteresis impact of the disturbance response and enable the arms to generate appropriate contact force against the box, we introduce variable stiffness admittance control. It regards the contact behavior of the end as a second-order spring-damping system

$$M\ddot{e}_{y,\chi} + D\dot{e}_{y,\chi} + K_{v,\chi}e_{y,\chi} = F_{w,\chi}, \quad (15)$$

where,  $M$ ,  $D$  and  $K_v$  are the arm's inertia matrix, Coriolis and centrifugal matrices and gravity respectively.  $\chi = \{l, r\}$  represents left arm and right arm respectively.  $F_{w,\chi}$  is the interaction force between the left or right arm and the outside world, and  $e_{y,\chi}$  is the position error of the left or right arm along the y-axis. The variable stiffness is further designed as [33]

$$K_{v,\chi} = \frac{k_p e_{f,\chi} + k_d \dot{e}_{f,\chi}}{e_{y,\chi}}, \quad (16)$$

where  $e_{f,\chi} = F_{w,\chi} - F_d$  is the force error,  $F_d$  is the desired contact force. Combining (15) and (16), we can get the dynamic equation of the desired position  $y_c$  on the y-axis

$$\begin{cases} \ddot{y}_{c,\chi}(t) = \frac{(k_p + 1)e_{f,\chi}(t) + k_d \dot{e}_{f,\chi}(t) + F_d(t) - D\dot{y}_{c,\chi}(t)}{M\dot{y}_{c,\chi}(t)}, \\ \dot{y}_{c,\chi}(t+1) = y_{c,\chi}(t) + \dot{y}_{c,\chi}(t)\Delta t \end{cases}, \quad (17)$$

where  $\Delta t$  is the discrete calculation step of the computer. We get the final hybrid force-position controller

$$p_{c,\chi} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} p_\chi + \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} y_{c,\chi} \quad (18)$$

The introduction of a hybrid force-position controller in the end-effector coordinate system can enhance the compliance of box contact and maintain human-like gripping ability. When one arm experiences a disturbance, due to the sharp reduction in contact force, the impedance control gives the other arm more distance compensation on the y-axis, which greatly alleviates the slip caused by state lag.

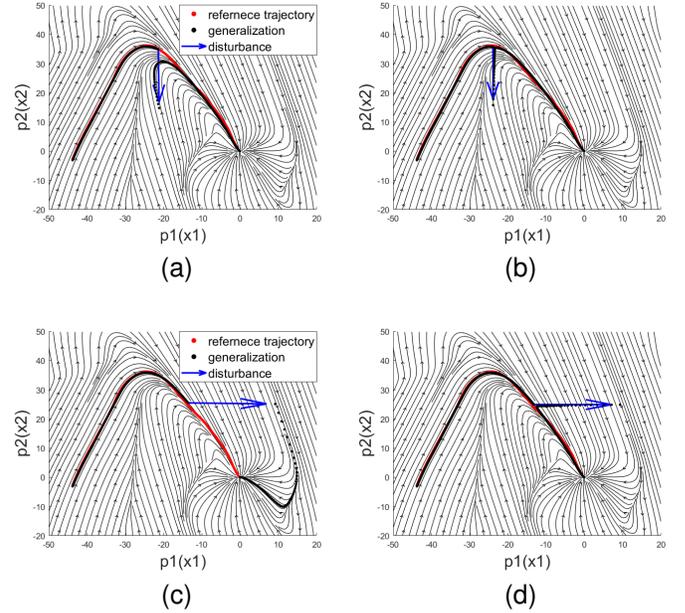


Fig. 2: Rollouts under disturbances at different time. (a), (c) represent the learning results of the original LPV-DS, and (b), (d) represents the learning results of the CLPV-DS.

