

Real-Time Operator Takeover for Visuomotor Diffusion Policy Training

Nils Ingelhag^{*1}, Jesper Munkeby^{*1}, Michael C. Welle^{*1,2}, Marco Moletta¹, Danica Kragic¹

Abstract—In this work, we introduce a Real-Time Operator Takeover (RTOT) paradigm for imitation learning-based methods, alongside novel insights into leveraging the Mahalanobis distance to automatically detect undesirable states. RTOT enables operators to seamlessly take control of a live visuomotor diffusion policy, guiding the system back into desirable states or reinforcing specific demonstrations. Once the operator has intervened and redirected the system, the control is seamlessly returned to the policy, which resumes generating actions until further intervention is required. We demonstrate that incorporating these targeted takeover demonstrations significantly improves policy performance compared to training solely with an equivalent number of, but longer, initial demonstrations. Furthermore, we provide an in-depth analysis of using the Mahalanobis distance to detect out-of-distribution states, illustrating its utility for identifying critical failure points during execution. Supporting materials, including videos of initial and takeover demonstrations and all rice-scooping experiments, are available on the project website¹.

I. INTRODUCTION

Imitation learning has achieved great success in recent years in a variety of challenging robotic tasks, such as cooking shrimp and wiping wine [1], serving rice and opening bottles using a bottle opener [2], and 6Dof mug flipping, sauce pouring, and spreading [3].

A key component of imitation learning methods, such as Visuomotor Diffusion Policies [3], Action Chunking with Transformers [4], or Visual Imitation through Nearest Neighbor [5], is the expert demonstrations that serve as training data. One approach to enhance these demonstrations is to purify potential noise in imperfect data after collection [6]. Another more prominent approach is to facilitate the collection of high-quality demonstrations using various teleoperation methods, such as virtual reality controllers or hand tracking [7]–[10], augmented reality [11], [12], and leader-follower programming techniques [13]. If significantly boosts the performance of imitation learning methods, they do not address a fundamental problem: how to cover sufficient variations in demonstrations/tasks to make the learned policy more robust. Enhancing robustness in imitation learning is an active area of research [15], [16]. Although such methods improve the handling of scene variations, they still struggle when encountering states that are not present in the training data. Policies often fail in such scenarios, getting stuck due to the lack of demonstrations for recovering from undesirable states.

^{*}These authors contributed equally (listed in alphabetical order).

¹KTH Royal Institute of Technology, Sweden, {ingelhag, munkeby, mwelle, moletta, dani}@kth.se

²INCAR Robotics AB, Sweden, michael.welle@incar-robotics.se

¹Operator-Takeover.github.io

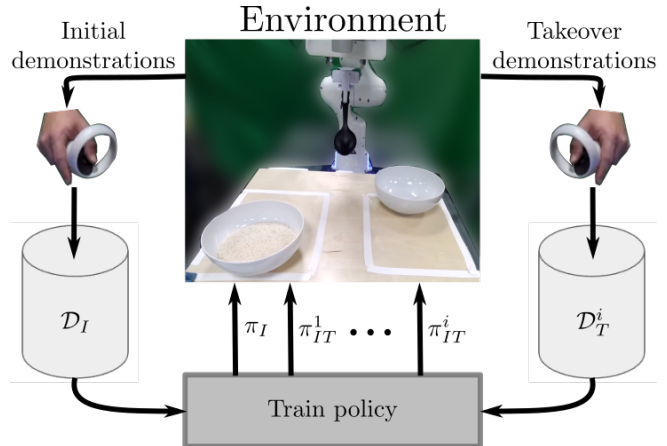


Fig. 1. Our Real-Time Operator Takeover paradigm. After training a policy with a small number of initial demonstrations, we run the policy in the environment with the operator on standby. As soon as the policy enters an undesirable state, the operator seamlessly takes over, and only the takeover portion is recorded as new demonstrations. A new policy is trained, and the paradigm can be repeated until the desired performance is achieved.

