CHG-DAgger: Interactive Imitation Learning with Human-Policy Cooperative Control

Taro Takahashi
Toyota Motor Corporation
taro_takahashi@mail.toyota.co.jp

Yutaro Ishida
Toyota Motor Corporation
yutaro_ishida@mail.toyota.co.jp

Takayuki Kanai
Toyota Motor Corporation
texttttakayuki_kanai@mail.toyota.co.jp

Naveen Kuppuswamy
Toyota Research Institute
naveen.kuppuswamy@tri.global

Abstract: This paper presents a novel approach to improve the usability of Interactive Imitation Learning (IIL) for end-to-end visuomotor control policies. The proposed framework, Cooperative-HG-DAgger (CHG-DAgger), allows the expert human and learned policy to collaborate in continuing the task upon task failure without switching control between the policy and human. As a result, human intervention time is reduced because the human can correct the motion while being guided by the policy, and they can understand when corrections are no longer needed through physical interaction. To achieve cooperative control, we adopted multilateral control, an extension of bilateral control, designed to avoid instability even with low-cost hardware and long reference trajectory update cycles. Ensuring seamless integration with the Diffusion Policy, recent advances in the visuomotor imitation learning method, the proposed method achieved high success rates through retraining by leveraging the recovery motion data. Additionally, it was shown that intervention time can be reduced in minor adjustment where human operation is close to the policy, when prior knowledge of the policy is limited. Our results indicate that the proposed method offers a more intuitive and efficient way of handling task failures, paving the way for continuous learning and robust robot autonomy.

Keywords: Interactive Imitation Learning, Cooperative Control

1 Introduction

Recently, there has been a significant rise in research on end-to-end imitation learning for visuomotor control policies. In this field, performing tasks when there are Out of Distribution (OOD), such as variations in the objects being manipulated or lighting conditions, presents a challenging problem. As a result, much research has focused on constructing large-scale datasets to train [1, 2], efficient data collection methods [3, 4], and generalization strategies [5, 6] to address these challenges. Despite various efforts, predicting all potential OOD cases, including failures, and gathering all demonstrations before rollout is still infeasible. Consequently, the learned policy has suffered from OOD until today.

One promising approach to addressing these issues is Interactive Imitation Learning (IIL) [7, 8]. IIL is a branch of the imitation learning where human feedback is provided intermittently during robot execution (i.e., *after* rollout), enabling the robot's behavior to be improved online or retrained offline. This IIL can be categorized into two types. The first is robot gated IIL [9, 10], where the robot autonomously determines when human intervention is needed. This approach offers the advantage of high autonomy, reducing the burden on human operators. The second type is human gated IIL [11], where humans intervene when they judge intervention is necessary. This method is advantageous in its ability to handle complex, advanced tasks and unforeseen situations.

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

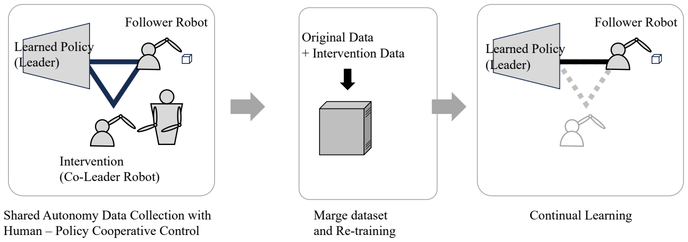


Figure 1: Overview of Proposed CHG-DAgger: In Interactive Imitation Learning (IIL), the policy is retrained with data corrected by a human expert. To enhance the usability for IIL for end-to-end robot motion generation, the proposed method allows cooperative control between the human and the learned policy. Humans can correct motion through co-leader robot under the policy's guidance and notice when it is no longer necessary through physical interaction.

A representative method of human-gated IIL is HG-DAgger [11] and its application to large-scale datasets, BC-Z [12]. In these methods, when the human judges that intervention is necessary, the control by policy is switched to the control by humans to correct robot motion. Once the human judges that the corrections are complete, the control by the humans is switched back to the control by the policy. However, there are two limitations related to this control switching. The first limitation is full reliance on human control during the intervention: when the human is correcting the robot motions, even if the learned policy can execute the task roughly. In other words, the learned policy is unable to assist while the human is performing the task accurately. The second limitation is that it is difficult for humans to decide "when to switch back to learned policy control" after correcting the motion. To decide this, humans must be confident that the learned policy will function effectively after the switch. However, this is difficult, so as a result, the human may end up controlling the robot for longer than necessary. In other words, humans are unable to determine the optimal timing for switching back to the policy while evaluating the learned policy's task-execution capabilities.

