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]
Shanghai Jiao Tong University[1] University of Illinois Urbana-Champaign[2]

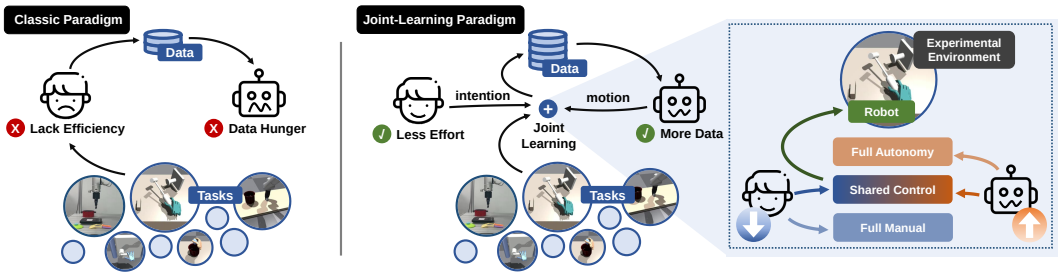


Figure 1: Traditional frameworks typically separate human and agent training, requiring operators first to learn the task environment before data collection. This often leads to inefficiencies due to delayed and insufficient data gathering. In our framework, we integrate human and agent training from 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. 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.

Keywords: Human-Robot Collaboration, Teleoperation, Robotic Manipulation

1 Introduction

Inspired by shared autonomy [1, 2, 3], 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 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.

8th Conference on Robot Learning (CoRL 2024), Munich, Germany.

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 simulation environments using two end-effectors: a dexterous hand and a gripper, as well as three real-world tasks. Results show our system significantly enhances data collection efficiency, increasing success rate by 30% and nearly doubling collection speed. Additionally, data collected in shared autonomy mode is as effective for downstream tasks as data collected directly from teleoperation. Our main contributions are summarized as follows:

- Proposing a human-agent joint learning paradigm to reduce human adaptation while maintaining data quality in teleoperation data collection.
- Developing a system that enables concurrent development of the human operator and assistive agent, streamlining learning and expediting autonomous robot manipulation.
- 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.

2 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. Second, using these data, we train a diffusion-model-based assistive agent (Sec. 2.1) to establish shared control between the human operator and the agent, thereby improving the efficiency of the data collection process (Sec. 2.2).

2.1 Diffusion-Model-Based Assistive Agent.

After collecting data via teleoperation, 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 Gaussian distribution to predict the desired action

$$a = f(a^k|s, k). \tag{1}$$

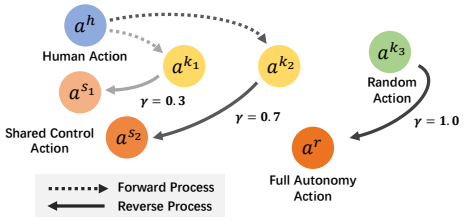


Figure 2: **Diffusion based shared control.** To achieve shared control between the human and agent, we blend the action from the human operator a^h using the forward and reverse process. The parameter γ governs the control ratio, where a lower γ results in the action better aligning with the human operator's intention. In contrast, a higher γ allows the learned agent to exert more influence over the blended action.

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 [4]: $a^s = \gamma a^h + (1-\gamma)a^r$, where a^r is generated by the learned agent. However, considering that the agent operates as a diffusion policy (Fig. 2), 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: $a^k = a^h + \epsilon^k$. Subsequently, a reverse process denoises the action a^k:

$$a^s = f(a^k|s, k). \tag{2}$$

Applying action a enables shared control between the human and the assistive agent. The control ratio $\gamma = k/K$ adjusts this sharing by varying k. When $\gamma = 0$, $a^s = a^h$, and the robot is entirely human-controlled. As γ approaches 1, a^s becomes the autonomous action a^r. Thus, a higher γ grants the agent more control to stabilize and direct the dexterous hand.

2.2 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.

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. 2 (shown in Fig. 2).