### E. Reactive collision avoidance

Skill generalization means the reproduction of skills in new contexts. However, when a new obstacle is present in the environment, the generalization of LPV-DS is not reactive enough: its behavior only depends on the current state but does not effectively respond to the new environment. In this section we introduce a reactive obstacle avoidance strategy for CLPV-DS, aiming to minimize the impact on global asymptotic stability. When considering the motion of state  $x$  as the motion of a charged particle, similar to the Lorentz force, we introduce an additional control for obstacle avoidance

$$f_o(x) = \frac{x_g - x}{\|x_g - x\|} \times B, \quad (19)$$

where  $B$  is the artificial magnetic field

$$B = c \times \dot{d}, \quad (20)$$

where  $c$  is the artificial current, and  $d = x_o - x$  is the distance between the box and the center of the obstacle  $x_o$ . Its selection is flexible and produces different obstacle avoidance effects. Here we design as

$$c = \dot{d} - \frac{d}{\|d\|} \left( \frac{d}{\|d\|} \cdot \dot{d} \right) \quad (21)$$

We incorporate (19) into CLPV-DS

$$\dot{x} = [f(x) + k_o f_o(x)]\delta(x, \hat{p}) \quad (22)$$

For the coefficient  $k_o$ , we design as

$$k_o = \frac{\|d\|}{\|d\| - r}, \quad (23)$$

where  $r$  is the safe distance from the center of the obstacle  $x_o$ . As the box approaches an obstacle, a greater magnitude

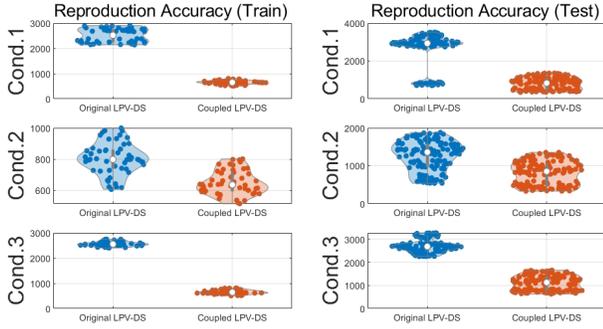


Fig. 3: The generalization accuracy comparison between original LPV-DS and CLPV-DS on the LASA dataset under three types of disturbances.

of escape velocity will be superimposed in the deviation direction.

We can easily verify the stability of the proposed DS using Lyapunov stability theory. Consider a differentiable energy-inspired Lyapunov candidate function of the following form

$$V(x) = (x - x_g)^T(x - x_g) + (\hat{p} - x)^T(\hat{p} - x), \quad (24)$$

where  $x_g$  is the attractive equilibrium point. Its gradient can be further derived

$$\begin{aligned} \dot{V}(x) &= (x - x_g)^T \dot{x} + \dot{x}^T (x - x_g) + \\ &\quad (\hat{p} - x)^T (\dot{\hat{p}} - \dot{x}) + (\dot{\hat{p}} - \dot{x})^T (\hat{p} - x) \\ &= (x - x_g)^T \dot{x} + \dot{x}^T (x - x_g) + C(\hat{p} - x)^T (x - \hat{p}) + \\ &\quad C(x - \hat{p})^T (\hat{p} - x) - (\dot{\hat{p}} - \dot{x})^T \dot{x} - \dot{x}^T (\dot{\hat{p}} - \dot{x}) \\ &< (x - x_g)^T \dot{x} + \dot{x}^T (x - x_g) \\ &= \{(x - x_g)^T [f(x) + k_o f_o(x)] + \\ &\quad [f(x) + k_o f_o(x)]^T (x - x_g)\} \delta(x, \hat{p}) \\ &= \{(x - x_g)^T \sum_{k=1}^K \gamma_k(x) (A_k x + b_k) + \\ &\quad \sum_{k=1}^K \gamma_k(x) (A_k x + b_k)^T (x - x_g) + \\ &\quad \frac{2k_o}{\|x_g - x\|} (x - x_g)^T [(x_g - x) \times B]\} \delta(x, \hat{p}) \\ &= \{(x - x_g)^T \sum_{k=1}^K \gamma_k(x) (A_k + A_k^T) (x - x_g)\} \delta(x, \hat{p}) \\ &\leq 0 \end{aligned} \quad (25)$$

Therefore, our reactive obstacle avoidance strategy is globally asymptotically stable for the attractive point  $x_g$ .

### III. EXPERIMENTS

#### A. Simulations

In this section, we verify the robust generalization and reactive obstacle avoidance strategy of our proposed CLPV-DS through 2D simulations. We validate the effectiveness in the LASA handwriting dataset [34] and provide visual representations of them.

