Human-Agent Joint Learning for Efficient Robot Manipulation Skill Acquisition

Shengcheng Luo[1*], Quanquan Peng[1*], Jun Lv[1*], Kaiwen Hong[2],
Katherine Rose Driggs-Campbell[2], Cewu Lu[1], Yong-Lu Li[1]

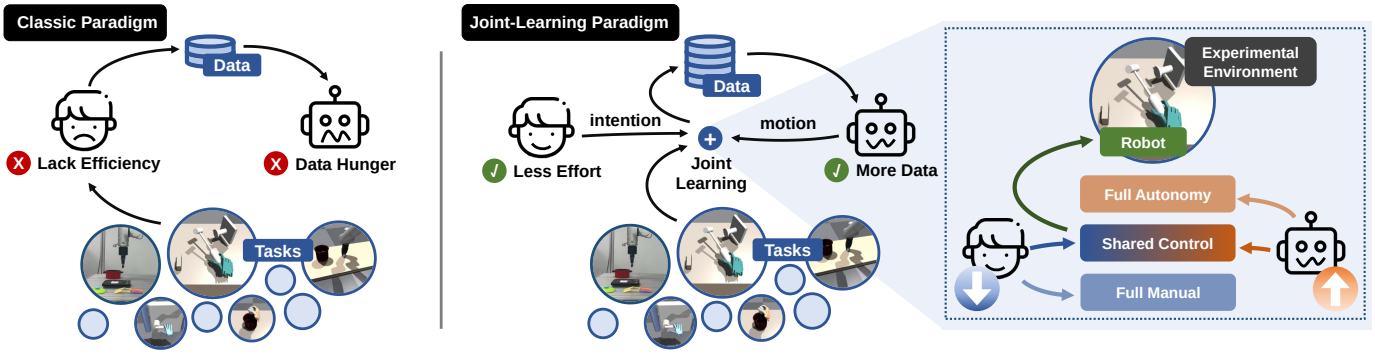


Fig. 1: **Human-agent joint learning overview.** Traditional frameworks typically separate human and agent training, requiring operators first to learn the task environment before data collection. This often leads to inefficiencies due to delayed and insufficient data gathering. In our framework, we integrate human and agent training from the start in a joint learning model. This enables simultaneous development and adapts the agents to human operation more effectively, enhancing overall efficiency and promoting better collaboration between humans and machines allowing for human effortless adaptation data collection.

***Abstract*—Employing a teleoperation system for gathering demonstrations offers the potential for more efficient learning of robot manipulation. However, teleoperating a robot arm equipped with a dexterous hand or gripper, via a teleoperation system presents inherent challenges due to the task's high dimensionality, complexity of motion, and differences between physiological structures. In this study, we introduce a novel system for joint learning between human operators and robots, that enables human operators to share control of a robot end-effector with a learned assistive agent, simplifies the data collection process, and facilitates simultaneous human demonstration collection and robot manipulation training. As data accumulates, the assistive agent gradually learns. Consequently, less human effort and attention are required, enhancing the efficiency of the data collection process. It also allows the human operator to adjust the control ratio to achieve a trade-off between manual and automated control. We conducted experiments in both simulated environments and physical real-world settings. Through user studies and quantitative evaluations, it is evident that the proposed system could enhance data collection efficiency and reduce the need for human adaptation while ensuring the collected data is of sufficient quality for downstream tasks. *For more details, please refer to our webpage https://norweig1an.github.io/HAJL.github.io/*.**

I. INTRODUCTION

In recent years, significant progress has been made on learning robot manipulation policies from demonstrations. Previous studies have utilized teleoperation systems [1–6] to collect human demonstrations, and learning-based policies [7–9] have been formulated using the gathered data. Despite the notable advancements, several challenges still need to be addressed. For example, in vision-based teleoperation systems, even with state-of-the-art 3D hand pose estimation algorithms [10–13], errors persist that significantly degrade the teleoperation. Additionally, discrepancies between the structures of human hands and robot end-effectors, along with the lack of haptic feedback during contact-rich manipulation, also pose challenges. As a result, current teleoperation systems demand substantial human effort, in addition, collecting high-quality datasets remains a challenging and labor-intensive task in many scenarios.

Naturally, a question was raised: *in data-collection, how to make human effort less while improving the data quality?* Here, we aim to address this question and argue that human-agent joint learning can help. That said, an effective and efficient teleoperation system should be designed to preferentially capture the operator's intentions for directing a robot end effector and pose the *main frame*, while concurrently enabling an autonomous agent to help us ensure motion stability and *interpolate* the details. To this end, we propose a framework that achieves shared control between the human and a learned assistive agent during data collection. As shown in Fig. 1, our *human-agent joint learning* framework seeks to integrate the data collection and policy learning to enhance the efficiency of the whole process, reducing human effort, and improving the data quality.

Given our human-agent joint learning approach, we allow the data acquisition agent to grow and learn along with the human operator. Inspired by shared autonomy [14–16], we introduce a novel teleoperation system that enables collaboration between humans and learning-based agents to control a robot jointly during the data collection and learning process. In particular, our proposed system provides the flexibility to adjust a "control ratio" between the human operator and a

* denotes equal contribution
[1]Shanghai Jiao Tong University, [2]University of Illinois Urbana-Champaign.

learning-based agent. A lower control ratio, in the beginning, emphasizes the human's role in teaching the agent finer-grained knowledge under the structure of human intention and principal actions. As the agent's learning improves, a higher ratio indicates greater autonomy from the learned agent to replace the human effort to "inpaint" the whole process given only human intention and principal actions.

With the proposed system, the human effort will be reduced due to the shared control during data collection. Additionally, the agent learning process is integrated with the data collection, improving the efficiency of the whole process. In addition, the quality of the collected data is also improved, benefiting different kinds of downstream tasks. We conducted experiments in six different simulation environments using two types of end-effectors: a dexterous hand and a gripper. Additionally, we performed experiments on three real-world tasks to validate our findings. Evaluation results indicate that our proposed system significantly enhances data collection efficiency, increasing the collection success rate by 30% and nearly doubling the collection speed. Additionally, data collected in shared autonomy mode is as effective for downstream tasks and models as data collected directly from the teleoperation system, demonstrating comparable validity. Our main contributions are summarized as follows:

- We study how to reduce human adaptation while keeping data quality in teleoperation data collection and propose a human-agent joint learning paradigm.
- We build a system that fosters concurrent development between the human operator and assistive agent, which not only streamlines the learning process but also expedites the robot's ability to perform robot manipulations autonomously.
- Conducting both simulation and real-world experiments to demonstrate the efficiency and effectiveness of our proposed system. Our system achieved significant performance improvements, including a **30%** increase in data collection success rate and **double** the collection speed.

II. RELATED WORKS

Teleoperation for Data Collection. Data has always been a crucial foundation, and robots are no exception. Teleoperation serves as a significant source for collecting robot data [7, 17–22]. Some works achieve teleoperation through wearable devices [1–4, 23], and vision-based teleoperation systems offer a low-cost and easily developed alternative [5, 6, 24, 25]. For instance, [25] utilizes neural networks for markerless vision-based teleoperation of dexterous robotic hands from depth images. [5] set up a vision-based teleoperation system to control the Allegro Hand, accomplishing various contact-rich manipulation tasks in the real world. Recently, [6] introduced AnyTeleop, a unified teleoperation system designed to accommodate various arms, hands, realities, and camera setups within a singular framework. In this paper, we introduce a joint learning paradigm to assist teleoperation by sharing control between the human operator and a learning-based agent, improving the efficiency of data collection using teleoperation.

Interactive robot learning. Collecting fine-grained human demonstration data for robotic manipulation is an effective but labor-intensive and time-consuming way to enable robots to complete a wide range of tasks [26, 27]. Previous work uses shared autonomy to assist people with disability in performing tasks by arbitrating human inputs and robot actions [28]. Many of the shared autonomy algorithms aim to estimate human intents from a set of pre-defined goals [29–32], using clothoid curves to parametrize the state and control [33] or by mapping low-dimension control input to high-dimension robot actions [28, 34]. In this work, we introduce a system that integrates the agent's learning process with data collection, facilitating both data collection and robot learning.