In this work, we tackle this challenge with our Real-Time Operator Takeover (RTOT) paradigm. The core idea is that even the most skilled operator may struggle to anticipate and demonstrate recovery scenarios or failure cases the policy might encounter. Our RTOT approach addresses this limitation by first training an initial policy with a smaller set of demonstrations than typically used in the literature. This policy is then deployed in the environment while the operator remains on standby. When the operator observes the system entering or approaching an undesirable state, they seamlessly take control to teleoperate the robot back to a desirable state. This *takeover* demonstrations are recorded, and a new policy is trained using both the initial and the newly recorded takeover demonstrations. As shown in Fig. 1, this paradigm can be repeated as many times as necessary to achieve the desired performance for any given task.

In detail, our contributions are as follows:

- 1) Introduction and validation of the RTOT paradigm on a real-world rice-scooping task.
- 2) Analysis of the Mahalanobis distance (a measure of distance from a point to a distribution) and its potential applications in imitation learning-based methods.
- 3) Extensive videos showcasing the initial demonstration collection, the takeover demonstration collection, and all 50 rice-scooping experiments, available on the project website¹.

II. BACKGROUND & RELATED WORK

A. Visuomotor Diffusion Policies

Diffusion models [17], initially introduced in generative image modeling, have recently gained popularity in robotics as a versatile and powerful framework. Compared to traditional discriminative models, diffusion models demonstrate remarkable generalization capabilities [18]. Fundamentally, visuomotor diffusion-based policy models generate complex structured data distributions iteratively, by learning a stochastic transport map from a known prior distribution (e.g. Gaussian noise) to a desired target distribution. In robotics, the target distribution often consists of action sequences that enable a robot to accomplish specific tasks.

A key strength of diffusion models lies in their ability to learn behaviors using a limited amount of demonstrations [3], as well as requiring minimal explicit modeling of tasks or environments, making these models highly adaptable for a wide range of applications. Recent work has leveraged diffusion models to address planning problems [19]–[21] and perform various robotic manipulation tasks [3], [22], [23], often relying on images as input. Other applications include control of bi-manual mobile manipulation systems for navigating indoor environments and performing kitchen tasks [1], as well as manipulating deformable objects in robot-assisted surgical scenarios [24].

B. Imitation Learning and Out-of-Distribution Detection.

Imitation learning enables robots to acquire skills from expert demonstrations [25] by framing the problem as supervised learning from observations to actions. Recent advancements have incorporated historical context [26]–[29], alternative training objectives [5], [30], multi-task and few-shot learning [31]–[33], and language-conditioned policies for semantic understanding [34]. These methods have improved generalization to task variations [28], [29], [35] and enabled fine-manipulation on low-cost hardware [4].

A key challenge in imitation learning is the compounding error [36], which leads robots to difficult-to-recover, out-of-distribution (OOD) states [37], [38]. To mitigate this, various strategies have been explored, including on-policy expert corrections in DAgger [37] and its variants [39], [40], reward function modifications [41], [42], offline synthetic data generation [43]–[45], and training of recovery policies [46]. However, expert interventions can be time-consuming and impractical with current teleoperation interfaces [44], motivating the proposed seamless real-time takeover paradigm.

Despite efforts in OOD state detection [47], [48], diffusion-based models [3] generally do not integrate these mechanisms and rather rely only on demonstrations which extensively cover the training manifold. A critical limitation remains in detecting distribution shifts during deployment while demonstrations are still being collected. To address this, we propose a human-in-the-loop framework that enables seamless intervention during OOD occurrences, facilitating targeted data collection.

C. Human-in-the-Loop and Real-Time Takeover.

Human-robot interaction (HRI) has advanced considerably in recent years, fundamentally transforming how humans and robots collaborate and coexist [49]. These developments are largely driven by machine learning and the integration of multimodal data [50]. In robotic manipulation, teleoperation enables natural human control of robots, leveraging human cognitive abilities and domain expertise alongside the physical capabilities of humanoid robots [51].