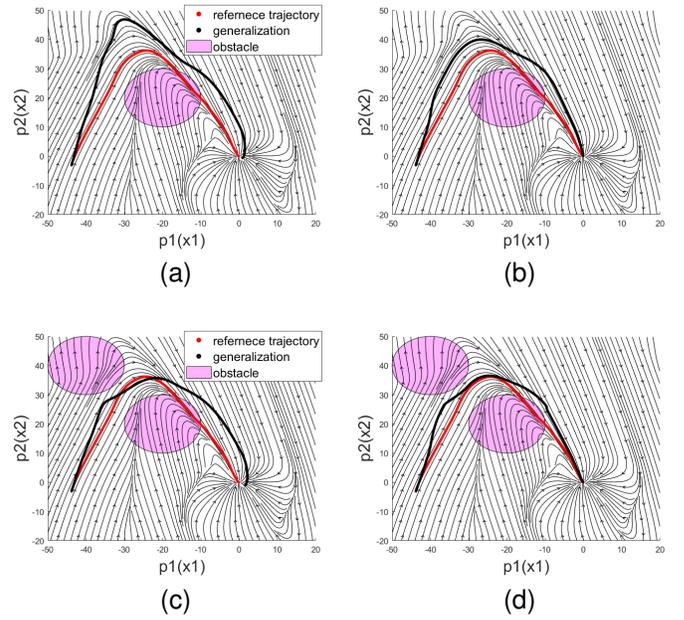


Fig. 4: Rollouts under various number of obstacles. (a), (c) represent the APF, and (b), (d) represent our reactive obstacle avoidance strategy (ROAS).

Fig. 2 illustrates the learning effect of CLPV-DS on the “Angle” group in LASA. The flow field indicates a high fitting accuracy and global asymptotic stability of the model. Disturbances of  $[0, -20]^T$  (Fig. 2a,2b) and  $[20, 0]^T$  (Fig. 2c,2d) are applied to the original LPV-DS (left column) and CLPV-DS (right column) respectively. The trajectory of the original LPV-DS may be significantly distorted after encountering a disturbance deviation (as shown in Fig. 2c), affecting its fitting accuracy. In contrast, with the introduction of the coupling formula, the state of the system is protected, fully restoring it to the pre-disturbance state. CLPV-DS exhibits spatial generalization capabilities even with different initial positions, protecting only the state where unexpected deviations occur during generalization. Fig. 3 further shows the comparison results between original LPV-DS and CLPV-DS across the entire LASA dataset. Dynamic Time Warping Distance (DTWD) [35] is employed to assess the similarity of trajectories with different lengths

$$E = \frac{1}{T_i} \sum_{t=1}^{T_i} \mathcal{S}(x_{d,t}^{(i)}, x_{d,t+1}^{(i)}, x_t^{(i)}, x_{t+1}^{(i)}), \quad (26)$$

where  $\mathcal{S}$  denotes the closed area between four vertices. Similarly, we assess the error by applying the following three types of disturbances with different amplitudes and directions at the midpoint of the path: (Cond. 1)  $[0, -20]^T$ ; (Cond. 2)  $[0, -10]^T$ ; (Cond. 3)  $[20, 0]^T$ . The training set and test set are divided in a 3:4 ratio. The results consistently show that our method has good generalization accuracy under different disturbances.

In Fig. 4, we compare the obstacle avoidance effects of the Artificial Potential Field (APF) (Fig. 4a,4b) and our reactive