In order to avoid these limitations, we propose a physical cooperative control system between humans and policies for human intervention. This system would allow the human to correct actions more easily by sensing the policy's intent through physical interaction, while deciding when to end the intervention based on the policy's behavior. We call this framework Cooperative-HG-DAgger (CHG-DAgger). Figure 1 shows an overview of the system. To implement the physical cooperative control between humans and the policy, we utilized multilateral control [13], which allows multiple users to cooperate in remote control, adapting it to our proposed imitation learning setup. Furthermore, our design of the multilateral control system offers a more straightforward solution that can maintain stability even in low-cost robot hardware setups and with long command update cycles typical of visuomotor control policies.

The purpose of this study is to demonstrate the feasibility of IIL with human-policy cooperative control and to empirically study its user friendlyness in real-world tasks. We set up three real-world tasks for evaluation and confirmed that the success rate, which dropped below 16.6% due to OOD, recovered to more than 77.8% after re-training with the proposed method. The usability of IIL is difficult to evaluate because it depends on each user's sensitivity. Therefore, in this study, we conducted experiments on 10 subjects using ”intervention time,” which is the time required for a human to correct the robot's motion after a task failure, as an evaluation index. As a result, it was confirmed that while intervention time increased for ”major correction” of motion where human manipulation and policy control diverged significantly, intervention time was reduced in 80% of subjects for ”minor adjustment” of motion where human actions and learned policy control were closely aligned. Furthermore, among them, the proposed method reduced intervention time by about 8.5% on average in scenarios with no prior knowledge of the learned policy.

The contributions of this paper are as follows:

1. We introduce Cooperative-HG-DAgger, a novel HG-DAgger strategy that facilitates IIL while reducing the supervision data collection load.

2. We propose stable multilateral control on low-cost hardware to enable a human-policy cooperative control framework, ensuring seamless integration with recent advances in visuomotor imitation learning methods, such as Diffusion Policy [14].
3. We experimentally demonstrate that the multilateral control framework for IIL reduces intervention time.

The structure of this paper is as follows: Chapter 2 discusses related work, Chapter 3 presents the proposed method, Chapter 4 details the experimental setup, Chapter 5 discusses the experimental results, and Chapter 6 provides conclusion and limitations.

2 RELATED WORK

Imitation Learning. In end-to-end imitation learning for visuomotor control policies, executing tasks under different distributions from the training data, such as in OOD scenarios, remains a challenging issue. To address this issue, research has been conducted on efficient demonstrations, data collection using simulators, and generalization of imitation learning, all aimed at collecting data that covers changes in objects, lighting, and other environmental factors. However, it is difficult to collect recovery actions by assuming all possible failure scenarios. In our study, we use IIL incorporated with multilateral control to address this issue. For end-to-end imitation learning of visuomotor control policies, cutting edge models such as Action Chunking with Transformers (ACT) [15] and Diffusion Policy (DP) [14] have been proposed. We integrated DP [14] to our system without any modifications. DP is capable of multimodal representation and can train multiple behaviors with the same objective. This feature might allows for the representation of various recovery motions depending on the person.

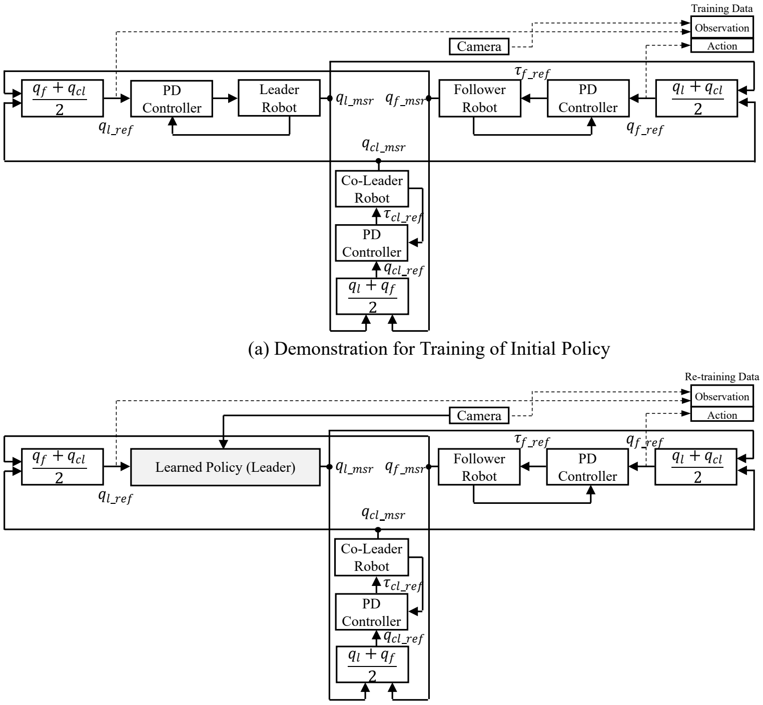
Bilateral Control. In visuomotor control policy imitation learning, training data is often collected using VR-based controllers or unilateral devices for expert demonstrations. For dexterous tasks, bilateral control may also be used, which allows the human to feel the manipulation force while operating. On the other hand, to enable cooperative operations by multiple human operators, multilateral control [13] has been proposed as an extension of bilateral control. However, this method assumes the use of high-performance hardware and short, jitter-free control cycles, such as those found in industrial manipulators. The low-cost, low-performance hardware often used in imitation learning research typically has lower-resolution angle encoders, imprecise measurements or estimations of contact forces at the end-effector and joint torques, and lower frame and joint stiffness, along with backlash in the reduction gears. Additionally, command values derived from visuomotor control policies, such as DP, are updated at significantly longer intervals than the control cycles of the robot's feedback control, which can lead to instability. Therefore, we adapt the multilateral control and design to address these challenges.

3 PROPOSED METHOD

We propose the following control algorithm to correct the actions through physical cooperative control between humans and the learned policy. Additionally, we will discuss several related considerations.

3.1 Control Algorithm

First, in conventional imitation learning using bilateral control [16], training data is collected through remote control between a leader robot and a follower robot, and during rollout with inference, the leader robot is replaced by a learned policy. The proposed method introduces a "co-leader robot" to enable physical cooperative control between human and policy. During rollout based on inference, the co-leader robot can performs interventions with human at any time if necessary (Figure 2(b)). The training data for the initial model is collected through multilateral control [13] (Figure 2(a)) with a leader, co-leader, and follower robot. It should be noted that the training data for the initial



(a) Demonstration for Training of Initial Policy

(b) Rollout and Intervention for Re-training

Figure 2: Control System Block Diagram. To implement the physical cooperative control between humans and the policy, the multilateral control [13] was adaped to our IIL system. Furthermore, our design offers a more straightforward solution that can maintain stability even in low-cost robot hardware setups and with long command update cycles typical of visuomotor control policies.

model can also be obtained using bilateral control, which requires fewer robots, but the proposed method results in fewer differences in the control systems, thus reducing OOD caused by the control systems.

To prevent instability due to low-cost hardware and long command update cycles due to inference time, two simplifications were made to the multilateral control system:

- removing force coupling from parallel structure of position coupling and force coupling
- removing minor loop of acceleration control with disturbance observer

The block diagram of the controller is shown in Figure 2(b). Based on these, the proposed multi-material control is expressed by the following equations (1) to (6).

$$q_{l\_ref} = \frac{q_{f\_msr} + q_{l\_msr}}{2} \tag{1}$$

$$q_{cl\_ref} = \frac{q_{l\_msr} + q_{f\_msr}}{2} \tag{2}$$

$$q_{f\_ref} = \frac{q_{l\_msr} + q_{cl\_msr}}{2} \tag{3}$$

$$\tau_{l\_ref} = (q_{l\_ref} - q_{l\_msr}) \times K_P + (\dot{q}_{l\_ref} - \dot{q}_{l\_msr}) \times K_D \tag{4}$$

$$\tau_{cl\_ref} = (q_{cl\_ref} - q_{cl\_msr}) \times K_P + (\dot{q}_{cl\_ref} - \dot{q}_{cl\_msr}) \times K_D \tag{5}$$

$$\tau_{f\_ref} = (q_{f\_ref} - q_{f\_msr}) \times K_P + (\dot{q}_{f\_ref} - \dot{q}_{f\_msr}) \times K_D \tag{6}$$

where q is the joint angle, τ is the joint torque, and the subscripts l, cl, and f represent the leader, co-leader, and follower robots, respectively. ref is the reference value, msr is the measured value, K_P is the proportional gain, and K_D is the differential gain.

Generally, force feedback control and bilateral control with force coupling are prone to be more unstable than position control and position coupling. In addition, with the low-cost hardware used in this study, the disturbance observer exhibited unstable behavior at cutoff frequencies larger than $30rad/sec$, while $1100rad/sec$ was set in previous studies [17, 18]. This means that the disturbance observer could only perform at lower bandwidths in the low-cost hardware, which did not contribute much to improved performance in multilateral control. However, even with these modified designs, the force could be felt due to the position coupling inherent in multilateral control, and the task performed in this study could be performed.

3.2 Learning Algorithm

As a nominal behavior cloning policy, we adopt DP [14] without any modifications as one of the contenders for state-of-the-art performance. To connect the periodically inferred action space smoothly, methods similar to the weighted averaging mentioned in ACT [15] and the countermeasures for delay in UMI [3] were adopted. Here, the weights and delay times were determined through an experimental adjustment. The observation space (input) and Action space (output) related to multilateral control were set as follows. First, in conventional bilateral control-based imitation learning [16], the behavior of the leader robot, including the human expert's actions, is modeled. The observation space is the command value for the leader (i.e., the follower's measured value), and the output is the next moment's measured value of the leader (i.e., the follower's command value). However, if the same values are used in multilateral control, the human's intervention operation will not be reflected. Therefore, in proposed method, in addition to the leader robot, the co-leader robot is also considered as the values on the control block diagram shown in Figure 2.

3.3 Data Collection and Shared Autonomy Workflow

The initial policy learning is conducted with data collected using remote operations using multilateral control by a human expert using the leader robot. During this process, the co-leader robot works but does not make contact with any person or object. Using this initial policy, tasks can be executed autonomously. If the task fails, or if task failure is predicted and human intervention is deemed necessary, a human intervenes to continue the task and collect data through cooperative control with the learned policy. If human intervention is not required, the system operates autonomously with the policy. In the intervention evaluation experiment, an additional motion data collected, including succeeded data without interventions, are integrated with the original motion data and re-trained. In the re-training, both original and additional data sets are treated equally. Models trained in the past are not reused.

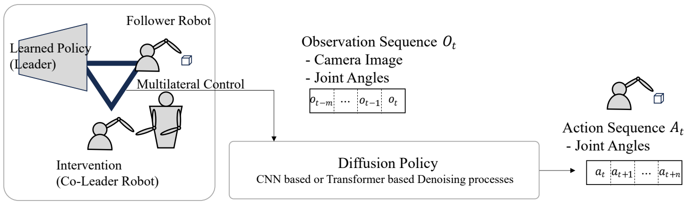


Figure 3: Training Model for CHG-DAgger. As a nominal behavior cloning policy, we adopt DP [14].

4 EXPERIMENTAL SETUP

Our experimental setup is shown in Figure 5. The leader robot used for demonstrations, the co-leader robot for intervention, and the follower robot for task execution were all equipped with 3D Systems 3-degree-of-freedom haptic device, 3D Touch [19]. All robots were connected to a single control PC, and multilateral control was performed at a 1 msec cycle. Torque commands were sent at each control cycle, and joint angle measurements were collected. Additionally, a separate inference PC equipped with an NVIDIA RTX A6000 GPU was prepared. As a fixed scene camera, an Intel RealSense D435 was connected to the inference PC. The inference PC and control PC communicated via UDP at a 1 msec cycle. Inference was performed every $800msec$, generating reference joint trajectories every $100msec$.

Table 1 and Figure 4 show the three robot tasks conducted in this study. For each task, we prepared a scenario to occur the OOD for the evaluation experiment. One model is trained for each task. In the "Move Tape" task, the robot inserts the its end-effector into the hollow part of an electrical tape and slides it on a desk to moves it to the red area. For evaluation, the initial position of the tape was set on the right half area during the collection of training data for the initial policy, and rollouts were performed with the initial position on the left half area. The "Cube Rotate" task is a simple manipulation task that involves rotating a cube. The direction of rotation and the color of the face at the end of task are predefined. The robot is trained with data of starting from the direction required one or two rotations. In the rollout, the task starts from a direction required three rotations to complete it. In the "Raise Bottle" task, the robot pushes the top of a fallen spice bottle and lifts it up. The robot is trained using data that the initial position is close to the robot and the initial orientation is limited to a vertical direction. However, in the OOD evaluation, the bottle starts from a distant or tilted position. In many cases, during the early stages of the rollout, the bottle rotates away from the vertical position, leading to failure, requiring the operator to reposition the bottle vertically to continue the task.

The project of this investigation was approved and conducted according to the "Ethical Guidelines for Research" of Toyota Motor Corporation. Informed consent was obtained from all participants including details of the experimental procedures and our privacy protection policies.

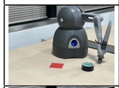
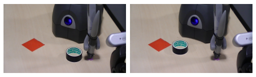
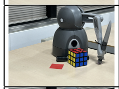
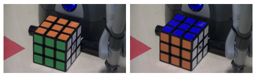
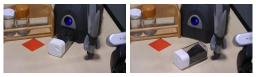
	Task Name	Task Description	Distribution Shift	ID and OOD
	Move Tape	Move the electrical tape. The robot inserts the robot end into the center of the tape and moves the tape to the red area.	Tape was set on right half area for training and left half for rollout.	
	Rotate Cube	Rotating a Rubik's Cube. The direction of rotation and the color of the face at the end are defined.	Different directions of the cube and different color combinations.	
	Raise Bottle	Stand up fallen spice bottle by pushing and pulling up the robot's end. The bottom of the bottle is facing the robot.	Different bottle positions and directions.	

Table 1: Tasks and Distribution Shift for Evaluation

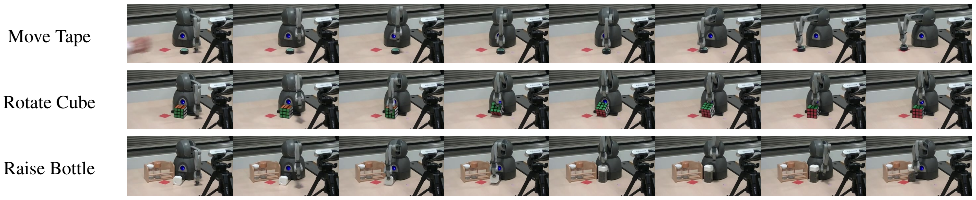


Figure 4: Photo Sequence of Tasks

5 EXPERIMENTAL RESULTS

We empirically demonstrate that our proposed method, CHG-DAgger, can realize IIL through cooperative control between humans and policies for some tasks in the real world. Furthermore, we show that the intervention time is shorter than the conventional method HD-DAgger.

5.1 Retraining with IIL

IIL-based retraining was performed using the proposed method for the tasks "Move Tape" and "Rotate Cube" from Table 1. The success rates are shown in Table 2. The initial policy was trained on expert-collected demonstrations with 100 episodes. And an additional 100 motion data collected in

OOD with intervention are integrated with the original 100 motion data and re-trained. The succeeded data without interventions in additional data was less than 10% of the motion because the experiment was conducted in a situation where the task was likely to fail due to OOD. The success rate was measured over approximately 25 rollouts. The randomized OOD conditions, for example, initial positions, were predetermined and fixed in advance for the success rate evaluations under (b) OOD conditions and (c) post-retraining success rates to ensure consistency. The success rates of the initial policy are shown in Table 2(a). For each task, the success rate exceeded 96.0% when there were no OOD conditions. In the ”Move Tape” task, the success rate dropped to 16.6% when OOD conditions were introduced but increased to 77.8% after retraining. For the ”Rotate Cube” task, the success rate decreased to 12.0% under OOD conditions but rose to 80.0% after retraining. These results indicate that retraining was achieved using the proposed method.

5.2 Intervention Time in IIL

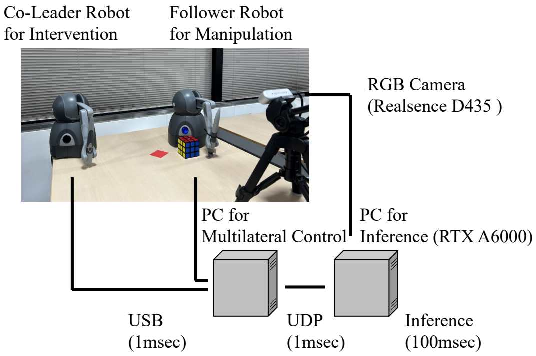
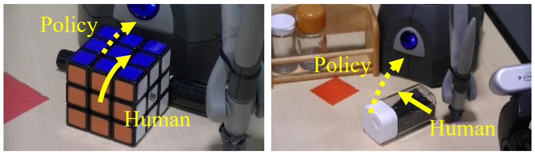

Co-Leader Robot for Intervention
Follower Robot for Manipulation
RGB Camera (Realsence D435)
PC for Multilateral Control
PC for Inference (RTX A6000)
USB (1msec)
UDP (1msec)
Inference (100msec)

Figure 5: Experimental Hardware Setup



(a) Minor Adjustment (b) Major Correction

Figure 6: Human and Policy Trajectory

We compared the intervention time required for the conventional HG-Dagger method and the proposed CHG-Dagger method to recover from failures in the tasks “rotating a cube” and “lifting a bottle”. Although there was no significantly different decrease in intervention time across all tasks, a marginally significant trend ($p = 0.06$) was observed for the CHG-DAgger-first group on the Rotate Cube task.

The subjects were 10 individuals aged 20 to 60 and were divided into two groups to consider the ordeting effect: (A) a group that performed HG-Dagger first, and (B) a group that performed CHG-Dagger first. In group (B). Hence, if the shared autonomy framework is more user-friendly, the intervention time by the proposal is expected to be shorter than that of the baseline, regardless of the ordering effect. The average time for 10 interventions using HG-Dagger and the average time for 10 interventions using CHG-Dagger were calculated. This was done for each subject and each task. As shown in Figure 6, when OOD occurred in the “Rotate Cube” task, the trajectory generated by the policy was almost the same as the one generated without OOD, but shifted approximately 2 cm upward. Therefore, the human expert needed to make a “minor adjustment” by correcting downward while operating in the same direction as the policy. On the other hand, “major correction” was necessary in the “Raise Bottle”. The initial position and direction of the bottle needed to be corrected. However, the policy had not learned the action of correcting the initial position and direction, so it generated an action to push it up. Therefore, the expert human needed to correct in a direction different from the trajectory generated by the policy.

Though the result indicates ”a marginal” trend in significance, the intervention time is reduced by our proposal for a major part, 80% of the subjects. The figure 7 indicates that the distribution of intervention time increases or decreases for the ”Rotate Cube” task, which required ”minor adjustment,” is shown in Figure 7. The ratio of the average intervention times of each subject for CHG-Dagger and HG-Dagger was calculated. Values less than 1 indicate that CHG-Dagger had shorter intervention times. The table 3 shows intervention time for each task sand group. In the ”Rotate Cube” task, where ”simple corrections” were needed, the intervention times were almost the same in Group (A), but in Group (B), the intervention time for CHG-Dagger was 8.5% shorter. A t-test of the results for Group (B) showed that the results did not reach the 5% significance level, but it showed a significant difference in the 6% level of significance. On the other hand, in the ”Raise Bottle” task, which required ”complex corrections,” the intervention time for CHG-Dagger was 1.6 times longer in group (A), and 1.7 times in group (B). In summary, while the intervention time was longer for the ”Raise Bottle” task, which required ”major correction,” the intervention time using the proposed CHG-Dagger was 8.5% shorter than HG-Dagger when prior knowledge of the policy was limited in the ”Rotate Cube” task, where only ”minor adjustment” was needed.

Task	**Success Rate (%)**		
	(a) Initial Policy without OOD	(b) Initial Policy with OOD	(c) Retrained Policy
Move Tape	96.2	16.6	77.8
Rotate Cube	96.0	12.0	80.0

Table 2: Success Rate

Task	Intervention Time (sec)			
	(A)HG-DAgger First		(B)CHG-DAgger First	
	HG-DAgger (baseline))	CHG-DAgger (proposed))	HG-DAgger (baseline))	CHG-DAgger (proposed))
Rotate Cube	5.5	5.4	4.7	4.3
Raise Bottle	8.7	13.8	8.9	15.2

Table 3: Intervention Time

6 Conclusion and Limitations

In this paper, we focused on the usability of IIL and presented a new approach to IIL through physical cooperative control between humans and learned policies for visuomotor control policies. In the proposed method, multilateral control was adopted and simplified for low-cost hardware and long update cycles time of reference trajectory by policy’s inference, connecting it to a representative visuomotor policy, DP. The proposed system was empirically evaluated in multiple tasks on intervention time. Although the intervention time with the proposed CHG-DAgger was longer than HG-DAgger for the ”Raise Bottle” task, which required ”major correction,” the intervention time was shorter in 80% subject for ”Rotate Cube” task which required only ”minor adjustment.” In addition, it was 8.5% shorter for users with little prior knowledge of the policy. These results suggest that CHG-DAgger provides a more intuitive and efficient way to handle task failures in real-world applications, paving the way for continuous learning.

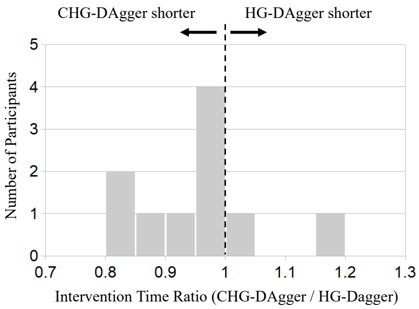


Figure 7: Histogram of Intervention Time Ratio of Rotate Cube Task

However, our system has several limitations. First, while intervention times were shorter with the proposed method for simple tasks, they became longer for more complex tasks. This is because, in simple tasks, the actions performed by the human and the learned policy are similar, making it easier for humans to operate. In contrast, for complex tasks, the movements executed by the learned model differ from those performed by the human, introducing unintended actions that make it more difficult for humans to operate. In such cases, it is easier to fully switch control to the human, but the timing for switching control back to the learned policy is difficult to determine. Further research is needed on how to implement cooperative control to resolve this trade-off. For example, previous research on shared autonomy may provide useful insights. Second, evaluating only intervention time as a performance metric is insufficient, as we cannot fully assess usability. Furthermore, this study did not evaluate whether the system can learn multimodal recovery methods that vary between individuals. Nevertheless, we expect this research to be a stepping stone for future studies.

References

[1] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, J. Dabis, C. Finn, K. Gopalakrishnan, K. Hausman, A. Herzog, J. Hsu, J. Ibarz, B. Ichter, A. Irpan, T. Jackson, S. Jesmonth, N. Joshi, R. Julian, D. Kalashnikov, Y. Kuang, I. Leal, K.-H. Lee, S. Levine, Y. Lu, U. Malla, D. Manjunath, I. Mordatch, O. Nachum, C. Parada, J. Peralta, E. Perez, K. Pertsch, J. Quiambao, K. Rao, M. Ryoo, G. Salazar, P. Sanketi, K. Sayed, J. Singh, S. Sontakke, A. Stone, C. Tan, H. Tran, V. Vanhoucke, S. Vega, Q. Vuong, F. Xia, T. Xiao, P. Xu, S. Xu, T. Yu, and B. Zitkovich. Rt-1: Robotics transformer for real-world control at scale. In *arXiv preprint arXiv:2212.06817*, 2022.

[2] A. Brohan, N. Brown, J. Carbajal, Y. Chebotar, X. Chen, K. Choromanski, T. Ding, D. Driess, A. Dubey, C. Finn, P. Florence, C. Fu, M. G. Arenas, K. Gopalakrishnan, K. Han, K. Hausman, A. Herzog, J. Hsu, B. Ichter, A. Irpan, N. Joshi, R. Julian, D. Kalashnikov, Y. Kuang, I. Leal, L. Lee, T.-W. E. Lee, S. Levine, Y. Lu, H. Michalewski, I. Mordatch, K. Pertsch, K. Rao, K. Reymann, M. Ryoo, G. Salazar, P. Sanketi, P. Sermanet, J. Singh, A. Singh, R. Soricut, H. Tran, V. Vanhoucke, Q. Vuong, A. Wahid, S. Welker, P. Wohlhart, J. Wu, F. Xia, T. Xiao, P. Xu, S. Xu, T. Yu, and B. Zitkovich. Rt-2: Vision-language-action models transfer web knowledge to robotic control. In *arXiv preprint arXiv:2307.15818*, 2023.

[3] C. Chi, Z. Xu, C. Pan, E. Cousineau, B. Burchfiel, S. Feng, R. Tedrake, and S. Song. Universal manipulation interface: In-the-wild robot teaching without in-the-wild robots. In *Proceedings of Robotics: Science and Systems (RSS)*, 2024.

[4] Z. Fu, T. Z. Zhao, and C. Finn. Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation. In *Conference on Robot Learning (CoRL)*, 2024.

[5] H. Kim, Y. Ohmura, and Y. Kuniyoshi. Gaze-based dual resolution deep imitation learning for high-precision dexterous robot manipulation. *IEEE Robotics and Automation Letters*, 6(2): 1630–1637, 2021.

[6] Y. Ishida, Y. Noguchi, T. Kanai, K. Shintani, and H. Bito. Robust imitation learning for mobile manipulator focusing on task-related viewpoints and regions. In *International Conference on Intelligent Robots and Systems (IROS)*, 2024.

[7] S. Ross and J. A. Bagnell. Efficient reduction for imitation learning. In *International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2010.

[8] S. Ross, G. J. Gordon, and J. A. Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2011.

[9] R. Hoque, A. Balakrishna, E. Novoseller, A. Wilcox, D. S. Brown, and K. Goldberg. Thriftydagger: Budget-aware novelty and risk gating for interactive imitation learning. In *Proceedings of 5th Annual Conference on Robot Learning (CoRL)*, 2021.

[10] J. Zhang and K. Cho. Query-efficient imitation learning for end-to-end simulated driving. In *Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI)*, 2017.

[11] M. Kelly, C. Sidrane, K. Driggs-Campbell, and M. J. Kochenderfer. Hg-dagger: Interactive imitation learning with human experts. In *Proceedings of the 2019 International Conference on Robotics and Automation (ICRA)*, Montreal, 2019.

[12] E. Jang, A. Irpan, M. Khansari, D. Kappler, F. Ebert, C. Lynch, S. Levine, and C. Finn. Bc-z: Zero-shot task generalization with robotic imitation learning. In *Proceedings of the 5th Conference on Robot Learning (PMLR)*, volume 164, pages 991–1002, 2022.

[13] S. Katsura, Y. Matsumoto, and K. Ohnishi. Realization of ”law of action and reaction” by multilateral control. *IEEE Transactions on Industrial Electronics*, 52(5):1196–1205, Sep 2005.

[14] C. Chi, Z. Xu, S. Feng, E. Cousineau, Y. Du, B. Burchfiel, R. Tedrake, and S. Song. Diffusion policy: Visuomotor policy learning via action diffusion. In *Proceedings of Robotics: Science and Systems (RSS)*, 2023.

[15] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn. Learning fine-grained bimanual manipulation with low-cost hardware. In *Proceedings of Robotics Science and Systems (RSS)*, 2023.

[16] T. Adachi, K. Fujimoto, S. Sakaino, and T. Tsuji. Imitation learning for object manipulation based on position/force information using bilateral control. In *Proceedings of 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018.

[17] T. Tsuji, K. Natori, H. Nishi, and K. Ohnishi. A controller design method of bilateral control system. *European Power Electronics and Drives Journal (EPE Journal)*, 16(2):22–27, 2006.

[18] Y. Kuroki, Y. Kosaka, T. Takahashi, E. Niwa, H. Kaminaga, and Y. Nakamura. Cr-n alloy thin-film based torque sensors and joint torque servo systems for compliant robot control. In *2013 IEEE International Conference on Robotics and Automation (ICRA)*, Karlsruhe, Germany, 2013.

[19] 3d systems touch haptics device. `https://www.3dsystems.com/haptics-devices/touch`.