III. PROPOSED METHOD

The primary contribution of this work is the development of a novel and highly efficient data collection method. To achieve this, the system is designed in two key stages: first, the proposed system allows human operators to control the robot via a teleoperation system to gather an initial but insufficient training dataset, which serves as the foundation for the second stage (Sec. III-B). Second, using these data, we train a diffusion-model-based assistive agent (Sec. III-C) to establish shared control between the human operator and the agent, thereby improving the efficiency of the data collection process (Sec. III-D). This approach mirrors the concept of "bootstrapping" [35], where, as more data is accumulated, the system progressively reduces the effort required from human operators, facilitating further data collection and iterative system improvement. Additionally, once sufficient data has been gathered, the system offers the option to transition the shared control agent to full autonomy.

A. Preliminary

To get a learned agent in Sec. III-C, enabling human-agent joint learning, we follow the Denoising Diffusion Probabilistic Model (DDPM) [36] training paradigm. Here we first briefly introduce the DDPM algorithm. The *forward process* of the Diffusion Model can be regarded as adding Gaussian noise to the data x^0 according to a variance schedule $\beta_{1:K}$ by

$$x_k = \sqrt{\alpha_k} x_{k-1} + \sqrt{1-\alpha_k}\varepsilon, \tag{1}$$

where $\varepsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \alpha_k = 1 - \beta_k$. DDPM models the output generation as a denoising process (Stochastic Langevin Dynamics). A line of works [8, 37–39] use diffusion model to generate the action for agents: given x^K sampled from Gaussian noise $\mathcal{N}(\mathbf{0}, \mathbf{I})$, it utilizes a parameterized diffusion process to model how x^K is denoised in order to get noise-free action x^0 by

$$p_\theta(x^0) = \int p(x^K) \prod_{k=1}^{K} p_\theta(x^{k-1}|x^k) \mathrm{d}x^{1:K}, \tag{2}$$

where $p_\theta(x^{k-1}|x^k) = \mathcal{N}(\mu_\theta(x^k, k), \Sigma(x^k, k))$ is usually referred as *reverse process*. [40] shows that $p_\theta(x^{t-1}|x^k)$ becomes

tractable when conditioned on x_0 and Eq. 2 can be reformulated as minimizing the error in the noise prediction. [36] simplify the training loss function as

$$\mathcal{L} := \mathbb{E}_{k,x_0,\varepsilon\sim\mathcal{N}(\mathbf{0},I)}\left[\|\varepsilon - \varepsilon_\theta(x_k(x_0,\varepsilon),k)\|_2^2\right], \tag{3}$$

where step k is sampled uniformly as $k \in [1,K]$, ε_θ is the noise prediction model. During the inference phase, we can generate x_0 by recursively sample $z \sim \mathcal{N}(\mathbf{0},I)$:

$$x_{k-1} = \mu_\theta(x_k,k) + \sigma_k z. \tag{4}$$

Similar to [8, 41], with the collected trajectory $\{(s_i,a_i)\}_{i=0}^T$, we aim to train an agent to imitate the trajectory, accomplishing a specific task $\mathcal{T}$. Therefore, we utilize DDPM to capture the conditional distribution of $p(a|s)$ and the training loss in Eq. 3 shall be modified as

$$\mathcal{L} := \mathbb{E}_{k,(s_i,a_i),\varepsilon\sim\mathcal{N}(\mathbf{0},I)}\left[\|\varepsilon - \varepsilon_\theta(a_i+\varepsilon,s_i,k)\|_2^2\right]. \tag{5}$$

B. Teleoperation System.

Our system initially captures the raw sensory signal $\mathcal{I}$. Human hand pose $\mathcal{P}^h$ can be obtained from the captured signal using off-the-shelf 3D hand pose estimation [10, 11, 13]. The pose $\mathcal{P}^h$ consists of the positions of the human hand's keypoints. Then, employing an inverse kinematic function f_{IK}, we compute the action of the robot $a \in \mathbb{R}^m$, such that $a = f_{IK}(\mathcal{P}_t^h, \mathcal{P}_{t+1}^h)$, where it is calculated upon the change in the hand pose. Given this teleoperation system, the human operator will move the hand to produce a sequence of hand poses $\{\mathcal{P}_i^h\}_{i=0}^T$ to teleoperate the robot with an action sequence $\{a_i\}_{i=0}^T$ to achieve the task $\mathcal{T}$. The human collected demonstration trajectory $\{(s_i,a_i)\}_{i=0}^T$, where $s \in \mathbb{R}^n$ is the robot state, could be used for downstream tasks.

C. Diffusion-Model-Based Assistive Agent.

After collecting data via teleoperation mentioned in Sec. III-B, we train a diffusion-model-based assistive agent to learn how to assist humans in collecting data in a shared control manner.

At an abstract level, the diffusion-model-based assist agent, noted as $f(\cdot|\cdot)$, is provided with the state s, denoising step number k, and a noise action a^k, which could be an imperfect action gathered from the teleoperation system or sampled from a Gaussian distribution, to predict the desired action

$$a = f(a^k|s,k). \tag{6}$$

During data collection, the proposed system offers the option to control the robot in a shared control mode rather than directly applying the collected action a^h from the teleoperation system. This leads to a reduced human workload during the data collection process. The classical shared autonomy method is achieved through the equation [29]:

$$a^s = \gamma a^h + (1-\gamma)a^r, \tag{7}$$

where a^r is generated by the learned agent. However, considering that the agent operates as a diffusion policy (Fig. 3), we blend the action from the human with the forward and reverse processes. Given action a^h, a forward process diffuses the action as follows: $a^k = a^h + \varepsilon^k$. Subsequently, a reverse process denoises the action a^k:

$$a^s = f(a^k|s,k). \tag{8}$$

By applying action a, the control of the robot is shared between the human and the diffusion-model-based assistive agent. We can adjust the control ratio $\gamma = k/K$ between the human operator and the diffusion-model-based assistive agent by varying k. When $\gamma = 0$, the action a^s represents the teleoperation action a^h, which is the dexterous robot directly controlled by a human operator. As γ approaches 1.0, the action a^s transitions to full autonomy a^r. A higher γ value indicates a higher level of autonomy, allowing the learning-based agent more control rights to stabilize and direct the dexterous hand.

Algorithm 1 Overall Process

Require: The human operator $\mathcal{H}$;
Ensure: The collected dataset $\mathcal{D}$; assistive agent f; control ratio γ;
1: Initialization: $\mathcal{D} \leftarrow \emptyset, \gamma \leftarrow 0$;
2: **while** $|\mathcal{D}|$ is small **do** ▷ not enough data is collected
3: $\mathcal{H}$ collects data d under f's help; ▷ see III-D2 for control ratio adjustment
4: **if** d is valid **then**
5: $\mathcal{D} \leftarrow \{d\} \cup \mathcal{D}$;
6: **end if**
7: Finetune f with $\mathcal{D}$;
8: **end while**
9: **return** $\mathcal{D}$ and f;

D. Integrating Data Collection and Manipulation Learning.

In this section, we show how to integrate data collection and manipulation learning into a unified framework that progressively reduces human effort and enhances robot autonomy.

1) Detailed Algorithm Explanation

We outline the overall process in Algo. 1. The assistive agent is trained in three steps as follows:

Step 1. Initially, we collect a dataset for pre-training agent f under full manual control by human operators, *i.e.*, with the control ratio $\gamma = 0$.

Step 2. Given the initial dataset, we train a relatively low performance assistive agent to aid in further data collection. The training process has been formulated in Eq. 5 and Eq. 6, where a neural network ε_θ is trained to predict noise ε out of the noisy action a^k.