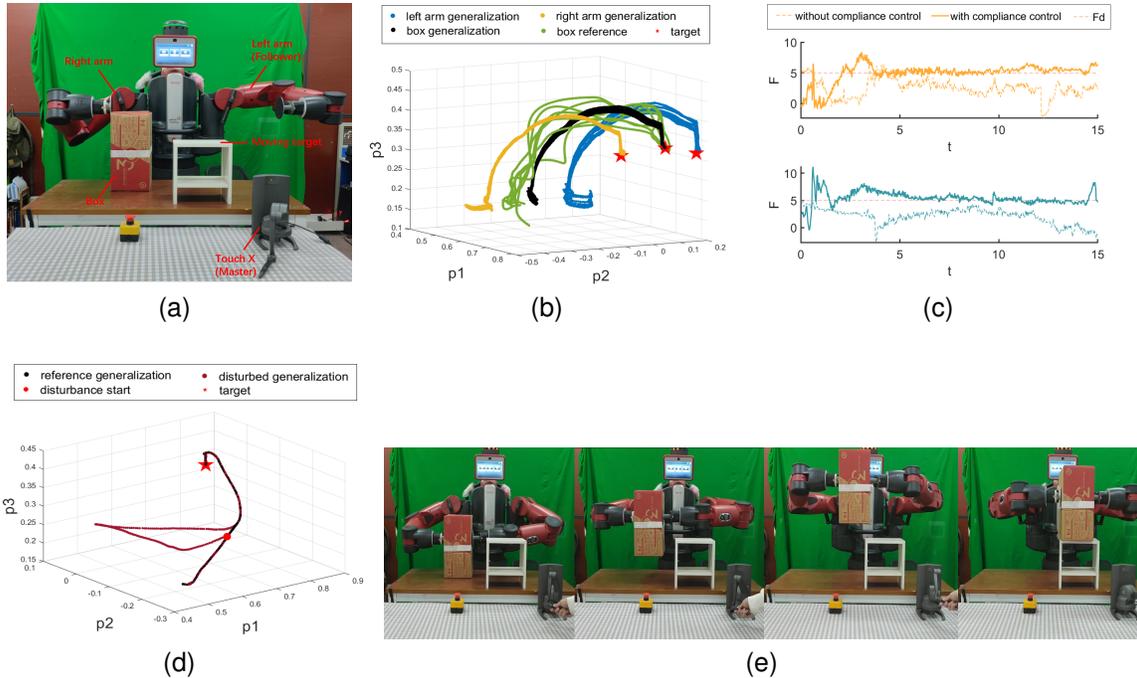


Fig. 5: (a) Experiment setup. (b) Demonstration and normal generalization. (c) Comparison of z-axis forces on motion control and compliance control, including the left arm (top) and the right arm (bottom). (d) A generalization under a teleoperated disturbance. (e) Shared teleoperation process during demonstration.

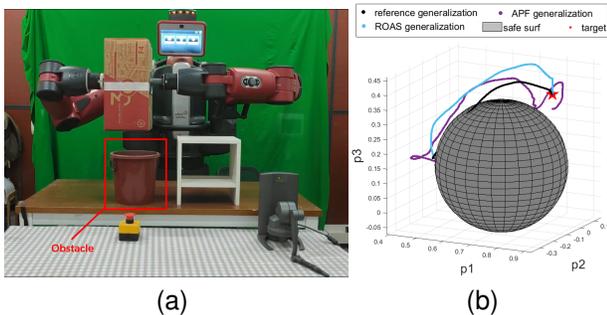


Fig. 6: (a) A flowerpot is added to the scene to act as the obstacle. (b) The comparison results of APF and ROAS.

obstacle avoidance strategy (Fig. 4c,4d). Obviously, APF only considers the obstacle avoidance repulsion, which causes a shift in the global attractive point. Our method still guarantees the initial global attractive point and has less impact on the trajectory.

### B. Real-world box-moving task

In this section, we validate the effectiveness of our proposed robust imitation learning framework through a dual-arm box-moving task. The experimental setup, as shown in Fig. 5a, includes the Baxter robot, Touch X, a box and a shelf platform serving as the moving target. Touch X serves as the master for shared teleoperation demonstrations, controlling the left arm of the Baxter as the follower. The task involves moving the box to

the shelf platform safely. The process of shared teleoperation is shown in Fig. 5e. As shown in Fig. 5b, 7 demonstrations are performed, recording and storing the midpoints of both arms for CLPV-DS learning. During generalization, we perform 7 rollouts and they all successfully moving the box to the shelf platform, and the trajectories are smooth and stable. To achieve smooth and fast tracking, we select  $\dot{p}_{max} = 10$ ,  $k_c = 100$  in CLPV-DS. Setting  $\dot{p}_{max}$  to an excessively large value may lead to instabilities and unsafe motion. Conversely, if it is too small, it will cause tracking to be too slow and affect the execution time of the task. Compliance control with  $k_p = 25$ ,  $k_d = 2$  and a desired force  $F_d = 5N$  is applied to ensure secure clamping without damaging the box. Fig. 5c shows the forces on the left and right arms during generalization. With compliance control, the overall force on both arms remains close to the desired force. But without compliance control, the force on the arms is smaller, making the box prone to rotation or falling during movement, leading to task failure. To demonstrate the disturbance recovery capabilities of our generalization, we apply disturbances to the left arm through teleoperation, as shown in Fig. 5d. Despite the disturbance causing the left arm to deviate, the mutual following strategy allows the box to remain securely gripped. After releasing the left arm, the box returns to its position before the deviation occurs and continues to complete the task.