One prevalent class of methods for teleoperation employs virtual-, augmented-, and mixed-reality (VAM) frameworks, which serve as a common interface between human and robot. These frameworks have been applied to diverse areas in robotics, including motion planning and HRI [52], or often used to control robot manipulators via position or velocity control [53], [54]. Applications include object handover [55], visualizing robot intentions during delivery tasks [56], and executing cloth manipulation tasks [7].

To the best of our knowledge, however, there is virtually no work in the realm of HRI for robotic manipulation that explicitly permits a human to seamlessly assume control to correct the robot’s movement while simultaneously recording the new trajectory to make it available for training. Although some studies have employed takeover request (TOR) frameworks in autonomous driving [57] or other have implemented action corrections to ensure successful task completion [7], [58], these approaches do not incorporate the detection of out-of-distribution (OOD) states or the explicit recording of the takeover as a new demonstration. This motivates our RTOT paradigm, which we are going to explain in more details in the following section.

III. METHOD

In this section, we describe our framework for enabling real-time operator takeover to improve visuomotor diffusion policy training.

The core concept, illustrated in Fig.3, involves collecting an initial set of demonstrations (\mathcal{D}_I) to train an initial policy (π_I). The policy is deployed in the environment while the operator remains on standby, ready to takeover if the system approaches an undesirable state. In such cases, the operator can take control in real-time by pressing a button on the VR controller. The operator intervenes to prevent the robot from spilling rice (fourth frame), guides it to successfully deposit the rice into the bowl, and then returns the control back to the policy. These takeover interventions are recorded as new demonstrations and a subsequent policy π_{IT}^1 is trained using the combined dataset of initial and takeover demonstrations. This iterative process ensures that the policy continuously improves by addressing failure cases encountered during execution, rather than relying solely on the initial demonstrations of the operator.

A. Real-Time Takeover

A visuomotor diffusion policy [3] is initially trained on a demonstration dataset \mathcal{D}_I . The demonstrations are collected by teleoperating the robot using the Quest2ROS [8]

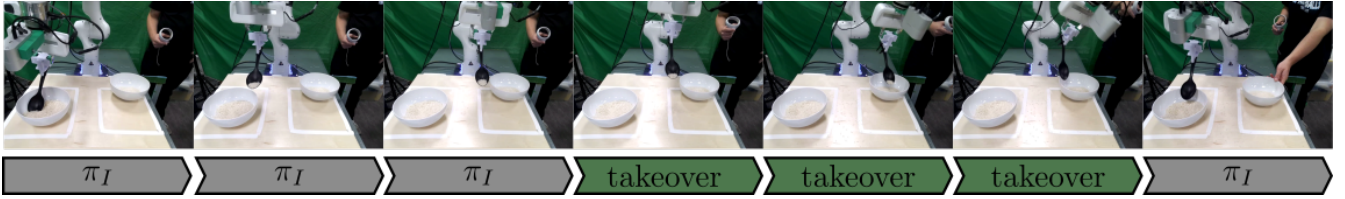


Fig. 2. Real-Time Operator Takeover in action: The policy π_I controls the robot until a state is reached where the operator must take over to avoid spilling rice on the table. After the operator completes the intervention (depositing the rice in the bowl), control is seamlessly returned to π_I .

application, which relays the VR controller’s velocities to a Cartesian velocity controller on the robot. The operator controls the robot by pressing the side trigger button on the VR controller, and only actions executed while the trigger is pressed are recorded as 6D end-effector velocities. Any pauses in operation—where the trigger is released, allowing the operator to reposition their hand or stop the robot—are omitted from the recorded data. This ensures a clean dataset by removing unnecessary idle information.

Once the initial policy π_I is trained, it is deployed in the environment, where it generates 6D velocity commands for the robot. While the policy operates, a ring buffer continuously records observations. Each observation includes the end-effector camera’s RGB view and proprioceptive information such as the robot’s pose. If the operator observes the policy moving toward an undesirable state (e.g., spilling rice), they press the VR controller’s trigger button to intervene. When this occurs, the takeover actions are immediately applied to the robot, overriding the policy-generated commands.