Step 3. The trained agent assists in a second data collection round, aiming for higher efficiency and success. We then refine the agent using data from both rounds to enhance its performance. This cycle repeats until the agent achieves full autonomy and the required data volume is collected.

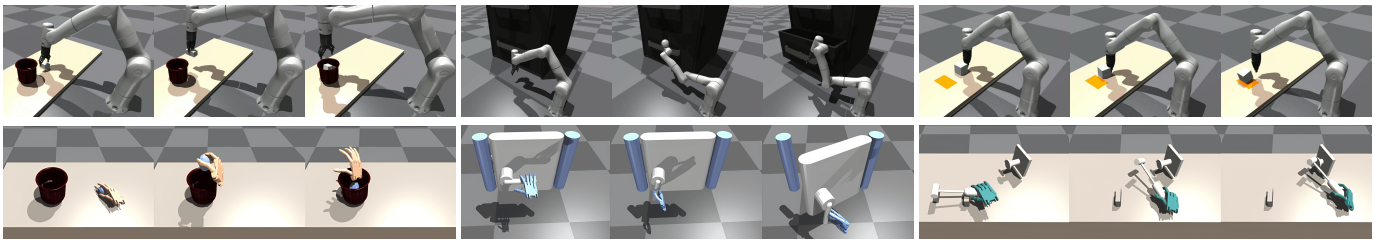


Fig. 2: **Simulation tasks overview.** Here are six task settings and their task flow for Pick-and-Place *(left)*, Articulated-Manipulation *(middle)*, Gripper-Push *(upper-right)* and Dexterous-Tool-Use *(bottom-right)*.

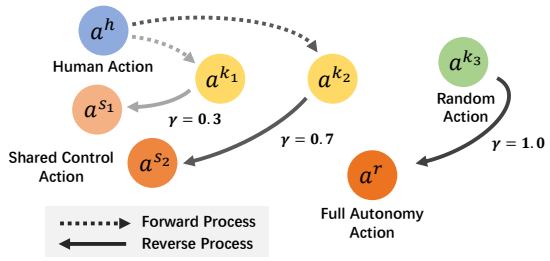


Fig. 3: **Diffusion based shared control.** To achieve shared control between the human and agent, we blend the action from the human operator a^h using the forward and reverse process. The parameter γ governs the control ratio, where a lower γ results in the action better aligning with the human operator's intention. In contrast, a higher γ allows the learned agent to exert more influence over the blended action.

2) Control Ratio Adjustment

For each data collection, we offer users two options to adjust the control ratio γ: (1) Users can empirically adjust γ based on their needs. (2) Alternatively, set $\gamma = \frac{1}{2}(1+\cos\theta)$, where θ is obtained by calculating the dot product of the previous timestep's human action a^h and shared action a^s. This assesses alignment, increasing γ for positive alignment to enhance agent control, and reducing it for misalignment to increase user control.

After obtaining the control ratio γ, we calculate the shared action a^s, using the human operator's action a^h as input, as defined in Eq. 8 (shown in Fig. 3).

IV. EXPERIMENTS

In this section, we introduce the settings for both real world tasks and simulation tasks, along with the experimental results and data validation. Due to page limitations, we have included some detailed information such as training details and ablation study results on our webpage.

A. Tasks.

We adopt six multi-stage manipulation tasks (Fig. 2). *Pick-and-Place* aims at picking an object on the table and placing it into a container. *Articulated-Manipulation*'s objective for the dexterous hand is to grasp and unscrew a door handle to open it, while for the gripper, it is to grab a drawer handle and pull the drawer open. *Push-cube* requires the robot to push the cube to the target position. *Tool-Use* aims at picking a hammer and using it to drive a nail into a board.

B. Efficiency of Data Collection.

Our proposed system leverages shared control between human operators and learned agents to enhance the efficiency of data collection. To learn how the assistant agent could improve the data collection process, we conducted a user study.

In the user study, 10 human operators participate, collecting data under two modes: one where control is shared between the operator and the learned agent (*w/ Ours*), and the other where control is directly by the operator alone (*w/o Ours*). Each participant is instructed to collect as much data as possible within three minutes under two different modes for three dexterous hand tasks. Three metrics are evaluated: *Success Rate* (Percent) indicates the percentage of attempts where data collection was successful. *Horizon Length* (Steps per Sample) measures the length of each collected trajectory, with a lower horizon length indicating smoother data collection. *Collection Speed* (Samples per Hour) refers to the number of successful trajectories that can be collected in one hour.

In Tab. I, by sharing control between humans and learned agents, our system shows improvements in both success rate and collection speed, while the average horizon length of the collected trajectories is reduced. This suggests that our system enhances the efficiency of data collection by facilitating a process that is easier to succeed, faster, and more fluid in terms of trajectory smoothness. To ensure the fairness of the experiment wasn't compromised, we equally divided the user group into two parts, *Group 1* first collected data directly by themselves *(w/o Ours)* and then collected data with an assistive agent *(w/ Ours)*, while the *Group 2* reversed the order, first *(w/ Ours)* mode and then *(w/o Ours)* mode.

C. Quantitative Evaluation.

To gain deeper insight into how the learned agent assists the human operator, we visualize several keyframes from the data collection process of three dexterous hand tasks. From Fig. 4, it is evident that human operators are not required to provide too precise control with the assistive agent facilitating shared control over the dexterous hand. Instead, they only need to convey high-level intentions, such as the direction of hand movement or finger grasp motions. In multi-stage tasks, like picking up a hammer and then using it to drive a nail, operators only need to provide a *trigger action* to guide the agent to transition from one sub-stage to the next. As a result, less effort and attention are required, making the data collection easier to execute successfully and speeding it up.

		Pick-and-Place			*Door-Open*			*Tool-Use*		
		Success Rate ↑	*Horizon Length* ↓	*Collection Speed* ↑	*Success Rate* ↑	*Horizon Length* ↓	*Collection Speed* ↑	*Success Rate* ↑	*Horizon Length* ↓	*Collection Speed* ↑
Group1	*w/ Ours*	**86.96**	**219.01**	**320**	**87.11**	**142.29**	**460**	**66.50**	**232.17**	**200**
	w/o Ours	51.53	378.49	176	62.49	258.27	252	42.38	487.95	129
Group2	*w/ Ours*	**94.06**	**214.16**	**324**	**80.29**	**134.16**	**424**	**55.55**	**275.71**	**172**
	w/o Ours	45.42	471.48	120	53.45	317.21	176	34.47	511.03	124

TABLE I: User studies on three dexterous hand tasks.

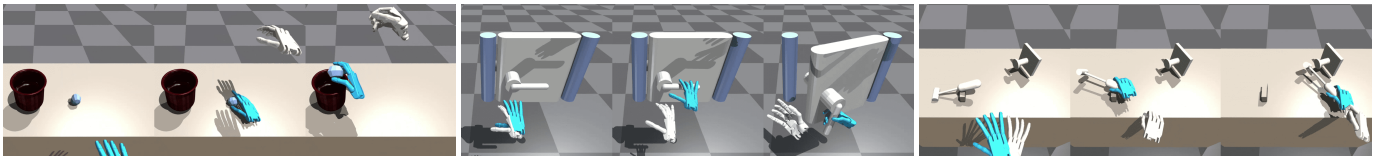


Fig. 4: **Shared control process overview.** The white one is the hand controlled purely by the *human operator*, while the cyan one is under *shared control* between the human and the assistive agent.

When the learned agent shares control with users, the system effectively corrects imperfect human control signals to accomplish specific tasks. Given the challenge of directly measuring the level of imperfection in user signals and the correction ability of our system, we simulate human input using a baseline agent trained with Behavior Cloning (BC) as a proxy for user control.

In Fig. 5, additional data collected under our framework effectively contributes to training improvements. The graph illustrates that with limited data availability, the agent can assist the simulated operator more effectively. As the agent gains access to and trains on more data, its ability to correct actions improves. These results indicate that our system gradually reduces the demand for the operator's attention and effort, thereby enhancing the overall efficiency of data collection process.