To assess the obstacle avoidance capability of our framework, we added a flowerpot to block the original generalization path, as shown in Fig. 6a. Without any obstacle avoidance capabilities, the box must collide with the flowerpot. Taking into account the volume of the flowerpot and the box, we

set the radius of the safety sphere  $r = 0.23m$ , that is, the trajectory of the box outside the sphere must not collide. Fig. 6b reveals the obstacle avoidance results between APF and our Reactive Obstacle Avoidance Strategy (ROAS). Consistent with the results in the simulations, the repulsive field of the APF affected the equilibrium position, causing the box to finally fail to land on the shelf platform. For APF, setting a small repulsion coefficient makes it challenging for the trajectory to stay away from obstacles, while a large coefficient leads to unstable pulses. ROAS enables CLPV-DS to plan a stable obstacle avoidance trajectory with improved smoothness.

#### IV. CONCLUSION

In summary, this paper proposes a novel robust dual-arm imitation learning framework capable of learning and generalizing the box moving skill. For data acquisition, we designed a shared teleoperation system based on task constraints to provide a simple and user-friendly interface. To improve robustness, we propose CLPV-DS capable of fast recovery from disturbances and generalization to both arms based on constraints. Incorporating the mutual following strategy with compliance control, even in the face of strong disturbances, the constraints between the arms and the appropriate contact force can be ensured, ensuring the smooth completion of the task. Finally, considering obstacles in complex environments, we design a reactive obstacle avoidance strategy to achieve obstacle avoidance while preserving global asymptotic stability. In the experimental section, we validate and intuitively compare the proposed imitation framework in the 2D simulation and the dual-arm box-moving task in the real environment, achieving good generalization effects.

However, our method still has limitations: the lack of sensing modules for dynamic obstacle detection; reliance on actual object parameters for box handling. In future work, we aim to address these shortcomings by introducing visual pose detection [36], [37] to actively identify and properties of the box and obstacles in the environment, aiming to avoid dynamic obstacles autonomously.