Importantly, the contents of the ring buffer (representing the sequence of observations leading up to the intervention) are also saved as part of the new takeover demonstration. This ensures that the collected data includes context from moments just before the operator’s takeover, providing the policy with more nuanced information about failure recovery. While the operator continues teleoperating the robot, subsequent observations and actions are recorded until the operator deems the system to be in a desirable state again. At this point, releasing the trigger seamlessly hands control back to the policy, which resumes generating actions.

This targeted approach allows the operator to focus only on correcting problematic scenarios, avoiding redundant data collection for states already well-handled by the policy. The result is a compact and highly relevant dataset that incrementally addresses failure cases as they arise. A detailed schematic of this process is illustrated in Fig. 3.

Through this iterative process, our RTOT paradigm ensures that the dataset evolves to cover a wide range of scenarios, including recovery from failure cases. This targeted, incremental improvement results in a more robust and capable policy compared to traditional imitation learning approaches.

IV. EXPERIMENTAL SETUP

Our primary goal in this section is to investigate whether training visuomotor policies using real-time operator

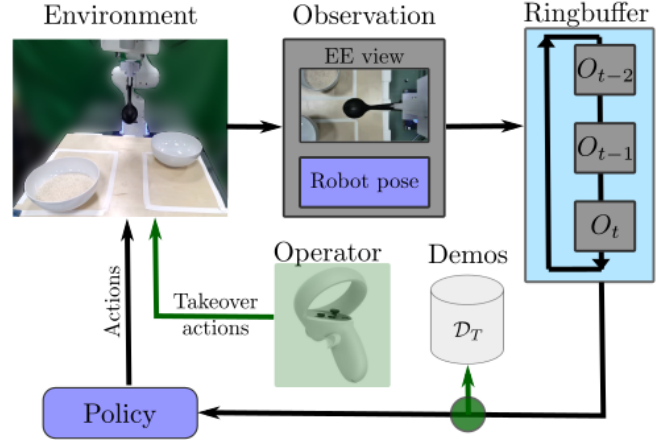


Fig. 3. Illustration of the Real-Time Takeover process: While the policy sends actions, observations (end-effector RGB view and robot pose) are continuously stored in a ring buffer. When the operator takes control using the VR controller, these ring buffer observations, along with subsequent data, are recorded as a new demonstration in \mathcal{D}_T .

takeover leads to improved task performance.

We evaluate this through a cyclic rice scooping task, as shown in Fig. 1. The objective is to scoop rice from the bowl on the left (positioned in the white rectangle) and deposit it into the bowl on the right (which can be placed anywhere within its respective rectangle). To address the cyclic nature of the task, the initial demonstrations are designed to both start and end at approximately the same position after successfully completing one scoop.

To begin, we collect an initial dataset \mathcal{D}_I^{60} consisting of 60 expert demonstrations. These demonstrations are performed while varying the locations of the rice bowls to introduce diversity in the training data. From this comprehensive dataset, we extract two smaller subsets: the first 20 demonstrations form \mathcal{D}_I^{20} , and the first 40 demonstrations form \mathcal{D}_I^{40} . Using these datasets, we train corresponding policies: π_I^{20} , π_I^{40} , and π_I^{60} .

Next, we deploy π_I^{20} in the rice scooping environment as described in Section ???. During this deployment, we collect 20 additional takeover demonstrations, denoted as \mathcal{D}_T^{20a} , while systematically varying the rice bowl positions. These takeover demonstrations are then combined with the initial dataset \mathcal{D}_I^{20} to form an augmented dataset \mathcal{D}_I^{20a} , which is used to train a new policy, π_{IT}^a .