Algorithm 1 Overall Process

Require: The human operator $\mathcal{H}$;
Ensure: The collected dataset $\mathcal{D}$; assistive agent f; control ratio γ;
Initialization: $\mathcal{D} \leftarrow \emptyset, \gamma \leftarrow 0$;
while $|\mathcal{D}|$ is small **do** ▷ not enough data is collected
  $\mathcal{H}$ collects data d under f's help; ▷ see Sec. 2.2 for control ratio adjustment
  if d is valid **then**
    $\mathcal{D} \leftarrow \{d\} \cup \mathcal{D}$;
  end if
  Finetune f with $\mathcal{D}$;
end while
return $\mathcal{D}$ and f;

3 Experiments

Tasks. We consider six multi-stage manipulation tasks (Fig. 3): *Pick-and-Place* involves picking an object and placing it into a container. In *Articulated-Manipulation*, the dexterous hand unscrews a door handle; the gripper pulls open a drawer. *Push-cube* requires pushing a cube to a target position. *Tool-Use* involves picking up a hammer to drive a nail into a board.

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. Visualizations in Fig. 4 show that with the assistive agent, human operators need only provide high-level intentions, like movement direction or grasp motions, rather than precise control. In multi-stage tasks, such as using a hammer, operators merely provide *trigger actions* to guide transitions between sub-stages, reducing effort and speeding up data collection.

To evaluate the agent's ability to correct imperfect human control, we simulate human input using a baseline Behavior Cloning (BC) agent as a proxy. As shown in Fig. 5, with limited data, the assistive agent effectively aids the simulated operator. As more data is collected and the agent is further trained, its ability to correct actions improves, reducing operator effort and enhancing data collection efficiency.

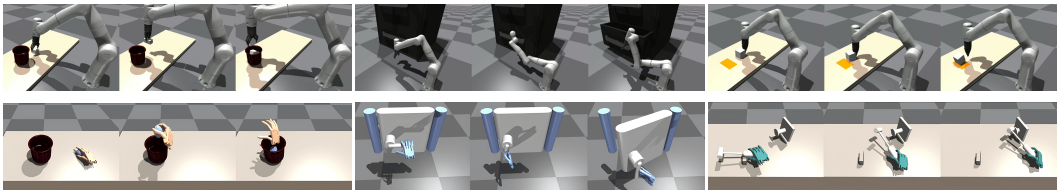


Figure 3: **Simulation tasks overview.** Here are six task settings and their task flow for Pick-and-Place *(left)*, Articulated-Manipulation *(middle)*, Gripper-Push *(upper-right)* and Dexterous-Tool-Use *(bottom-right)*.

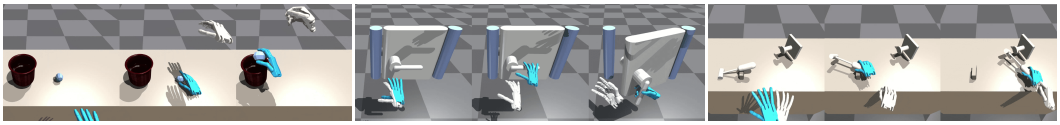
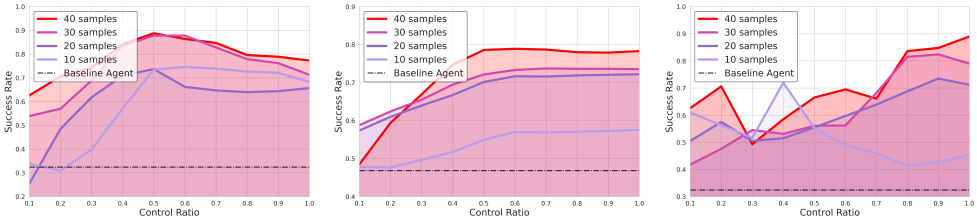


Figure 4: **Shared control process overview.** The white one is the hand controlled purely by the *human operator*, while the cyan one is under *shared control* between the human and the assistive agent.



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

Figure 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 ($\gamma = 1$).

Once sufficient data is collected, the assistive agent can transition to full autonomy by setting $\gamma = 1$ and generating actions from Gaussian noise. We achieve success rates of 0.76, 0.78, and 0.89 across three dexterous manipulation tasks, indicating the agent can effectively operate autonomously.

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 1: 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 2: Tool-Use and Push-cube task success rate under increasing data.

Data Quality on Downstream Task. We show that shared control does not compromise data quality. We collect dexterous hand and gripper demonstrations in two modes: human-only control ($\mathcal{H}$) and shared control ($\mathcal{S}$) between the operator and assistive agent. These datasets are used to train agents like BC, BC-RNN [5], and Diffusion Policy (DP) [6].

In Tab. 1, data collected through shared control between the human and the assistive agent shows comparable or even better performance with BC and BC-RNN compared to expert human demonstrations. Their results are comparable with DP, possibly as DP can better fit the tasks, which is in line with [6].

In Tab. 2, adding more shared control data significantly improves policy learning, yielding similar performance to human-expert data on tool-use and push-cube tasks at a lower cost. This confirms the quality and efficiency of data collected under our shared control approach for downstream tasks.

4 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.

Acknowledgments

If a paper is accepted, the final camera-ready version will (and probably should) include acknowledgments. All acknowledgments go at the end of the paper, including thanks to reviewers who gave useful comments, to colleagues who contributed to the ideas, and to funding agencies and corporate sponsors that provided financial support.

References

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

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

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

[4] A. D. Dragan and S. S. Srinivasa. A policy-blending formalism for shared control. *The International Journal of Robotics Research*, 32(7):790–805, 2013.

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

[10] 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*, 52(5):1206–1219, 2005.

[11] 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*, pages 879–893. PMLR, 2018.

[12] 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)*, pages 15031–15038. IEEE, 2024.

[13] 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)*, pages 5962–5969. IEEE, 2023.

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

[15] 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)*, pages 6617–6624. IEEE, 2017.

[16] 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)*, pages 5180–5186. IEEE, 2019.

[17] J. I. Lipton, A. J. Fay, and D. Rus. Baxter's homunculus: Virtual reality spaces for teleoperation in manufacturing. *IEEE Robotics and Automation Letters*, 3(1):179–186, 2017.

[18] 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)*, pages 9164–9170. IEEE, 2020.

[19] 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.

[20] 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*, pages 287–301. Springer, 2019.

[21] 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)*, pages 416–422. IEEE, 2019.

[22] 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.

[23] 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*, pages 1723–1736. PMLR, 2023.

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

[25] 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*, 37(7):717–742, 2018.

[26] 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*, 41:1401–1422, 2017.

[27] 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)*, pages 66–73. IEEE, 2016.

[28] C. Mower, J. Moura, and S. Vijayakumar. Skill-based shared control. In *Robotics: Science and Systems XVII*. The Robotics: Science and Systems Foundation, July 2021. doi:10.15607/RSS.2021.XVII.028. URL `https://roboticsconference.org/`. Robotics: Science and Systems 2021, R:SS 2021 ; Conference date: 12-07-2021 Through 16-07-2021.

[29] 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*, 46(1):115–147, 2022.

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

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

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

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

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

[35] 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.

[36] K. Shaw, A. Agarwal, and D. Pathak. Leap hand: Low-cost, efficient, and anthropomorphic hand for robot learning. *arXiv preprint arXiv:2309.06440*, 2023.

Appendix

In this appendix, we provide additional details to support our main paper. We briefly discuss related work in teleoperation for data collection and interactive robot learning in Sec. A. We then present background on the Denoising Diffusion Probabilistic Model (DDPM) in Sec. B. Next, we explain our algorithm's steps for training the assistive agent in Sec. C and efficiency of data collection in Sec. D. Additionally, in Sec. E, we demonstrated the applicability of our method across various human-machine interfaces (HMIs). Moreover, we conducted ablation studies in Sec. F and data analysis in Sec. G to further validate the effectiveness and robustness of our approach. Finally, we describe our real-world experiments and summarize user feedback from our evaluations in Sec. H.

A 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 [5, 7, 8, 9, 10, 11, 12]. Some works achieve teleoperation through wearable devices [13, 14, 15, 16, 17], and vision-based teleoperation systems offer a low-cost and easily developed alternative [18, 19, 20, 21]. For instance, [21] utilizes neural networks for markerless vision-based teleoperation of dexterous robotic hands from depth images. [18] set up a vision-based teleoperation system to control the Allegro Hand, accomplishing various contact-rich manipulation tasks in the real world. Recently, [19] 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 [22, 23]. Previous work uses shared autonomy to assist people with disability in performing tasks by arbitrating human inputs and robot actions [24]. Many of the shared autonomy algorithms aim to estimate human intents from a set of pre-defined goals [4, 25, 26, 27], using clothoid curves to parametrize the state and control [28] or by mapping low-dimension control input to high-dimension robot actions [24, 29]. In this work, we introduce a system that integrates the agent's learning process with data collection, facilitating both data collection and robot learning.

B Preliminary

To get a learned agent in Sec. 2.1, enabling human-agent joint learning, we follow the Denoising Diffusion Probabilistic Model (DDPM) [30] 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

$$x_k = \sqrt{\alpha_k} x_{k-1} + \sqrt{1-\alpha_k}\epsilon, \tag{3}$$

where $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \alpha_k = 1 - \beta_k$. DDPM models the output generation as a denoising process (Stochastic Langevin Dynamics). A line of works [31, 32, 6, 33] 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

$$p_\theta(x^0) = \int p(x^K) \prod_{k=1}^{K} p_\theta(x^{k-1}|x^k) \mathrm{d}x^{1:K}, \tag{4}$$

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*. [34] shows that $p_\theta(x^{t-1}|x^k)$ becomes tractable when conditioned on x_0 and Eq. 4 can be reformulated as minimizing the error in the noise prediction. [30] simplify the training loss function as

$$\mathcal{L} := \mathbb{E}_{k, \boldsymbol{x}_0, \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})} \left[\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\boldsymbol{x}_k(\boldsymbol{x}_0, \boldsymbol{\epsilon}), k) \|_2^2 \right], \tag{5}$$

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

$$x_{k-1} = \mu_\theta(x_k, k) + \sigma_k \boldsymbol{z}. \tag{6}$$

Similar to [6, 35], 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. 5 shall be modified as

$$\mathcal{L} := \mathbb{E}_{k,(s_i,a_i),\boldsymbol{\epsilon}\sim\mathcal{N}(\mathbf{0},\boldsymbol{I})}\left[\|\boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(a_i + \boldsymbol{\epsilon}, s_i, k)\|_2^2\right]. \tag{7}$$

C 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. 7 and Eq. 1, 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.

D Efficiency of Data Collection.

We conducted a user study to evaluate how our shared control system enhances data collection efficiency. Ten participants collected data under two modes: shared control with the learned agent (*w/ Ours*) and direct control by the operator alone (*w/o Ours*). They had three minutes per mode for three dexterous hand tasks, aiming to collect as much data as possible. We measured *Success Rate* (Percent), *Horizon Length* (Steps per Sample), and *Collection Speed* (Samples per Hour).

As shown in Tab. 3, our system improved success rate and collection speed while reducing horizon length, indicating enhanced efficiency. To ensure fairness, participants were split into two groups with reversed mode orders. *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.

		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 3: User studies on three dexterous hand tasks.

E 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. 4, we repeated the dexterous articulated-manipulation experiment with Leap Hand [36] 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 4: Articulated-Manipulation task success rate under increasing data with Quest3.

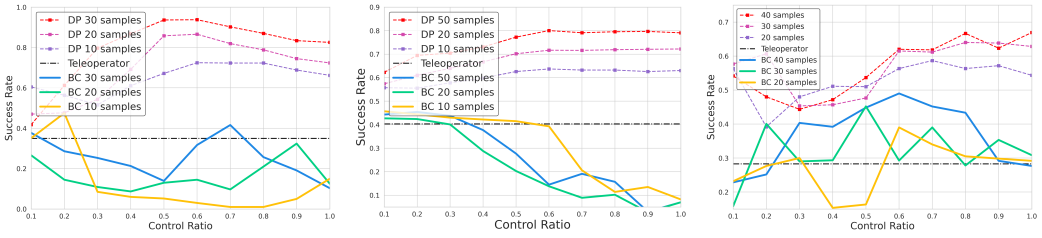
VR Dexterous	*Articulated-Manipulation* *BC*	*DP*
$10\mathcal{H}$	0.04	0.10
$10\mathcal{H}$ + $10\mathcal{H}$	0.15	0.25
$10\mathcal{H}$ + $20\mathcal{H}$	0.26	0.26
$10\mathcal{H}$ + $30\mathcal{H}$	0.40	0.30
$10\mathcal{H}$ + $10\mathcal{S}$	0.34	0.28
$10\mathcal{H}$ + $20\mathcal{S}$	0.30	0.35
$10\mathcal{H}$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

F 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 [4]. We use it in the ablation study to blend BC policy with pure human action to achieve shared control in Fig.6. 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.

As shown in Fig.6, 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.



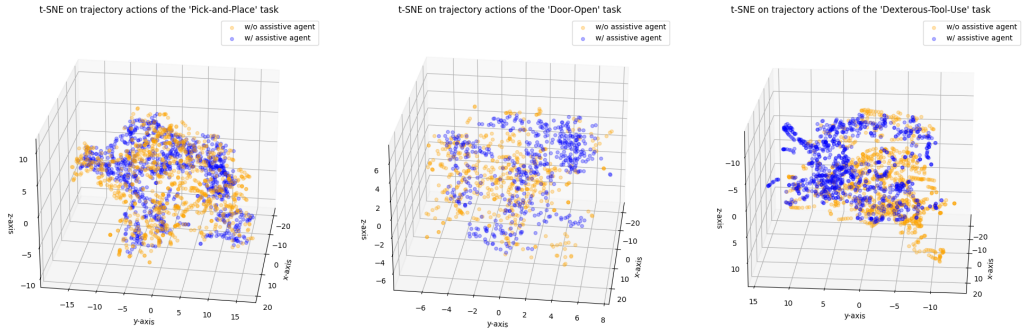
(a) *Pick-and-Place* (b) *Articulated-Manipulation* (c) *Tool-Use*

Figure 6: Ablation on different dexterous agents trained with different compositions of data.

G Data Analysis

We use t-SNE to visualize the distribution of trajectory actions on different dexterous tasks, as shown in Fig. 7. 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.



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

Figure 7: t-SNE visualization of distributions of trajectory actions (w/ and w/o assistive agent).

H 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. 8. Following the same rules as Sec. D, 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. 5. 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. 6.

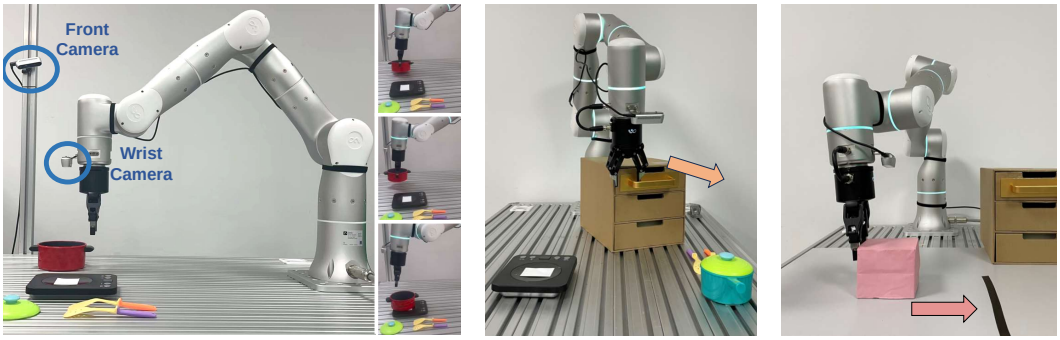


Figure 8: **Real world setting.** 1. *Pick-and-Place*: use the gripper to pick the red pot up and place it onto the black induction cooker. 2. *Articulated-Manipulation*: use the gripper to open the drawer. 3. *Push-cube*: use the gripper to push the cube across the black line.

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

Table 5: Real world parallel gripper Pick-and-Place task user study.

	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 6: Real world parallel gripper experiments of data quality.

We have developed a questionnaire comprising shown in Tab. 7 to capture various dimensions of user experience and ergonomics, and we invited 10 volunteers to rate our system based on their feedback.

***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 7: Subjective Measures

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.