#### REFERENCES

- [1] R. Wu and A. Billard, "Learning from demonstration and interactive control of variable-impedance to cut soft tissues," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 5, pp. 2740–2751, 2021.
- [2] D. Kappler, F. Meier, J. Issac, J. Mainprice, C. G. Cifuentes, M. Wüthrich, V. Berenz, S. Schaal, N. Ratliff, and J. Bohg, "Real-time perception meets reactive motion generation," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1864–1871, 2018.
- [3] A. Ajoudani, A. M. Zanchettin, S. Ivaldi, A. Albu-Schäffer, K. Kosuge, and O. Khatib, "Progress and prospects of the human–robot collaboration," *Autonomous Robots*, vol. 42, pp. 957–975, 2018.
- [4] J. Zhu, M. Gienger, G. Franzese, and J. Kober, "Do you need a hand?—a bimanual robotic dressing assistance scheme," *IEEE Transactions on Robotics*, vol. 40, pp. 1906–1919, 2024.
- [5] H. Huang, C. Zeng, L. Cheng, and C. Yang, "Toward generalizable robotic dual-arm flipping manipulation," *IEEE Transactions on Industrial Electronics*, 2023.
- [6] L. Sun, G. Aragon-Camarasa, S. Rogers, and J. P. Siebert, "Autonomous clothes manipulation using a hierarchical vision architecture," *IEEE Access*, vol. 6, pp. 76646–76662, 2018.
- [7] M. Alles and E. Aljalbout, "Learning to centralize dual-arm assembly," *Frontiers in Robotics and AI*, vol. 9, p. 830007, 2022.
- [8] K. Takata, T. Kiyokawa, I. G. Ramirez-Alpizar, N. Yamanobe, W. Wan, and K. Harada, "Efficient task/motion planning for a dual-arm robot from language instructions and cooking images," in *2022 IEEE/RISJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 12058–12065.
- [9] C. Yang, C. Zeng, P. Liang, Z. Li, R. Li, and C.-Y. Su, "Interface design of a physical human–robot interaction system for human impedance adaptive skill transfer," *IEEE Transactions on Automation Science and Engineering*, vol. 15, no. 1, pp. 329–340, 2017.
- [10] C. Zeng, C. Yang, Q. Li, and S. Dai, "Research progress on human-robot skill transfer," *Acta Automatica Sinica*, vol. 45, no. 10, pp. 1813–1828, 2019.
- [11] Y. Du, J. Jian, Z. Zhu, D. Pan, D. Liu, and X. Tian, "A syntactic method for robot imitation learning of complex sequence task," *Robotic Intelligence and Automation*, vol. 43, no. 2, pp. 132–143, 2023.
- [12] S. Islam, P. X. Liu, A. E. Saddik, R. Ashour, J. Dias, and L. D. Seneviratne, "Artificial and virtual impedance interaction force reflection-based bilateral shared control for miniature unmanned aerial vehicle," *IEEE Transactions on Industrial Electronics*, vol. 66, no. 1, pp. 329–337, 2019.
- [13] C. Yang, Y. Zhang, G. Zhao, and L. Cheng, "Human-robot shared control system based on 3d point cloud and teleoperation," *Science China Technological Sciences*, pp. 1–9, 2023.
- [14] Y. Zhu, C. Yang, Z. Tu, Y. Ling, and Y. Chen, "A haptic shared control architecture for tracking of a moving object," *IEEE Transactions on Industrial Electronics*, vol. 70, no. 5, pp. 5034–5043, 2023.
- [15] D. Rakita, B. Mutlu, M. Gleicher, and L. M. Hiatt, "Shared control-based bimanual robot manipulation," *Science Robotics*, vol. 4, no. 30, p. eaaw0955, 2019.
- [16] R. Rahal, F. Abi-Farraj, P. R. Giordano, and C. Pacchierotti, "Haptic shared-control methods for robotic cutting under nonholonomic constraints," in *2019 IEEE/RISJ international conference on intelligent robots and systems (IROS)*. IEEE, 2019, pp. 8151–8157.
- [17] A. J. Ijspeert, J. Nakanishi, H. Hoffmann, P. Pastor, and S. Schaal, "Dynamical movement primitives: learning attractor models for motor behaviors," *Neural computation*, vol. 25, no. 2, pp. 328–373, 2013.
- [18] D.-H. Zhai, Z. Xia, H. Wu, and Y. Xia, "A motion planning method for robots based on dmeps and modified obstacle-avoiding algorithm," *IEEE Transactions on Automation Science and Engineering*, 2022.
- [19] Z. Lu, N. Wang, and D. Shi, "Dmeps-based skill learning for redundant dual-arm robotic synchronized cooperative manipulation," *Complex & Intelligent Systems*, vol. 8, no. 4, pp. 2873–2882, 2022.
- [20] C. Zeng, Y. Li, J. Guo, Z. Huang, N. Wang, and C. Yang, "A unified parametric representation for robotic compliant skills with adaptation of impedance and force," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 2, pp. 623–633, 2021.
- [21] C. Zeng, H. Su, Y. Li, J. Guo, and C. Yang, "An approach for robotic leaning inspired by biomimetic adaptive control," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 3, pp. 1479–1488, 2021.
- [22] Z. Lu, N. Wang, and C. Yang, "A constrained dmeps framework for robot skills learning and generalization from human demonstrations," *IEEE/ASME Transactions on Mechatronics*, vol. 26, no. 6, pp. 3265–3275, 2021.
- [23] J. Li, M. Cong, D. Liu, and Y. Du, "Enhanced task parameterized dynamic movement primitives by gmm to solve manipulation tasks," *Robotic Intelligence and Automation*, vol. 43, no. 2, pp. 85–95, 2023.
- [24] L.-H. Kong, W. He, W.-S. Chen, H. Zhang, and Y.-N. Wang, "Dynamic movement primitives based robot skills learning," *Machine Intelligence Research*, vol. 20, no. 3, pp. 396–407, 2023.
- [25] J. Dong, W. Si, and C. Yang, "A novel human-robot skill transfer method for contact-rich manipulation task," *Robotic Intelligence and Automation*, 2023.
- [26] E. Pignat and S. Calinon, "Bayesian gaussian mixture model for robotic policy imitation," *IEEE Robotics and Automation Letters*, vol. 4, no. 4, pp. 4452–4458, 2019.
- [27] S. M. Khansari-Zadeh and A. Billard, "Learning stable nonlinear dynamical systems with gaussian mixture models," *IEEE Transactions on Robotics*, vol. 27, no. 5, pp. 943–957, 2011.
- [28] N. B. Figueroa Fernandez and A. Billard, "A physically-consistent bayesian non-parametric mixture model for dynamical system learning," *Proceedings of Machine Learning Research*, 2018.
- [29] Z. Jin, D. Qin, A. Liu, W.-A. Zhang, and L. Yu, "Learning neural-shaped quadratic lyapunov function for stable, accurate and generalizable human–robot skills transfer," *Robotics and Computer-Integrated Manufacturing*, vol. 82, p. 102526, 2023.
- [30] H. C. Ravichandar, I. Salehi, and A. P. Dani, "Learning partially contracting dynamical systems from demonstrations," in *CoRL*, 2017, pp. 369–378.