This new policy, π_{IT}^a , is subsequently deployed again in

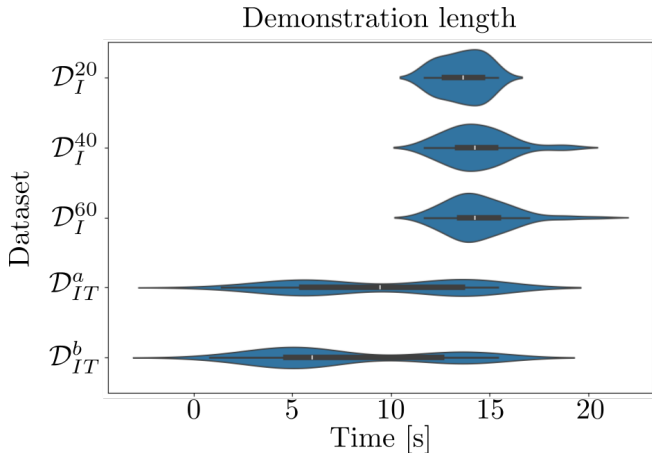


Fig. 4. Demonstration lengths for the datasets used to train the evaluated visuomotor policies. The takeover paradigm produces significantly shorter demonstrations on average, highlighting its efficiency.

the rice scooping environment to collect an additional 20 takeover demonstrations, \mathcal{D}_{IT}^{20b} . The final dataset, \mathcal{D}_{IT}^b , comprises the initial demonstrations and both sets of takeover demonstrations: $\mathcal{D}_{IT}^b = \mathcal{D}_I^{20} \cup \mathcal{D}_I^{20a} \cup \mathcal{D}_I^{20b}$. The corresponding policy trained on this dataset is the final policy evaluated in this study.

Analysis of Dataset Lengths. We analyze the lengths of the different datasets used to train the policies. Fig. 4 shows a violin plot illustrating these differences. Notably, the initial datasets \mathcal{D}_I^{20} , \mathcal{D}_I^{40} , and \mathcal{D}_I^{60} have similar mean demonstration times of 13.6s, 14.4s, and 14.5s, respectively, with total demonstration times of 272.0s, 574.3s, and 869.0s. In contrast, the takeover dataset \mathcal{D}_{IT}^a achieves a mean demonstration time of 9.5s, resulting in a total demonstration time of 377.8s, which is 34% shorter than \mathcal{D}_I^{40} despite containing the same number of demonstrations.

This efficiency is even more pronounced in the final dataset, \mathcal{D}_{IT}^b , which has the shortest mean demonstration time of 7.76s and a total demonstration time of 465.6s—a reduction of 46% compared to its non-takeover counterpart, \mathcal{D}_I^{60} . These results demonstrate the efficiency of the takeover paradigm in generating compact and targeted training datasets without sacrificing task performance.

Additionally, the combination of initial and takeover demonstrations helps to address failure cases more effectively, as takeover demonstrations specifically focus on challenging scenarios encountered during policy deployment. This iterative process not only improves policy robustness but also reduces the overall training time required to achieve high task performance.

Videos showcasing the initial demonstrations, takeover demonstrations, and task evaluations are available on the project’s website¹.

V. EXPERIMENTAL EVALUATION

We evaluate the performance of five different policies (π_I^{20} , π_I^{40} , π_{IT}^a , π_I^{60} , π_{IT}^b) on the cyclic rice scooping task. The

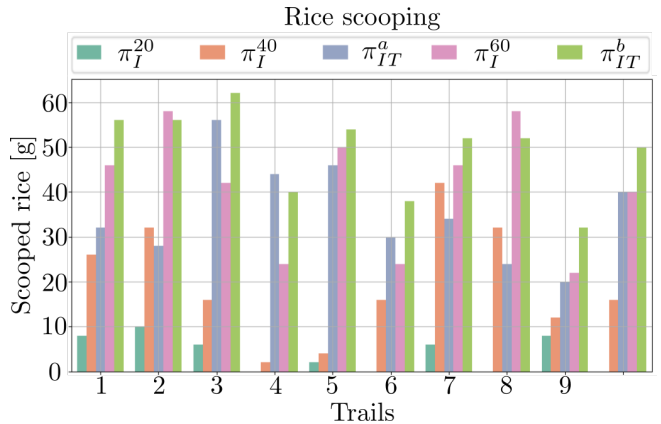


Fig. 5. Detailed results of the cyclic rice scooping experiments. The amount of rice (in grams) is shown for each of the 10 trials across all five evaluated policies.

goal of the task is to transfer as much rice as possible from a full bowl to an empty bowl within 45 seconds, measured in grams. For each policy, we conduct 10 trials, each with different placement locations for the bowls. Videos of all 50 experiments are available on the project website.

The detailed results of these experiments are shown in Fig.5, and the mean and standard deviation (std) of the scooped rice for each policy are summarized in TableV.

Model	# Demos	Mean \pm std [g]
π_I^{20}	20	4.0 \pm 4.0
π_I^{40}	40	19.8 \pm 12.9
π_{IT}^a	40	35.4 \pm 11.0
π_I^{60}	60	41.0 \pm 13.5
π_{IT}^b	60	49.2 \pm 9.4

TABLE I
MEANS AND STANDARD DEVIATIONS OF THE RICE SCOOPED IN 45 SECOND OVER THE 10 TRIALS.

Results and Analysis. The results reveal a clear trend: the performance of baseline policies improves with the number of initial demonstrations. For example, π_I^{20} , trained on only 20 demonstrations, achieved an average of 4.00 grams of rice scooped across 10 trials. Increasing the dataset to 40 demonstrations boosted the performance significantly, with π_I^{40} averaging 19.80 grams. Using all 60 initial demonstrations, π_I^{60} achieved a substantial improvement, scooping an average of 41.00 grams per trial. This highlights the importance of larger datasets for improving baseline performance in visuomotor policy training.

To address the question of whether employing the real-time operator takeover paradigm leads to better task performance, we compare the performance of the baseline π_I^{40} with that of the takeover-enhanced policy π_{IT}^a . Remarkably, π_{IT}^a achieved an average of 35.40 grams scooped, representing a 79% improvement over π_I^{40} . Similarly, the final policy, π_{IT}^b , trained on both initial and two sets of takeover demonstrations, outperformed π_I^{60} by 20%, with an average of 49.20

grams scooped per trial.

Efficiency of Takeover Demonstrations. These results are particularly noteworthy given that the total demonstration time for the takeover datasets is significantly lower. For example, the total demonstration time for \mathcal{D}_{IT}^b is 465.6 seconds, which is 46% shorter than the total time for \mathcal{D}^{60} (869.0 seconds). Despite this reduction, π_{IT}^b outperforms the baseline trained on the full dataset of initial demonstrations. This indicates that the targeted nature of takeover demonstrations, focusing on failure cases, provides more valuable information for policy training compared to redundant initial demonstrations.

Key Insights. These findings highlight two important aspects. First, redundant demonstrations contribute only marginally to further performance improvement once the policy has mastered the easier portions of the task. Second, it is extremely challenging to anticipate all potential failure cases during the initial demonstration phase. The real-time operator takeover paradigm addresses these challenges by iteratively improving the policy. By taking control only when the system encounters undesirable states, the operator provides precise, context-specific demonstrations that were missing from the initial dataset.

VI. TO TAKEOVER OR NOT TO TAKEOVER

In the current framework, the decision of when to perform a takeover is left entirely to the operator. We argue that this is a sensible approach, given that the quality of the initial demonstrations—and more broadly, the overall task performance—is highly influenced by the quality of these demonstrations [5]. However, having a more systematic way of assessing how close the robot is to undesirable states, or detecting when such states occur, could provide significant benefits.

Recall that we assume all expert demonstrations represent desirable states. The Denoising Diffusion Probabilistic Model (DDPM) employed here is trained to minimize the KL-divergence between the conditional distribution $p(A_t|O_t)$ —where actions A and observations O are derived from the expert demonstrations—and the distribution of samples generated by the DDPM. Intuitively, this means the policy struggles to produce meaningful actions if the current observations (or more precisely, the embeddings of those observations) differ significantly from the data used in training. Improving the robustness of visuomotor diffusion policies against such variations in observations remains an active area of research [6], [16].

This motivates the need for a metric that evaluates how far observations obtained during inference deviate from the training distribution. Such a metric could serve as an out-of-distribution (OOD) measure to better inform intervention decisions.

A. Mahalanobis Distance as an OOD Measure

The Mahalanobis distance provides a statistical measure of the distance between a point and a distribution, accounting for correlations between variables. It is computed as:

$$d_M = \sqrt{(x - \mu)^\top \Sigma^{-1} (x - \mu)} \quad (1)$$

Here, μ is the mean vector of the distribution, Σ is the covariance matrix, and x is the queried data point.

Given the high dimensionality of the RGB images, we instead compute embeddings of the images concatenated with the robot pose. Each observation is encoded into a 1056-dimensional vector (two 512-dimensional embeddings for the images and two 16-dimensional embeddings for the pose). These embeddings serve as the conditioning input for the diffusion model.

To evaluate the Mahalanobis distance, we first encode all datasets \mathcal{D}^{20} , \mathcal{D}^{40} , \mathcal{D}^{60} , \mathcal{D}_{IT}^a , and \mathcal{D}_{IT}^b using their respective trained models (π_I^{20} , π_I^{40} , π_I^{60} , π_{IT}^a , and π_{IT}^b). This produces embeddings Z_I^{20} , Z_I^{40} , Z_I^{60} , Z_{IT}^a , and Z_{IT}^b . Similarly, all observations recorded during the experiments (\mathcal{E}) are encoded using these models, resulting in H_I^{20} , H_I^{40} , H_I^{60} , H_{IT}^a , and H_{IT}^b .

We estimate the mean (μ) and covariance matrix (Σ) of each embedding set using the Minimum Covariance Determinant Estimator implemented in scikit-learn [59]. For each encoded observation $h \in H$, the Mahalanobis distance d_M is calculated using Algorithm 1.

Algorithm 1 Mahalanobis Distance Calculation

- 1: **Input:** Datasets \mathcal{D}_I^{20} , \mathcal{D}_I^{40} , \mathcal{D}_I^{60} , \mathcal{D}_{IT}^a , \mathcal{D}_{IT}^b , trained models π_I^{20} , π_I^{40} , π_I^{60} , π_{IT}^a , π_{IT}^b , experimental observations \mathcal{E} .
- 2: **Output:** Mahalanobis distances d_m for each observation encoding $h \in H$ with respect to each embedding set.
- 3: **Step 1: Encode Dataset Embeddings**
- 4:

$$Z_I^{20} = \pi_I^{20}(\mathcal{D}_I^{20}), \quad Z_I^{40} = \pi_I^{40}(\mathcal{D}_I^{40}), \quad Z_I^{60} = \pi_I^{60}(\mathcal{D}_I^{60}), \\ Z_{IT}^a = \pi_{IT}^a(\mathcal{D}_{IT}^a), \quad Z_{IT}^b = \pi_{IT}^b(\mathcal{D}_{IT}^b)$$

- 5: **Step 2: Encode Experimental Observations**
- 6:

$$H_I^{20} = \pi_I^{20}(\mathcal{E}), \quad H_I^{40} = \pi_I^{40}(\mathcal{E}), \quad H_I^{60} = \pi_I^{60}(\mathcal{E}), \\ H_{IT}^a = \pi_{IT}^a(\mathcal{E}), \quad H_{IT}^b = \pi_{IT}^b(\mathcal{E})$$

- 7: **Step 3: Calculate Mahalanobis Distances**
 - 8: **for** each $Z \in \{Z_I^{20}, Z_I^{40}, Z_I^{60}, Z_{IT}^a, Z_{IT}^b\}$ and $H \in \{H_I^{20}, H_I^{40}, H_I^{60}, H_{IT}^a, H_{IT}^b\}$ **do**
 - 9: $\mu_Z, \Sigma_Z = \text{MinCovDet}(Z)$
 - 10: **for** each encoding $h \in H$ **do**
 - 11: $d_m(h, \mu_Z, \Sigma_Z) = \sqrt{(h - \mu_Z)^\top \Sigma_Z^{-1} (h - \mu_Z)}$
 - 12: **end for**
 - 13: **end for**
-

B. Mahalanobis Distance Analysis

The summarized results of Algorithm 1 are presented as violin plots in Fig. 6.

We observe that \mathcal{D}_I^{20} , the dataset with the highest average Mahalanobis distance, corresponds to the worst-performing

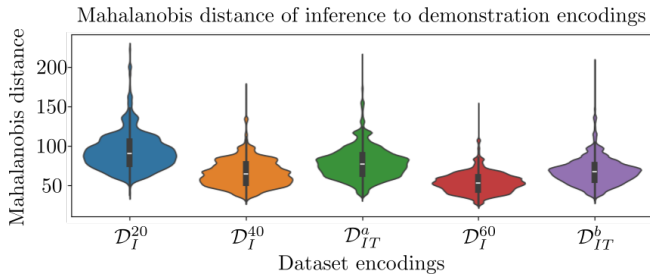


Fig. 6. Violin plots of Mahalanobis distances for experimental observations compared to the learned distribution for each dataset.

policy, π_I^{20} . This suggests that the Mahalanobis distance provides some indication of how far an observation during rollout deviates from the training distribution. Adding more demonstrations to the training set reduces the average Mahalanobis distance of observations gathered during experiments.

However, an interesting anomaly is observed when comparing \mathcal{D}_I^{40} to \mathcal{D}_{IT}^a or \mathcal{D}_I^{60} to \mathcal{D}_{IT}^b : the mean Mahalanobis distance does not directly correlate with task performance. For example, π_{IT}^b outperforms π_I^{60} in scooping rice but has a similar Mahalanobis distance profile.

Case Study: Trial 3. To explore this further, we analyze a specific trial in Fig. 7. For Trial number three, we compare the Mahalanobis distances of policies π_I^{20} , π_{IT}^b , and π_I^{60} . We observe that π_I^{20} starts off with the highest distance while π_{IT}^b , and π_I^{60} are both around 40. The better in task performing π_{IT}^b however then exhibits a slightly higher distance as the trial goes on however none of the states encountered would need require the takeover of the operator and the distance stays below 100 for the entire duration of the trial.

The poorly performing policy π_I^{20} exhibits the highest Mahalanobis distance peaking at around 160, around the 25s mark, when the robot’s spoon moves outside the target bowl and fails to recover. The robot remains stuck in this state until the trial ends, highlighting the relationship between high Mahalanobis distance and undesirable states. In contrast, both π_I^{60} and π_{IT}^b maintain relatively lower Mahalanobis distances throughout the trial, with no significant increases. While π_{IT}^b outperforms π_I^{60} in task performance, the Mahalanobis distance does not differentiate much between the two policies in this regard. We also observe that the sudden stops and therefore camera shaking have significantly influence as can be observed on the last peak in π_{IT}^b (orange), and π_I^{60} (green) around the 45s mark which was introduced by the sudden stop of the robot as the experiment was concluded. Furthermore, we see that the distance remains constant as long as no changes in the image are introduced via motion.

Key Insights While the Mahalanobis distance may not perfectly predict task performance (e.g., rice scooped), it successfully identifies undesirable states that lead to performance degradation, such as the robot becoming stuck. This insight highlights the potential for developing automated recovery strategies based on significant increases in

Mahalanobis distance. Additionally, these findings provide a deeper understanding of the role of demonstrations in improving policy robustness, underscoring the value of diverse and targeted training datasets.

VII. CONCLUSION

In this work, we introduced a real-time operator takeover (RTOT) paradigm for training visuomotor diffusion policies. This framework empowers users to seamlessly take control of the robot from a suboptimal policy, preventing failures during execution. These takeover demonstrations are subsequently integrated into the training process, iteratively refining the policy until the desired performance is achieved.

Our experimental evaluation demonstrates the effectiveness of this approach. