Assisted Teleoperation for Scalable Robot Data Collection

Shivin Dass*[1], Karl Pertsch*[1], Hejia Zhang[1], Youngwoon Lee[1], Joseph J. Lim[2], Stefanos Nikolaidis[1]

[1]University of Southern California, [2]KAIST

Abstract: Large-scale robotic datasets are essential for effective robot pre-training. However, collecting large-scale robotic data is expensive and slow as each operator can control only a single robot at a time. To make this costly data collection process efficient and scalable, we propose a novel assisted teleoperation system, which automates part of the demonstration collection process using a learned assistive policy. The assistive policy autonomously executes repetitive behaviors in data collection and asks for human input only when it is uncertain about which subtask or behavior to execute. We conduct teleoperation user studies both with a real robot and a simulated robot fleet and demonstrate that our assisted teleoperation system reduces human operators' mental load while improving data collection efficiency. Further, it enables a single operator to control multiple robots in parallel, which is a first step towards scalable robotic data collection.[2].

Keywords: Teleoperation, Shared Autonomy, Scalable Data Collection

1 Introduction

Recently, many works have shown impressive robot learning results from diverse, human-collected demonstration datasets [1, 2, 3, 4]. They underline the importance of scalable robot data collection for robot pre-training. Yet, the current standard approach for demonstration collection, human teleoperation, is tedious and costly: tasks need to be demonstrated repeatedly and each operator can control only a single robot at a time. Research in teleoperation has focused on exploring different interfaces, such as VR controllers [5] and smart phones [1], but does not address the aforementioned bottlenecks to scaling data collection. Thus, current teleoperation systems are badly equipped to deliver the scalability required by modern robot learning pipelines.

Our goal is to improve the scalability of robotic data collection by providing assistance to the human operator during teleoperation. We take inspiration from other fields of machine learning, such as semantic segmentation, where costly labeling processes have been substantially accelerated by providing human annotators with learned assistance systems, e.g., in the form of rough segmentation estimates, that drastically reduce the labeling burden [6, 7]. Similarly, we propose to train assistive *policies*, that can automate control of repeatedly demonstrated behaviors and ask for user input only when facing a novel situation or when unsure which behavior to execute. Thereby, we aim to reduce the mental load of the human operator and enable scalable teleoperation by allowing a single operator to perform data collection with multiple robots in parallel.

In order to build an assistive system for robotic data collection, we need to solve two key challenges: (1) we need to learn assistive policies from diverse human-collected data, which is known to be challenging [8], and (2) we need to learn when to ask for operator input while keeping such interventions at a minimum. To address these challenges, we propose to use a hierarchical stochastic policy that can learn effectively from diverse human data. Further, we use the policy's stochastic predictions to

[*] Denotes equal contribution. Correspondence to sdass@usc.edu, pertsch@usc.edu.

[2] For a more detailed version of this paper, see `https://clvrai.com/pato/`

6th Conference on Robot Learning (CoRL 2022), Auckland, New Zealand.

estimate its uncertainty about how to act in the current scene and which task to pursue. Then, we use this estimate to elicit operator input only if the assistive policy is uncertain about how to proceed.

The main contribution of this paper is a novel assisted teleoperation system, which enables scalable robotic data collection using a hierarchical assistive policy. We evaluate the effectiveness of our approach in a user study in which operators collect datasets of diverse kitchen-inspired manipulation tasks with a real robot. We find that our proposed *assisted* teleoperation approach reduces operators' mental load and improves their demonstration throughput. We further demonstrate that our approach allows a single operator to control data collection with multiple robots simultaneously in a simulated manipulation environment – a first step towards more scalable robotic data collection.

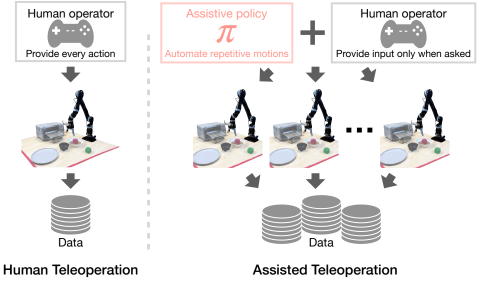


Figure 1: Policy-assisted teleoperation enables large-scale data collection by minimizing human operator inputs and mental efforts with an assistive policy, which autonomously performs repetitive subtasks. This allows a human operator to simultaneously manage multiple robots.

2 Related Work

Teleoperation is the most popular approach for collecting robot demonstrations [5, 1, 9, 4, 10]. Yet, none of these works explores active assistance of the human operator during teleoperation. The idea of sharing efforts between humans and robots when solving tasks has a rich history in the human-robot-interaction (HRI) community [11, 12, 13, 14, 15, 16, 17, 18], but approaches often require a pre-defined set of goals to infer the operator's intent. In contrast, more recent approaches explore joint human-robot data collection without such pre-defined goals [19, 20, 21, 22, 23, 24], but they only focus on the single-task case. We instead propose an approach for scalable collection of diverse, multi-task datasets. For a more detailed overview of related works, see Section A.

3 Approach

An *assistive policy* $\pi(a|s)$ produces actions a, e.g., robot end-effector displacements, given states s, e.g., raw RGB images. To enable scalable data collection of a dataset $\mathcal{D}$, the policy should control the robot and minimize required human inputs, which allows the human operator to divert attention away from the robot over contiguous intervals, e.g., to attend to other robotic agents collecting data in parallel. To train the assistive policy π we assume access to a pre-collected dataset $\mathcal{D}_{\text{pre}}$ of diverse agent experience, e.g., from scripted policies, previously collected data on different tasks or human play [25]. Crucially, we explicitly require our approach to handle scenarios in which the newly collected dataset $\mathcal{D}$ contains behaviors that are *not* present in $\mathcal{D}_{\text{pre}}$. Thus, it is *not* possible to fully automate data collection given the pre-training dataset. Instead, the system needs to request human input for unseen behaviors while providing assistance for known behaviors.

Learning Assistive Policies from Multi-Modal Data. We build on prior work in imitation of long-horizon, multi-task human data [26]. We propose to use a hierarchical policy with a subgoal predictor $p(s_g|s, z)$ and a low-level subgoal reaching policy $\pi_{\text{LL}}(a_t|s_t, s_g)$ (see Figure 5). We condition the subgoal predictor on a stochastic latent variable z to allow prediction of the *full distribution* of possible subgoals. We train the policy with behavior cloning, for details see appendix, Section B.

Deciding When to Request User Input. Intuitively, the policy should request help when it is uncertain about what action to take next. This can occur in two scenarios: (1) the policy faces a situation that is not present in the training data, so it does not know which action to take, or (2) the policy faces a seen situation, but the training trajectories contain multiple possible continuations and

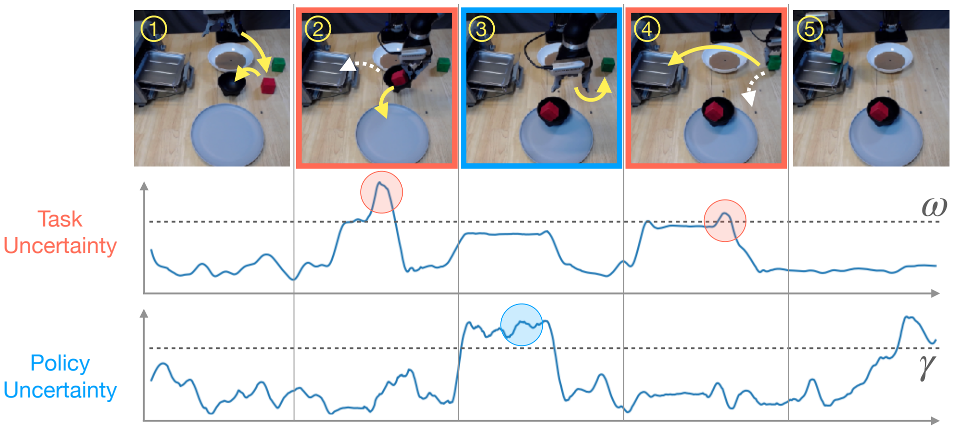


Figure 2: Visualization of our assistive teleoperation policy on a task from the real-robot user study. The policy autonomously executes familiar behaviors, but asks for user input in frames (2) and (4) when the task uncertainty surpasses the threshold ω to determine where to place bowl and green block (white vs. yellow arrow). Further, the policy asks for user input in frame (3) when the policy uncertainty estimate surpasses its threshold γ for the unseen transition between placing the bowl and picking up the green block. For qualitative video results, see https://youtu.be/TRW6yrG8k-A.

the policy is not sure which one to pick. The latter scenario commonly occurs during the collection of diverse datasets, since trajectories for different tasks often intersect.

Our hierarchical model allows us to separately estimate both classes of uncertainty. To estimate whether a given state is unseen, we train an ensemble of K low-level reaching policies: unseen states will have high disagreement $D(a^{(1)}, \dots, a^{(K)})$ between the actions predicted by these ensemble policies. To estimate the policy's certainty about the task we sample from the distribution of subgoals produced by the subgoal predictor and compute the inter-subgoal variance $Var(s_g^{(1)}, \dots, s_g^{(N)})$. We leverage both uncertainty estimates to decide on whether the assistive policy should continue controlling the robot or whether it should stop and ask for human input. We found a simple thresholding scheme sufficient, with threshold parameters γ, ω for the ensemble disagreement and subgoal variance, respectively (see Figure 7).

4 Experiments

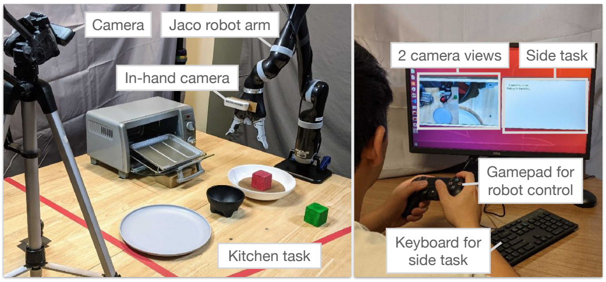


Figure 3: User study setup. **(left)** A Kinova Jaco arm, front-view and in-hand cameras, and objects for kitchen-inspired tasks are placed on the workspace. **(right)** A human operator can watch a monitor, which shows either the camera inputs or a side task. The operator uses a gamepad to control the robot, and uses a keyboard to solve the side task.

To evaluate the effectiveness of our assisted teleoperation, we conduct a user study ($N = 16$) in which users teleoperate a Kinova Jaco 2 robot arm to collect diverse robot manipulation demonstrations for kitchen-inspired long-horizon tasks, e.g., "place ingredients in bowl" and "place bowl in oven" (see Figure 3). Users teleoperate the robot's end-effector via joystick and buttons on a standard gamepad controller. The users are also asked to solve simple side tasks during teleoperation to measure their ability to divert attention and conduct other tasks. To train our assistive policy, we collect a pre-training dataset of 120 demonstrations. Crucially, during the user study the operators need to collect *unseen* long-horizon tasks. We compare our approach to (1) **teleoperation without assistance**, the current standard approach to collection robot demonstration data and (2) **ThriftyDAgger** [23], the closest prior work to ours for interactive human-robot data collection. ThriftyDAgger is designed to minimize human inputs during single-task demonstration collection by requesting human input only in *critical* states where a learned value function estimates low probability for reaching the goal.

We compare the teleoperation speed and number of completed side tasks in Table 1. Only the approaches *with assistance* allow the operator to divert their attention towards solving side tasks. Additionally, we find that our method enables the most effective teleoperation, since it requests user inputs at appropriate points in time (see Figure 2. We also measure users perceived mental load via the NASA TLX survey [27, 28] and their perception of the robot's intelligence, their satisfaction and trust in the system via a custom survey [16], which we administer after every teleoperation session. Study participants perceived the robot with our approach to be significantly more intelligent and trustworthy than with the comparison approaches, they were more satisfied with their collaboration and showed lower mental workload (for a detailed statistical analysis of the survey results, see Section C).

Table 1: Average number of completed side tasks and teleoperation time per demonstration during the real-robot teleoperation user study.

Approach	Avg. Num. of completed side tasks	Avg. teleop time (seconds)
Unassisted	0.25 (±0.66)	109.5 (±31.4)
ThriftyDAgger	**13.06 (±9.63)**	105.9 (± 29.5)
Ours	**15.88 (±7.11)**	**85.0 (±18.2)**

A key factor in the subjective differences between the two approaches is their ability to elicit user feedback at appropriate times: when the robot is at a decision point between two possible task continuations (see Figure 2, frames (2) and (4)). ThriftyDAgger's risk-based objective is not sensitive to such decision points and thus it rarely asks for user feedback. It instead executes one of the possible subtasks at random. In our study we found that this lead to erroneous skill executions in 48 % of cases. Such errors require tedious correction by the user, deteriorating their trust in the system and their teleoperation efficiency. In contrast, our approach leverages its estimate of task uncertainty (see Figure 2) to correctly elicit user feedback in 82 % of cases, leading to higher perceived levels of trust and reduced mental load.

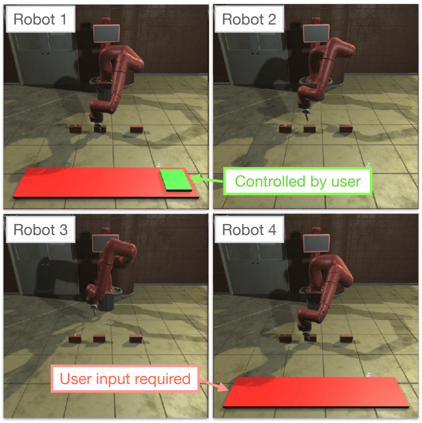
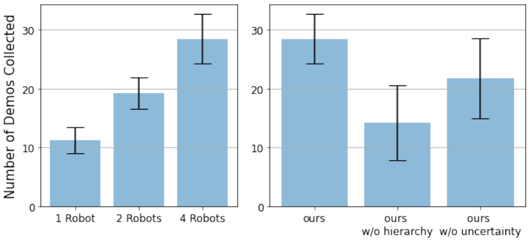


Figure 4: Setup for our multi-robot teleoperation study in simulation.

Scaling Data Collection to Multiple Robots. An important application of our approach is multi-robot teleoperation, in which a single operator performs data collection with multiple robots in parallel and periodically attends to different robots. To test this, we conduct a teleoperation study (N = 10) with *multiple* simulated robots in the realistic physics simulator [10] (see Figure 4). A human user is asked to collect demonstrations for a block stacking task with multiple robots *in parallel* via the same gamepad interface used in the real robot study. The user can switch control between different robots with a button press. We measure the total number of collected demonstrations across the robot fleet in a fixed time frame of T = 4 minutes.

We show that our approach enables strong scaling of data collection throughput (see right). As expected, the scaling is not linearly proportional, i.e., four robots do not lead to four times more demonstrations collected. This is because simultaneous teleoperation of a larger fleet requires more context switches between the robots, reducing the effective teleoperation time. We also perform two ablations of our method in the 4-robot setup: (1) **ours w/o hierarchy**, which trains an ensemble of *flat* stochastic policies $\pi(a|s)$, and (2) **ours w/o uncertainty**, which removes the uncertainty-based requesting of user input. Both ablations perform worse than our approach, indicating the importance of the introduced policy architecture and user feedback request mechanism.



5 Acknowledgement

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant (No.2019-0-00075, Artificial Intelligence Graduate School Program, KAIST) and National Research Foundation of Korea (NRF) grant (NRF-2021H1D3A2A03103683), funded by the Korea government (MSIT).

References

[1] A. Mandlekar, Y. Zhu, A. Garg, J. Booher, M. Spero, A. Tung, J. Gao, J. Emmons, A. Gupta, E. Orbay, S. Savarese, and L. Fei-Fei. Roboturk: A crowdsourcing platform for robotic skill learning through imitation. In *CoRL*, 2018.

[2] S. Cabi, S. G. Colmenarejo, A. Novikov, K. Konyushkova, S. Reed, R. Jeong, K. Zolna, Y. Aytar, D. Budden, M. Vecerik, O. Sushkov, D. Barker, J. Scholz, M. Denil, N. de Freitas, and Z. Wang. Scaling data-driven robotics with reward sketching and batch reinforcement learning. *RSS*, 2019.

[3] Y. Lu, K. Hausman, Y. Chebotar, M. Yan, E. Jang, A. Herzog, T. Xiao, A. Irpan, M. Khansari, D. Kalashnikov, and S. Levine. Aw-opt: Learning robotic skills with imitation and reinforcement at scale. In *CoRL*, 2021.

[4] 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. In *RSS*, 2022.

[5] T. Zhang, Z. McCarthy, O. Jow, D. Lee, X. Chen, K. Goldberg, and P. Abbeel. Deep imitation learning for complex manipulation tasks from virtual reality teleoperation. In *ICRA*, 2018.

[6] L. Castrejon, K. Kundu, R. Urtasun, and S. Fidler. Annotating object instances with a polygon-rnn. In *CVPR*, 2017.

[7] D. Acuna, H. Ling, A. Kar, and S. Fidler. Efficient interactive annotation of segmentation datasets with polygon-rnn++. In *CVPR*, pages 859–868, 2018.

[8] 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. *CoRL*, 2021.

[9] A. Gupta, V. Kumar, C. Lynch, S. Levine, and K. Hausman. Relay policy learning: Solving long-horizon tasks via imitation and reinforcement learning. *CoRL*, 2019.

[10] Y. Lee, E. S. Hu, Z. Yang, A. Yin, and J. J. Lim. IKEA furniture assembly environment for long-horizon complex manipulation tasks. *ICRA*, 2021. URL https://clvrai.com/furniture.

[11] S. Javdani, S. S. Srinivasa, and J. A. Bagnell. Shared autonomy via hindsight optimization. In *RSS*, 2015.

[12] M. Selvaggio, M. Cognetti, S. Nikolaidis, S. Ivaldi, and B. Siciliano. Autonomy in physical human-robot interaction: A brief survey. *IEEE Robotics and Automation Letters*, 2021.

[13] P. Berthet-Rayne, M. Power, H. King, and G.-Z. Yang. Hubot: A three state human-robot collaborative framework for bimanual surgical tasks based on learned models. In *ICRA*, pages 715–722. IEEE, 2016.

[14] A. Pichler, S. C. Akkaladevi, M. Ikeda, M. Hofmann, M. Plasch, C. Wögerer, and G. Fritz. Towards shared autonomy for robotic tasks in manufacturing. *Procedia Manufacturing*, 11: 72–82, 2017.

[15] Y. Gao and S. Chien. Review on space robotics: Toward top-level science through space exploration. *Science Robotics*, 2(7):eaan5074, 2017.

[16] S. Nikolaidis, Y. X. Zhu, D. Hsu, and S. Srinivasa. Human-robot mutual adaptation in shared autonomy. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 294–302. IEEE, 2017.

[17] M. Johns, B. Mok, D. Sirkin, N. Gowda, C. Smith, W. Talamonti, and W. Ju. Exploring shared control in automated driving. In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 91–98. IEEE, 2016.

[18] B. D. Argall. Autonomy in rehabilitation robotics: An intersection. *Annual Review of Control, Robotics, and Autonomous Systems*, 1:441, 2018.

[19] M. Kelly, C. Sidrane, K. Driggs-Campbell, and M. J. Kochenderfer. Hg-dagger: Interactive imitation learning with human experts. In *ICRA*, 2019.

[20] K. Menda, K. Driggs-Campbell, and M. J. Kochenderfer. Ensembledagger: A bayesian approach to safe imitation learning. In *IROS*, 2019.

[21] J. Zhang and K. Cho. Query-efficient imitation learning for end-to-end autonomous driving. *AAAI*, 2017.

[22] R. Hoque, A. Balakrishna, C. Putterman, M. Luo, D. S. Brown, D. Seita, B. Thananjeyan, E. Novoseller, and K. Goldberg. Lazydagger: Reducing context switching in interactive imitation learning. In *2021 IEEE 17th International Conference on Automation Science and Engineering (CASE)*, 2021.

[23] 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. *CoRL*, 2021.

[24] H. M. Clever, A. Handa, H. Mazhar, K. Parker, O. Shapira, Q. Wan, Y. Narang, I. Akinola, M. Cakmak, and D. Fox. Assistive tele-op: Leveraging transformers to collect robotic task demonstrations. *arXiv preprint arXiv:2112.05129*, 2021.

[25] C. Lynch, M. Khansari, T. Xiao, V. Kumar, J. Tompson, S. Levine, and P. Sermanet. Learning latent plans from play. In *CoRL*, 2020.

[26] A. Mandlekar, F. Ramos, B. Boots, L. Fei-Fei, A. Garg, and D. Fox. Iris: Implicit reinforcement without interaction at scale for learning control from offline robot manipulation data. *ICRA*, 2020.

[27] S. G. Hart. Nasa task load index (tlx). 1986.

[28] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In *Advances in psychology*, volume 52, pages 139–183. Elsevier, 1988.

[29] D. A. Pomerleau. Alvinn: An autonomous land vehicle in a neural network. In *NIPS*, pages 305–313, 1989.

[30] A. Billard, S. Calinon, R. Dillmann, and S. Schaal. Survey: Robot programming by demonstration. *Handbook of robotics*, 59(BOOK_CHAP), 2008.

[31] B. D. Argall, S. Chernova, M. Veloso, and B. Browning. A survey of robot learning from demonstration. *Robotics and autonomous systems*, 57(5):469–483, 2009.

[32] H. B. Amor, E. Berger, D. Vogt, and B. Jung. Kinesthetic bootstrapping: Teaching motor skills to humanoid robots through physical interaction. In *Annual conference on artificial intelligence*, 2009.

[33] D. P. Losey, K. Srinivasan, A. Mandlekar, A. Garg, and D. Sadigh. Controlling assistive robots with learned latent actions. In *ICRA*, 2020.

[34] H. J. Jeon, D. P. Losey, and D. Sadigh. Shared autonomy with learned latent actions. *RSS*, 2020.

[35] M. Fontaine and S. Nikolaidis. A quality diversity approach to automatically generating human-robot interaction scenarios in shared autonomy. *arXiv preprint arXiv:2012.04283*, 2020.

[36] M. C. Fontaine and S. Nikolaidis. Evaluating human–robot interaction algorithms in shared autonomy via quality diversity scenario generation. *ACM Transactions on Human-Robot Interaction (THRI)*, 11(3):1–30, 2022.

[37] S. Ross, G. Gordon, and D. Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *AISTATS*, pages 627–635, 2011.

[38] K. Sohn, H. Lee, and X. Yan. Learning structured output representation using deep conditional generative models. In *NIPS*, 2015.

[39] S. Nair, A. Rajeswaran, V. Kumar, C. Finn, and A. Gupta. R3m: A universal visual representation for robot manipulation. In *CoRL*, 2022.

[40] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.

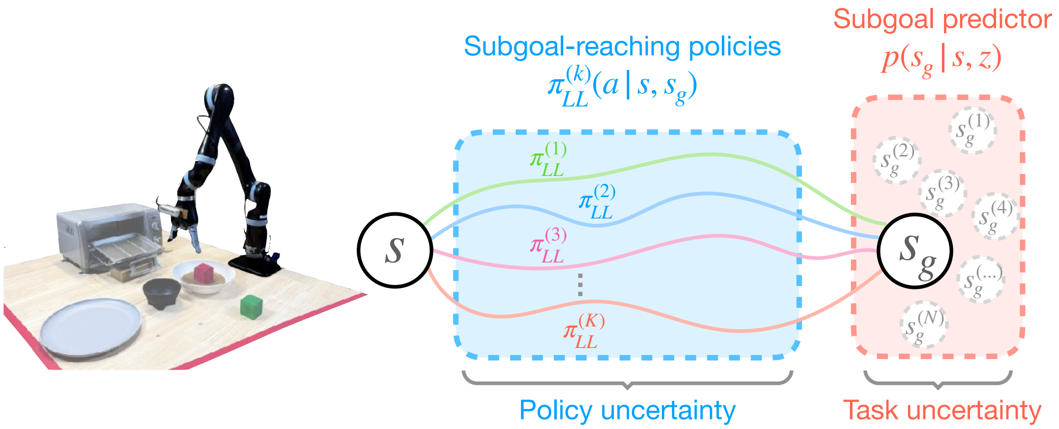


Figure 5: Our assistive policy is hierarchical: a high-level subgoal predictor $p(s_g|s,z)$ and a low-level subgoal-reaching policy $\pi_{LL}(a|s,s_g)$. To decide when to follow the assistive policy, we measure uncertainty of both high-level (subgoal predictor) and low-level (subgoal-reaching policy) decisions. The task uncertainty is estimated using the subgoal predictor's variance, and the policy uncertainty is estimated as a disagreement among an ensemble of subgoal-reaching policies.

A Extended Related Work

Robot Teleoperation. Demonstrations have played a key role in robot learning for many decades [29, 30, 31], thus many approaches have been explored for collecting such demonstrations. While initially kinesthetic teaching was common [32] in which a human operator directly moves the robot, more recently teleoperation has become the norm [5, 1, 9, 4, 10], since separating the human operator and the robot allows for more comfortable human control inputs and is crucial for training policies with image-based inputs. Research into teleoperation systems has focused on exploring different interfaces like VR headsets [5, 4], joysticks [2] and smartphones [1]. Yet, none of these works explores active assistance of the human operator during teleoperation. Others have investigated controlling high-DoF manipulators via low-DoF interfaces through learned embedding spaces [33, 34] to allow people with disabilities to control robotic arms. In contrast, our approach trains assistive policies that automate part of the teleoperation process with the goal of enabling more scalable data collection.

Shared Autonomy. The idea of sharing efforts between humans and robots when solving tasks has a rich history in the human-robot-interaction (HRI) community [11, 12, 13, 14, 15, 16, 17, 18, 35, 36]. Approaches for such *shared autonomy* typically rely on a pre-defined set of goals and aim to infer the intent of the human operator to optimally assist them. Crucially, in the context of data collection, we cannot assume that all goals are known a priori, since a core goal of data collection is to collect previously unseen behaviors. Thus, instead of inferring the operator's intent over a fixed goal set, we leverage the model's estimate over its own uncertainty to determine when to assist and when to rely on operator input.

Interactive Human Robot Learning. In the field of robot learning, many approaches have explored leveraging human input in the learning loop and focused on different ways to decide when to leverage such input. Based on the DAgger algorithm [37], works have investigated having the human themselves decide when to intervene [19], using ensemble-based support estimates [20], using discrepancies between model output and human inputs [21, 22] or risk estimates based on predicted future returns [23]. Yet, all these approaches focus on training a policy for a single task, not on collecting a diverse dataset. Thus, they are not designed to learn from multi-modal datasets or estimate uncertainty about the desired task. We show in our user study that these are crucial for enabling scalable robot data collection.

Assisted Robot Data Collection. Clever et al. [24] aims to assist in robot demonstration collection via a learned policy. They visualize the projected trajectory of the assistive policy to enable the human operator to intervene if necessary. However, they focus on collection of single-task, short-horizon

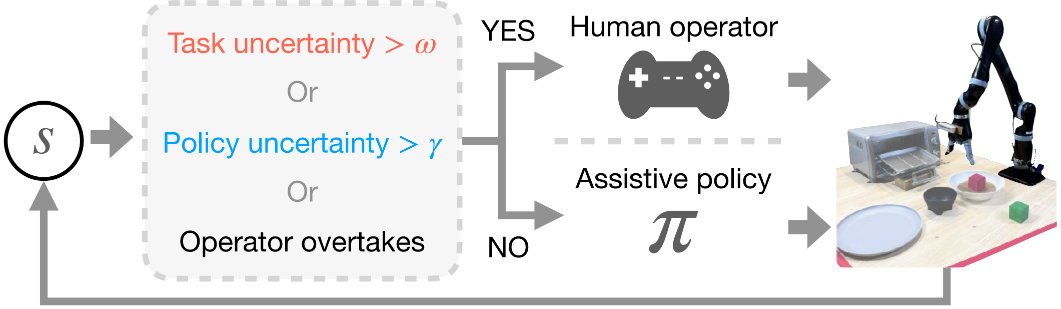


Figure 7: Our approach asks for human inputs when the assistive policy is uncertain about which subtask or action to take. If both the task uncertainty and policy uncertainty are lower than their thresholds, our assistive policy can reliably perform a subtask, reducing the workload of the human operator.

demonstrations and require the operator to constantly monitor the robot to decide when to intervene. In contrast, our system can collect diverse, multi-task datasets and learn when to ask the user for input, enabling more scalable data collection, e.g., with multiple robots in parallel.

B Policy Training Objective

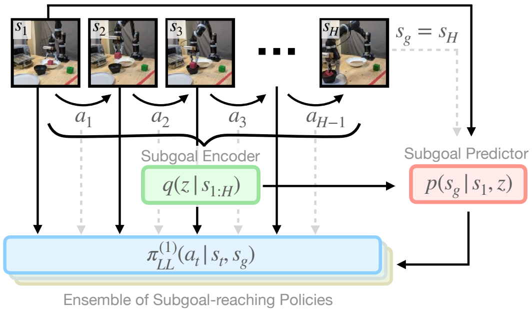


Figure 6: Our hierarchical assistive policy is trained using a pre-collected dataset $\mathcal{D}_{\text{pre}}$. From a sampled trajectory $(s_1, a_1, \ldots, a_{H-1}, s_H)$ of length H, a subgoal predictor $p(s_g|s_1, z)$ is trained as a conditional VAE to cover a multi-modal subgoal distribution, where $s_g = s_H$. Then, an ensemble of subgoal-reaching policies $\pi_{LL}^{(k)}(a_t|s_t, s_g)$ are trained to predict the ground truth actions.

We train the subgoal predictor as a conditional variational auto-encoder over subgoals [38]: given a randomly sampled starting state s_t from the pre-training dataset and a subgoal state $s_{t+H} = s_g$ H steps later in the trajectory, we use a learned inference network $q(z|s_t, s_g)$ to encode s_t and s_g into a latent variable z. We then use the subgoal predictor $p(s_g|s_t, z)$ to decode back to the original subgoal state. During training we apply a subgoal reconstruction loss, as well as a regularization loss on the latent variable z. Finally, the subgoal reaching policy is trained via simple behavioral cloning. We summarize the components of our training model in Figure 6. Our final training objective is:

$$\max_{\theta,\phi,\mu} \; \mathbb{E}_{\substack{(s,a,s_g)\sim\mathcal{D}_{\text{pre}} \\ z\sim q(\cdot|s,s_g)}} \underbrace{p_\theta(s_g|s,z)}_{\text{subgoal reconstruction}} + \underbrace{\pi_{\text{LL},\phi}(a|s,s_g)}_{\text{behavioral cloning}} - \underbrace{\beta D_{\text{KL}}\big(q(z|s,s_g),p(z)\big)}_{\text{latent regularization}} \tag{1}$$

We use θ, ϕ, μ to denote the parameters of the subgoal predictor, goal reaching policy, and inference network, respectively. β is a regularization weighting factor, D_{KL} denotes the Kullback-Leibler divergence, and we use a unit Gaussian prior $p(z)$ over the latent variable.

To execute our assistive policy $\pi(a|s)$, we first sample a latent variable z from the unit Gaussian prior, then pass z and s through the subgoal predictor $p(s_g|s, z)$ to generate a subgoal and then use the goal-reaching policy to predict executable actions $\pi_{\text{LL}}(a|s, s_g)$.

B.1 Implementation Details

For real robot experiments, we use both front-view and in-hand camera images of size $244 \times 244 \times 3$ as an observation. But, instead of using raw images, we use a 2048-dimensional pre-trained visual representation from R3M [39]. In addition, we include the end-effector pose and gripper state of the robot in the observation space. The action space consists of a 3D translational action and 1D discrete gripper action. The conditional-VAE for the subgoal predictor consists of encoder and decoder with 5-layer MLP with 128 hidden units, and ReLU activation. We use an ensemble of $K = 5$ LSTM subgoal-reaching policies with 128 hidden units and the Adam optimizer [40]. We use the same models for simulated experiments but the observation space consists of robot joint positions, velocities, and object poses.

C User Study: Statistical Analysis

During the study, participants agreed more strongly that they trusted the robot to perform the correct action at the correct time for the proposed approach (Wilcoxon signed-rank test, $p = 0.001$). Further, they found the robot to be significantly more intelligent with the proposed method (repeated-measures ANOVA, $F(1, 15) = 5.14, p = 0.039$, Cronbach's α = 0.95) and were significantly more satisfied with their collaboration with the robot ($F(1, 15) = 5.05, p = 0.040$, α = 0.91). Finally, during the NASA TLX survey, participants showed a lower mental workload using the proposed approach compared to the baseline ($F(1, 15) = 5.52, p = 0.033$).

Table 2: Post-execution survey (Likert scales with 7-option response format)

Trust:
Q1. I trusted the robot to do the right thing at the right time.
Robot intelligence ($\alpha = 0.95$)**:**
Q2. The robot was intelligent.
Q3. The robot perceived accurately what my goals are.
Q4. The robot and I worked towards mutually agreed upon goals.
Q5. The robot's actions were reasonable.
Q6. The robot did the right thing at the right time.
Human satisfaction ($\alpha = 0.91$)**:**
Q7. I was satisfied with the robot and my performance.
Q8. The robot and I collaborated well together.
Q9. The robot was responsive to me.