- [31] K. Neumann and J. J. Steil, "Learning robot motions with stable dynamical systems under diffeomorphic transformations," *Robotics and Autonomous Systems*, vol. 70, pp. 1–15, 2015.
- [32] N. B. Figueroa Fernandez and A. Billard, "A physically-consistent bayesian non-parametric mixture model for dynamical system learning," *Proceedings of Machine Learning Research*, 2018.
- [33] X. Zhang, H. Zhou, J. Liu, Z. Ju, Y. Leng, and C. Yang, "A practical pid variable stiffness control and its enhancement for compliant force-tracking interactions with unknown environments," *Science China Technological Sciences*, vol. 66, no. 10, pp. 2882–2896, 2023.
- [34] S. K.-Z. A. Billard, "Handwriting human motion dataset, in: Lasa data," <https://github.com/justagist/pyLasaDataset>, 2015.
- [35] H. Ravichandar, I. Salehi, and A. Dani, "Learning partially contracting dynamical systems from demonstrations," in *Proceedings of the 1st Annual Conference on Robot Learning*, 2017, pp. 369–378.
- [36] Y. Chen, J. Jiang, R. Lei, Y. Bekiroglu, F. Chen, and M. Li, "Graspada: Deep grasp adaptation through domain transfer," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 10 268–10 274.
- [37] Y. Tu, J. Jiang, S. Li, N. Hendrich, M. Li, and J. Zhang, "Posefusion: Robust object-in-hand pose estimation with selectlstm," *arXiv preprint arXiv:2304.04523*, 2023.



**Weiyong Wang** received the B.S. degree in the School of Automation Science and Engineering, South China University of Technology, Guangzhou, China, in 2021. He is currently pursuing the master's degree with the School of Automation Science and Engineering, South China University of Technology. His current research interests include robotics, intelligent control, and imitation learning.



**Chao Zeng** received the Ph.D. degree in pattern recognition and intelligent systems from the South China University of Technology, Guangzhou, China, in 2019. He visited the Department of Informatics, University of Hamburg (UHH), Hamburg, Germany, from 2018 to 2019. He has been a research fellow with UHH since July 2020. His research interests include robot learning and control, and physical human–robot interaction.



**Zhenyu Lu** (M'21) received the Ph.D degree in Northwestern Polytechnical University, Xi'an, China in 2019. He is currently working as a senior research fellow in Bristol Robotic Laboratory & University of the West of England, Bristol. His research interests include teleoperation, human–robot interaction and intelligent learning method.



**Chenguang Yang** (Fellow, IEEE) received the B.Eng. degree in measurement and control from Northwestern Polytechnical University, Xian, China, in 2005, and the Ph.D. degree in control engineering from the National University of Singapore, Singapore, in 2010. He performed postdoctoral studies in human robotics at the Imperial College London, London, U.K from 2009 to 2010. He is Chair in Robotics with Department of Computer Science, University of Liverpool, UK. He was awarded UK EPSRC UKRI Innovation Fellowship and individual

EU Marie Curie International Incoming Fellowship. As the lead author, he won the IEEE Transactions on Robotics Best Paper Award (2012) and IEEE Transactions on Neural Networks and Learning Systems Outstanding Paper Award (2022). He is the Corresponding Co-Chair of IEEE Technical Committee on Collaborative Automation for Flexible Manufacturing. His research interest lies in human robot interaction and intelligent system design.