Furthermore, once sufficient data is collected and the assistive agent is trained, it can transition into full autonomy mode by setting $\gamma = 1$ and denoising actions from the noise sampled from Gaussian distributions. Across three different dexterous manipulation tasks, we can achieve success rates of 0.76, 0.78, and 0.89, indicating that the assistive agent can effectively transform into an automated dexterous manipulation agent.

From our experiments, we have observed that the assistive agent significantly aids human operators in managing fine control, especially in scenarios where accurate observation by humans is challenging. For instance, tasks such as grasping an egg or moving a hammer present visual challenges. It can be difficult to visually confirm whether the egg is securely grasped or if there's a risk of it being dropped. This uncertainty makes it hard for human operators to react promptly to sudden changes. However, within our proposed joint learning framework, human operators are primarily required to focus on high-level intentions and task planning during data collection, while the assistive agent manages the detailed low-level actions. This division of labor significantly reduces the burden on human operators by clearly separating strategic planning from execution tasks, streamlining the collaboration between humans and machines.

Dexterous Hand	*Pick-and-Place*		*Articulated-Manipulation*		*Tool-Use*	
	$40\mathcal{H}$	$10\mathcal{H}$ + $30\mathcal{S}$	$40\mathcal{H}$	$10\mathcal{H}$ + $30\mathcal{S}$	$40\mathcal{H}$	$10\mathcal{H}$ + $30\mathcal{S}$
BC	0.30	**0.50**	0.22	**0.57**	0.39	**0.40**
BC-RNN	0.54	**0.67**	0.47	**0.50**	**0.27**	0.25
DP	0.73	**0.76**	0.77	**0.78**	0.88	**0.89**
Parallel Gripper	*Pick-and-Place*		*Articulated-Manipulation*		*Push-cube*	
	$40\mathcal{H}$	$10\mathcal{H}$ + $30\mathcal{S}$	$40\mathcal{H}$	$10\mathcal{H}$ + $30\mathcal{S}$	$40\mathcal{H}$	$10\mathcal{H}$ + $30\mathcal{S}$
BC	0.42	**0.44**	0.35	**0.37**	**0.88**	0.85
BC-RNN	**0.39**	0.36	0.71	**0.73**	0.59	**0.67**
DP	0.51	**0.60**	0.42	**0.67**	**0.83**	0.82

TABLE II: Data quality on downstream tasks.

	Dexterous Tool-Use		*Gripper Push-cube*	
	BC	*DP*	*BC*	*DP*
$10\mathcal{H}$	0.29	0.45	0.23	0.42
$10\mathcal{H}$ + $10\mathcal{H}$	0.28	0.67	0.37	0.78
$10\mathcal{H}$ + $20\mathcal{H}$	0.28	0.82	0.51	0.67
$10\mathcal{H}$ + $30\mathcal{H}$	0.39	0.88	0.88	0.83
$10\mathcal{H}$ + $10\mathcal{S}$	0.31	0.71	0.33	0.81
$10\mathcal{H}$ + $20\mathcal{S}$	0.30	0.79	0.61	0.62
$10\mathcal{H}$ + $30\mathcal{S}$	0.40	0.89	0.85	0.82

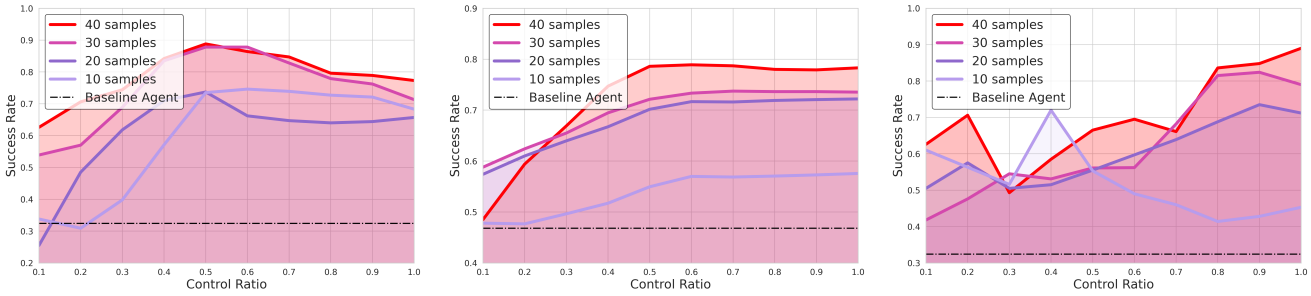
TABLE III: Agent performance under increasing data.

D. Data Quality on Downstream Task.

In this section, we illustrate that collecting data under shared control does not compromise the quality of the data. We gather dexterous hand and gripper manipulation demonstrations via the proposed system in two modes: fully controlling the robots by a human ($\mathcal{H}$) and sharing control ($\mathcal{S}$) between the human operator and the learned assistive agent. And utilize these data to train different kinds of agents, like BC, BC-RNN [7], and Diffusion Policy (DP) [8].

In Tab. II, compared to directly collecting human demonstrations from the expert human operator, who can achieve success rates and efficiency comparable to those with agent assistance, the data collected by sharing control between the human and the assistive agent can achieve comparable or even surprisingly better results with BC and BC-RNN. Their results are comparable with DP, possibly as DP can better fit the tasks, which is in line with [8].

In Tab. III, we compare the effects of using different sets of data to train BC and DP. We can find that utilizing more data collected under the shared control mode leads to comparable performance on the tool-use and push-cube tasks. This verifies



(a) *Pick-and-Place* (b) *Articulated* (c) *Tool-Use*

Fig. 5: **Agent performance over time.** The *x-axis* represents the control ratio γ and the *y-axis* represents the success rate. We train a simulated operator to evaluate our system, it shows that even with limited data, the learned assist agent can improve the success rate of data collection to improve the efficiency. With the data accumulated, the performance of the learned agent keeps rising. Moreover, the learned agent could be transitioned to a full autonomy agent ($\gamma = 1.0$).

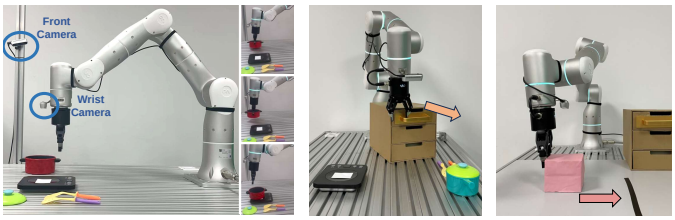


Fig. 6: **Real world setting.** 1. *Pick-and-Place*: use the gripper to pick the red pot up and place it onto the black induction cooker. 2. *Articulated-Manipulation*: use the gripper to open the drawer. 3. *Push-cube*: use the gripper to push the cube across the black line.

	Success Rate ↑	*Horizon Length* ↓	*Collection Speed* ↑
w/ Ours	**0.79**	**18.72**	**151**
w/o Ours	0.70	21.54	121

TABLE IV: Real world gripper Pick-and-Place task user study.

that the new data contributes significantly to policy learning and can achieve a similar effect compared to the data from human experts but at a much lower cost. These results indicate that the data collected under the proposed paradigm have sufficient quality and efficiency for downstream tasks.

E. Real World Experiment and User Feedback.

To better evaluate our system, we further conduct real-world experiments. Three tasks are adopted: Pick-and-Place, Articulated-Manipulation, and Push-cube in Fig. 6. Following the same rules as Sec. IV-B, four human volunteers are invited to participate in the user study to collect data under two modes: one where control is shared between the human operator and the learned agent (*w/ Ours*), and the other where control is directly by the human operator alone (*w/o Ours*). Our proposed system achieves significant improvements in success rate and collection speed by sharing control between human operators and learned agents, as demonstrated in Tab. IV. Additionally, data gathered under our proposed joint learning shared control mode yield performance on the three tasks that are comparable to those pure human datasets using BC and DP, further substantiated by the results presented in Tab. V.

We have developed a questionnaire comprising shown in Tab. VI to capture various dimensions of user experience and

	Pick-and-Place		*Articulated-Manipulation*		*Push-cube*	
	40$\mathcal{H}$	*20$\mathcal{H}$ + 20$\mathcal{S}$*	*30$\mathcal{H}$*	*10$\mathcal{H}$ + 20$\mathcal{S}$*	*20$\mathcal{H}$*	*10$\mathcal{H}$ + 10$\mathcal{S}$*
BC	13 / 20	14 / 20	18 / 20	19 / 20	15 / 20	15 / 20
DP	11 / 20	12 / 20	16 / 20	12 / 20	15 / 20	13 / 20

TABLE V: Real world gripper experiments of data quality.

Satisfaction: $\alpha = 0.769$
1. *It is fun to use.*
2. *It works the way I want it to work.*
3. *It is wonderful.*
4. *It helps me be more effective.*
5. *It is flexible.*

User-Friendly: $\alpha = 0.852$
6. *It is simple to use.*
7. *It is effortless.*
8. *I can use it without written instructions.*
9. *I do not notice any inconsistencies as I use it.*

TABLE VI: Subjective Measures.

ergonomics, and we invited 10 volunteers to rate our system based on their feedback.

This questionnaire assesses ease of use and overall satisfaction. The reliability of our questionnaire is supported by strong Cronbach's alpha values: $\alpha = 0.769$ for the satisfaction section and $\alpha = 0.852$ for the user-friendly section, indicating internal consistency.

V. CONCLUSION

In this paper, we introduce a novel human-agent joint learning paradigm that enables simultaneous human demonstration collection and robot manipulation teaching. This approach allows the human operator to share control with a diffusion-model-based assistive agent within a vision-based teleoperation system to control multiple robot end-effectors such as grippers and dexterous hands. Given our paradigm, the human operator can reduce the effort spent on data collection and adjust the control ratio between the human and agent based on different scenarios. Our system offers a more efficient and flexible solution for data collection and robot manipulation learning via teleoperation.

References

[1] S. P. Arunachalam, I. Güzey, S. Chintala, and L. Pinto, "Holo-dex: Teaching dexterity with immersive mixed reality," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2023, pp. 5962–5969.

[2] Z. Gharaybeh, H. Chizeck, and A. Stewart, *Telerobotic control in virtual reality*. IEEE, 2019.

[3] H. Liu, X. Xie, M. Millar, M. Edmonds, F. Gao, Y. Zhu, V. J. Santos, B. Rothrock, and S.-C. Zhu, "A glove-based system for studying hand-object manipulation via joint pose and force sensing," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2017, pp. 6617–6624.

[4] H. Liu, Z. Zhang, X. Xie, Y. Zhu, Y. Liu, Y. Wang, and S.-C. Zhu, "High-fidelity grasping in virtual reality using a glove-based system," in *2019 international conference on robotics and automation (ICRA)*. IEEE, 2019, pp. 5180–5186.

[5] A. Handa, K. Van Wyk, W. Yang, J. Liang, Y.-W. Chao, Q. Wan, S. Birchfield, N. Ratliff, and D. Fox, "Dexpilot: Vision-based teleoperation of dexterous robotic hand-arm system," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 9164–9170.

[6] Y. Qin, W. Yang, B. Huang, K. Van Wyk, H. Su, X. Wang, Y.-W. Chao, and D. Fox, "Anyteleop: A general vision-based dexterous robot arm-hand teleoperation system," *arXiv preprint arXiv:2307.04577*, 2023.

[7] A. Mandlekar, D. Xu, J. Wong, S. Nasiriany, C. Wang, R. Kulkarni, L. Fei-Fei, S. Savarese, Y. Zhu, and R. Martín-Martín, "What matters in learning from offline human demonstrations for robot manipulation," *arXiv preprint arXiv:2108.03298*, 2021.

[8] C. Chi, S. Feng, Y. Du, Z. Xu, E. Cousineau, B. Burchfiel, and S. Song, "Diffusion policy: Visuomotor policy learning via action diffusion," 2023.

[9] P. Florence, C. Lynch, A. Zeng, O. A. Ramirez, A. Wahid, L. Downs, A. Wong, J. Lee, I. Mordatch, and J. Tompson, "Implicit behavioral cloning," in *Conference on Robot Learning*. PMLR, 2022, pp. 158–168.

[10] J. Lv, W. Xu, L. Yang, S. Qian, C. Mao, and C. Lu, "Handtailor: Towards high-precision monocular 3d hand recovery," *British Machine Vision Conference (BMVC)*, 2021.

[11] Y. Rong, T. Shiratori, and H. Joo, "Frankmocap: A monocular 3d whole-body pose estimation system via regression and integration," in *IEEE International Conference on Computer Vision Workshops*, 2021.

[12] T. Schmidt, R. A. Newcombe, and D. Fox, "Dart: Dense articulated real-time tracking." in *Robotics: Science and systems*, vol. 2, no. 1. Berkeley, CA, 2014, pp. 1–9.

[13] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler, "Analysis of the accuracy and robustness of the leap motion controller," *Sensors*, vol. 13, no. 5, pp. 6380–6393, 2013.

[14] S. Javdani, S. S. Srinivasa, and J. A. Bagnell, "Shared autonomy via hindsight optimization," *Robotics science and systems: online proceedings*, 2015.

[15] S. Reddy, A. D. Dragan, and S. Levine, "Shared autonomy via deep reinforcement learning," *arXiv preprint arXiv:1802.01744*, 2018.

[16] C. Schaff and M. R. Walter, "Residual policy learning for shared autonomy," *arXiv preprint arXiv:2004.05097*, 2020.

[17] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, J. Dabis, C. Finn, K. Gopalakrishnan, K. Hausman, A. Herzog, J. Hsu *et al.*, "Rt-1: Robotics transformer for real-world control at scale," *arXiv preprint arXiv:2212.06817*, 2022.

[18] F. Ebert, Y. Yang, K. Schmeckpeper, B. Bucher, G. Georgakis, K. Daniilidis, C. Finn, and S. Levine, "Bridge data: Boosting generalization of robotic skills with cross-domain datasets," *arXiv preprint arXiv:2109.13396*, 2021.

[19] H.-S. Fang, H. Fang, Z. Tang, J. Liu, J. Wang, H. Zhu, and C. Lu, "Rh20t: A robotic dataset for learning diverse skills in one-shot," *arXiv preprint arXiv:2307.00595*, 2023.

[20] J. Kofman, X. Wu, T. J. Luu, and S. Verma, "Teleoperation of a robot manipulator using a vision-based human-robot interface," *IEEE transactions on industrial electronics*, vol. 52, no. 5, pp. 1206–1219, 2005.

[21] A. Mandlekar, Y. Zhu, A. Garg, J. Booher, M. Spero, A. Tung, J. Gao, J. Emmons, A. Gupta, E. Orbay *et al.*, "Roboturk: A crowdsourcing platform for robotic skill learning through imitation," in *Conference on Robot Learning*. PMLR, 2018, pp. 879–893.

[22] H. Fang, H.-S. Fang, Y. Wang, J. Ren, J. Chen, R. Zhang, W. Wang, and C. Lu, "Airexo: Low-cost exoskeletons for learning whole-arm manipulation in the wild," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2024, pp. 15 031–15 038.

[23] J. I. Lipton, A. J. Fay, and D. Rus, "Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 179–186, 2017.

[24] D. Antotsiou, G. Garcia-Hernando, and T.-K. Kim, "Task-oriented hand motion retargeting for dexterous manipulation imitation," in *Computer Vision–ECCV 2018 Workshops: Munich, Germany, September 8-14, 2018, Proceedings, Part VI 15*. Springer, 2019, pp. 287–301.

[25] S. Li, X. Ma, H. Liang, M. Görner, P. Ruppel, B. Fang, F. Sun, and J. Zhang, "Vision-based teleoperation of shadow dexterous hand using end-to-end deep neural network," in *2019 International Conference on Robotics and Automation (ICRA)*. IEEE, 2019, pp. 416–422.

[26] H. Liu, S. Nasiriany, L. Zhang, Z. Bao, and Y. Zhu, "Robot learning on the job: Human-in-the-loop autonomy and learning during deployment," *arXiv preprint arXiv:2211.08416*, 2022.

[27] H. R. Walke, K. Black, T. Z. Zhao, Q. Vuong, C. Zheng,

P. Hansen-Estruch, A. W. He, V. Myers, M. J. Kim, M. Du *et al.*, "Bridgedata v2: A dataset for robot learning at scale," in *Conference on Robot Learning*. PMLR, 2023, pp. 1723–1736.

[28] H. J. Jeon, D. P. Losey, and D. Sadigh, "Shared autonomy with learned latent actions," *arXiv preprint arXiv:2005.03210*, 2020.

[29] A. D. Dragan and S. S. Srinivasa, "A policy-blending formalism for shared control," *The International Journal of Robotics Research*, vol. 32, no. 7, pp. 790–805, 2013.

[30] S. Javdani, H. Admoni, S. Pellegrinelli, S. S. Srinivasa, and J. A. Bagnell, "Shared autonomy via hindsight optimization for teleoperation and teaming," *The International Journal of Robotics Research*, vol. 37, no. 7, pp. 717–742, 2018.

[31] K. Muelling, A. Venkatraman, J.-S. Valois, J. E. Downey, J. Weiss, S. Javdani, M. Hebert, A. B. Schwartz, J. L. Collinger, and J. A. Bagnell, "Autonomy infused teleoperation with application to brain computer interface controlled manipulation," *Autonomous Robots*, vol. 41, pp. 1401–1422, 2017.

[32] D. Sadigh, S. S. Sastry, S. A. Seshia, and A. Dragan, "Information gathering actions over human internal state," in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2016, pp. 66–73.

[33] C. Mower, J. Moura, and S. Vijayakumar, "Skill-based shared control," in *Robotics: Science and Systems XVII*. The Robotics: Science and Systems Foundation, Jul. 2021, robotics: Science and Systems 2021, R:SS 2021 ; Conference date: 12-07-2021 Through 16-07-2021. [Online]. Available: https://roboticsconference.org/

[34] D. P. Losey, H. J. Jeon, M. Li, K. Srinivasan, A. Mandlekar, A. Garg, J. Bohg, and D. Sadigh, "Learning latent actions to control assistive robots," *Autonomous robots*, vol. 46, no. 1, pp. 115–147, 2022.

[35] X. Chu, Y. Tang, L. H. Kwok, Y. Cai, and K. W. S. Au, "Bootstrapping robotic skill learning with intuitive teleoperation: Initial feasibility study," 2023. [Online]. Available: https://arxiv.org/abs/2311.06543

[36] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.

[37] M. Janner, Y. Du, J. B. Tenenbaum, and S. Levine, "Planning with diffusion for flexible behavior synthesis," 2022.

[38] A. Ajay, Y. Du, A. Gupta, J. Tenenbaum, T. Jaakkola, and P. Agrawal, "Is conditional generative modeling all you need for decision-making?" 2023.

[39] M. Xu, Z. Xu, C. Chi, M. Veloso, and S. Song, "Xskill: Cross embodiment skill discovery," 2023.

[40] C. Luo, "Understanding diffusion models: A unified perspective," 2022.

[41] T. Yoneda, L. Sun, G. Yang, B. C. Stadie, and M. R. Walter, "To the noise and back: Diffusion for shared autonomy," in *Robotics: Science and Systems XIX, Daegu, Republic of Korea, July 10-14, 2023*, 2023.

Supplementary Materials For Human-Agent Joint Learning for Efficient Robot Manipulation Skill Acquisition

Shengcheng Luo[1*], Quanquan Peng[1*], Jun Lv[1*], Kaiwen Hong[2], Katherine Rose Driggs-Campbell[2], Cewu Lu[1], Yong-Lu Li[1]

APPENDIX

I. Implementation details

Here we lay down the details of the data collection, training, and testing process. More technical details are given here to illustrate our method and implementations better.

A. Shadow Hand and Parallel Gripper Teleoperate System

To adapt to Isaac Gym and our vision system, we made certain modifications to the XML file of the Shadow Hand. We followed [**?** **?** **?**], removed the entire arm part, and added six degrees of freedom to the base mount of the Shadow Hand. This allows it to move freely in the virtual environment without depending on a base. Similarly, to obtain the rigid body Jacobian matrices of the five fingertips of the Shadow Hand, we added a massless rigid body to the tips of all five fingers of the Shadow Hand. This facilitates direct inverse kinematics calculations for the entire finger. In inverse kinematics (IK) calculations, we employed the Damped Least Squares (DLS) method [**?** **?**], this approach helps to prevent instability issues when approaching singularity points. Additionally, the DLS method supports real-time applications because it can provide fast and stable solutions, which is particularly crucial for teleoperation systems. Focusing solely on the five fingertips and wrist is regarded as the most balanced approach between computational efficiency and the precision required for complex hand movements in real-time applications. The system operates on a computer with an RTX 4070 graphics card and a monitor.

To mitigate the accumulation of errors, the process involves mapping hand motion from the real world into the virtual environment and then comparing each action with the action from the previous frame to calculate a delta action. The reason for calculating delta action is to identify and apply only the changes in movement from one frame to the next, rather than applying the absolute positions and orientations directly. This approach helps reduce the accumulation of errors that might occur due to discrepancies between the real-world movements and their representation in the simulated environment. By focusing on the changes (delta) rather than absolute values, the system can more accurately replicate the intended movements in the simulator, leading to more precise and consistent control of the shadow hand.

* denotes equal contribution
[1]Shanghai Jiao Tong University, [2]University of Illinois Urbana-Champaign.

B. Baselines

In this section, we provide the implementation details for BC and BC-RNN models. In Behavior Cloning (BC), the objective is to minimize $\mathbb{E}_{(s,a)\sim\mathcal{D}}||\pi_\theta(s)-a||^2$. We use a 3-layer multi-layer perception (MLP) with a ReLU activation function. All layers are fully connected layers with 128 hidden dimensions with a learning rate of $2\cdot10^{-3}$. We also use the AdamW [**?**] to be the optimizer. The training epoch in dexterous tasks Pick-and-Place, Articulated-Manipulation, and Tool-Use is $60, 100, 100$ separately.

As for BC-RNN, we use an LSTM as the backbone network for BC-RNN [**?**], which we find a slight performance improvement compared to the vanilla RNN model. Following [**?**], during the training phase, a state-action sequence $\{(s_i,a_i),\cdots,(s_{i+T-1},a_{i+T-1})\}$ of length T is sampled from the dataset $\mathcal{D}$ and the network will predict the action sequence based on the states as its input. During the inference phase $a_t, h_{t+1} = \pi_\theta(s_t, h_t)$ where h_t, h_{t+1} are the hidden states. Here we set the learning rate to be $2\cdot10^{-3}$, and the training epoch to be 60.

C. Diffusion-Model-Based Assistive Agent

The assistive agent's noise prediction model ε_θ's backbone network is a 4-layer multi-layer perception (MLP) with a Softplus activation function. All layers are fully connected layers with 128 hidden dimensions. Moreover, we set the diffusion steps $K=50$, $\beta_{\min}=10^{-4}, \beta_{\max}=0.1$ in Eq. 2 with Sigmoid scheduling and use Exponential Moving Average (EMA) to stabilize the training. The learning rate of ε_θ is 10^{-3}.

II. Experiment Setups

A. Tasks

Dexterous Hand Pick-and-Place aims at picking an object on the table and placing it into a container. The observation space is 24 dimensions, including the dexterous robot hand state (18-dim), the object's position (3-dim), and the container's position (3-dim). The dexterous robot hand state is the position of each fingertip (15-dim) and the wrist position (3-dim). The action space is 28 dimensions, including the state change of each joint (22-dim) and the wrist transformation (6-dim). The object's position is randomized for each attempt within a $10cm\times10cm$ square on the table.

Dexterous Hand Articulated-Manipulation aims at grasping and unscrewing the door handle to open the door. The observation space is 32 dimensions, including the dexterous

robot hand state (18-dim), the door handle's position (3-dim) and quaternion (4-dim), and the door base's position (3-dim) and quaternion (4-dim). In contrast, the action space is 28 dimensions. The door's position is randomized for each attempt within a $40cm \times 40cm$ square on the floor.

Dexterous Hand Tool-Use aims at picking a hammer and using it to drive a nail into a board. The observation space is 32 dimensions, including the dexterous robot hand state (18-dim), hammer's position (3-dim) quaternion (4-dim), and nail's position (3-dim). At the same time, the action space is 28 dimensions. The nail's position is randomized for each attempt within a $10cm \times 10cm$ square on the table.

Parallel Gripper Pick-and-Place aims at picking an object on the table and placing it into a container. The observation space is 27 dimensions, including the five rigid bodies of the gripper to object distances (15-dim), the distance between left and right grippers (3-dim), the object's position (3-dim), the distance between object and target (3-dim,) and the distance between flange and target (3-dim). The action space is 8 dimensions, including the state change of each joint (7-dim) and gripper (1-dim). The object's position is randomized for each attempt within a $10cm \times 10cm$ square on the table.

Parallel Gripper Articulated-Manipulation aims at picking an object on the table and placing it into a container. The observation space is 16 dimensions, including the five rigid bodies of gripper to object distances (15-dim), and the distance between object and target (1-dim). The action space is 7 dimensions, including the state change of each joint (7-dim). The object's position is randomized for each attempt within a $10cm \times 10cm$ square on the table.

Parallel Gripper Cube-Push aims at pushing an object on the table to the target position. The observation space is 22 dimensions, including the three rigid bodies of the gripper to object distances (9-dim), the flange's position (7-dim), the distance between object and target (3-dim,) and the distance between flange and target (3-dim). The action space is 7 dimensions, including the state change of each joint (7-dim). The object's position is randomized for each attempt within a $5cm \times 5cm$ square on the table.

B. Ablation study

We implement the shared control agent with different methods like the diffusion model and BC. BC adapts a classical way for blending policy to achieve shared control [**?**]. We use it in the ablation study to blend BC policy with pure human action to achieve shared control in Fig.1. Compared to the classical way which explicitly averages human action a^h and agent action a^r to get the shared action a^s, we instead use the diffusion model, which is a popular implicit model, to blend two actions. It models the process as the forward and reverse process. The forward/diffuse process is about adding Gaussian noise to human action a^h, and the reverse process uses a neural network $f(\cdot|\cdot)$ to denoise a^k to get the shared action a^s.

BC agent is trained using a specific sequence of data collection and fine-tuning steps to optimize performance across different levels of shared control. Initially, we collect data sets of 10, 10, and 20 episodes under various task conditions. These initial datasets are used to train a preliminary agent. Following this initial training phase, we employ the trained agent to assist in further data collection under three different control ratios represented by γ values of 0.25, 0.5, and 0.75. The data collected with the assistance of the agent under these γ settings are then used to fine-tune the agent.

TABLE I: Agent performance on human expert or amateur datasets.

Dexterous Hand	*Pick-and-Place*		*Articulated-Manipulation*		*Tool-Use*	
	Skilled	*Unskilled*	*Skilled*	*Unskilled*	*Skilled*	*Unskilled*
BC	0.45	0.02	0.43	0.18	0.40	0.05
BC-RNN	0.41	0.05	0.62	0.04	0.27	0.05
DP	0.71	0.01	0.68	0.10	0.81	0.03

TABLE II: Ablation study on DP performance between r.

	Pick-and-Place	*Articulated-Manipulation*	*Tool-Use*
$r = 0.0$	0.565	0.661	0.512
$r = 0.1$	**0.620**	**0.681**	**0.547**
$r = 0.2$	0.575	0.407	0.115
$r = 0.3$	0.435	0.216	0.029

As shown in Fig.1, experiments demonstrated that the success rate of an assistive agent based on BC is lower than that of an agent based on diffusion models, indicating a reduced capacity for assistance. In certain instances, the action even becomes worse at particular control ratios.

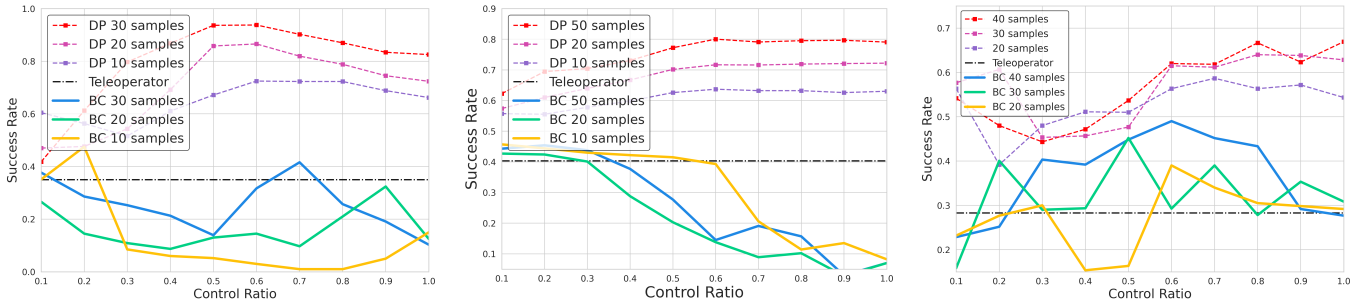
We test the performance of training with different data compositions. For a task, we gathered two manipulation datasets from both skilled and unskilled human operators. We consider operators to be skilled workers if they can practice for more than five hours and reach a success rate and efficiency comparable to those with assistive agents. As shown in Tab. I, the performance of agents trained on the dataset of unskilled operators is much lower than that on the dataset of skilled operators. Therefore, all the human operation datasets $\mathcal{H}$ we use in the main text are from skilled operators.

In our framework, r represents the modification ratio of noise between the state and action. Specifically, during the training, the noise added to state s satisfies $\varepsilon_s = u \cdot \mathcal{N}(\mathbf{0}, \mathbf{I})$ while the noise added to action a satisfies $\varepsilon_s = v \cdot \mathcal{N}(\mathbf{0}, \mathbf{I})$. Then $r = \frac{u}{v}$. We test different r as shown in Tab. II, to ensure the best agent performance. We default to using $r = 0.1$ in our model.

C. Real World Experiment

In this section, we evaluate the real-world performance of our method. We use the setup shown in Fig.2, which includes a Flexiv Rizon4 arm equipped with a gripper and two Intel RealSense D435i RGB-D cameras. One camera is mounted on the wrist of the robotic arm, while the second is positioned on the side. One task here is to pick the red pot shown in Fig.2 and place it onto the induction cooker.

During the real-world data collection phase, we estimate the human hand's pose using RGBD input. Considering the significant difference in morphology between the human hand



(a) *Pick-and-Place* (b) *Articulated-Manipulation* (c) *Tool-Use*

Fig. 1: Ablation on different dexterous agents trained with different compositions of data.

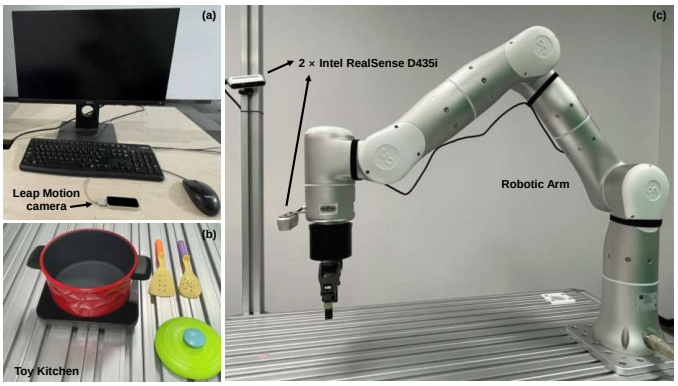


Fig. 2: Realworld Pick-and-Place Experiment. The hardware setup comprises (a) a Leap Motion camera utilized for teleoperation data collection, (b) a toy kitchen environment set up for the pick-and-place task, and (c) a Flexiv Rizon4 robotic arm equipped with a gripper and two cameras. One camera is mounted on the wrist of the robotic arm, while the second one is positioned on the side.

and a 7-DoF robotic arm, we chose to track the end effector's position by monitoring the position of the hand's wrist. Additionally, we used the action of closing or opening the human hand as the condition for determining whether to grasp or release an object. This approach leverages the greater dexterity of the human hand to enhance the control and precision of the robotic arm. We record RGB images from two camera views, joint poses (7-dim), gripper width (1-dim), the end effector's position (3-dim), and its quaternion (4-dim). The RGB images have a size of 640×480 pixels, each episode is sampled at a frequency of 10 Hz.

In real-world experiments, the network architecture is generally similar to the simulation environment's. Our input has changed from the original hand states and object states to the position and orientation of the robot arm end effector, as well as images from the first-person and third-person perspectives. We made two main modifications: 1) For the images, we used a ResNet-18 model. We used a standard ResNet-18 (without pretraining) as the encoder with its global average pooling replaced with a spatial softmax pooling to maintain spatial information. 2) We deepened the layer of the neural network, increased its hidden layer dimension, and expanded action horizon prediction from predicting the next frame action to predicting actions for the subsequent T frames, *i.e.*, $a_{t+1:t+T-1}$ (where $T = 8$).

III. Discussion and Limitation

A. Human-Machine Interface

Our approach has demonstrated success across a diverse set of Human Machine Interfaces(HMI), including:

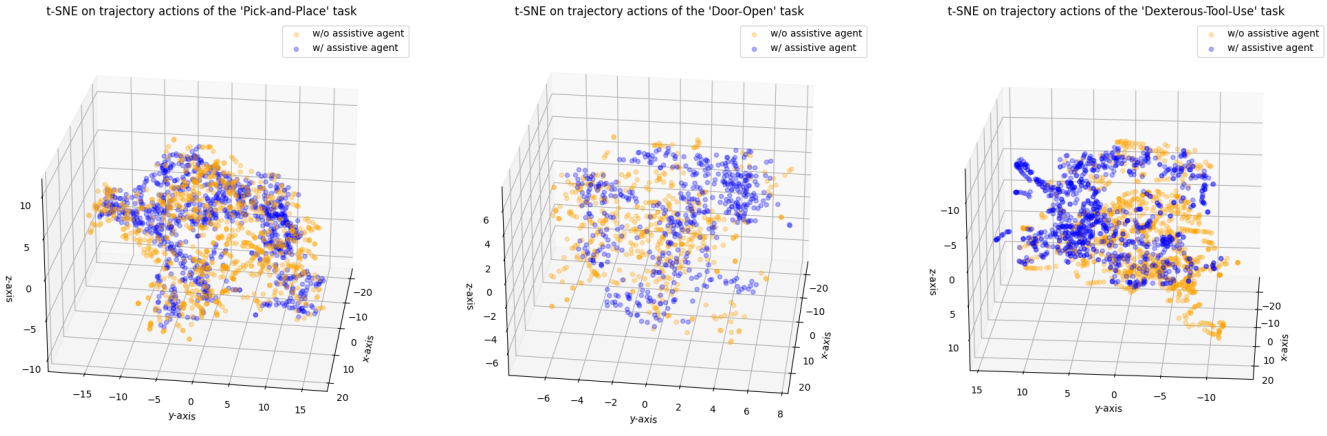
Sigma.7 Teleoperation Devices: Our system has successfully utilized Sigma devices to achieve precise control for tasks involving limited DoF. These devices require intricate control and feedback mechanisms, demonstrating our interface's robustness and effectiveness in physical UI scenarios.

RGB-D Cameras: Our system can accurately interpret spatial environments by leveraging depth perception, making it highly effective for freehand teleoperation. This capability lays the foundation for handling physical UIs with equal precision.

Virtual Reality (Meta Quest3): In VR environments, our interface provides an immersive and intuitive experience that closely mimics real-world interactions. This shows its capability to handle complex interfaces with precision and ease. As shown in Tab. III, we repeated the dexterous articulated-manipulation experiment with Leap Hand [**?**] in a VR environment and validated that our paradigm is applicable across different HMIs. This demonstrates the versatility of our approach, ensuring consistent operation across various human-machine interfaces.

TABLE III: Articulated-Manipulation task success rate under increasing data with Quest3.

VR Dexterous	*Articulated-Manipulation* *BC*	*DP*
$10\mathcal{H}$	0.04	0.10
$10\mathcal{H}$ + $10\mathcal{H}$	0.15	0.25
$10\mathcal{H}$ + $20\mathcal{H}$	0.26	0.26
$10\mathcal{H}$ + $30\mathcal{H}$	0.40	0.30
$10\mathcal{H}$ + $10\mathcal{S}$	0.34	0.28
$10\mathcal{H}$ + $20\mathcal{S}$	0.30	0.35
$10\mathcal{H}$ + $30\mathcal{S}$	0.44	0.63



(a) *Pick-and-Place* (b) *Articulated-Manipulation* (c) *Tool-Use*

Fig. 3: t-SNE visualization of distributions of trajectory actions (w/ and w/o assistive agent).

B. Data Analysis

We visualize the Preference Alignment [**?**] for dexterous hand articulated-manipulation task, as shown in Fig. 4. We find that as time progresses, the preference alignment increases across all three phases: reaching the door latch, twisting the door latch, and pulling. This indicates a growing synchronization between the user and the assistive agent throughout each stage of the task. Also, the preference alignment between the user and the assistive agent improves across different control ratios.

We use t-SNE to visualize the distribution of trajectory actions on different dexterous tasks, as shown in Fig. 3. Specifically, we have reduced the trajectory of actions to three dimensions using t-SNE, for both data collected by human operators with and without our system. To ensure a fair comparison, we uniformly sampled the same number of actions across both scenarios. We find that the distribution of the same task tends to cluster in the same space, whether with or without an assistive agent. This indirectly demonstrates that our system can enhance data collection speed and efficiency without compromising data quality.

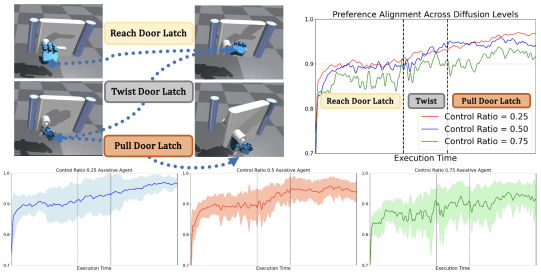


Fig. 4: The articulated-manipulation task consists of three phases: reaching the door latch, twisting it to the correct angle, and pulling it. We plotted the dot product between the user input action and the assistive agent's output action(Preference Alignment). In the plot, the red, green, and blue lines represent control ratios of 0.25, 0.5, and 0.75, respectively.

C. Limitations

Our current system's task-specific assistive agent, while effective for certain applications, does have its limitations. It currently can not handle tasks that involve multiple subtasks or targets that change dynamically, as these scenarios often require more flexibility, including the ability to adjust the control ratio throughout the sequence. To broaden the system's applicability, integrating large language models could also allow it to handle a wider range of robot learning tasks by conditioning on text input. Additionally, we think adding a learnable control ratio adjustment mechanism, especially for long-horizon tasks, could improve the system's adaptability and efficiency. We believe that our proposed joint-learning framework has the potential to leverage more powerful multi-task diffusion policies, allowing it to handle more complex scenarios in future enhancements.

ACKNOWLEDGMENT

The authors would like to thank anonymous reviewers for helping improve exposition. This work is supported in part by the National Natural Science Foundation of China under Grants 62306175.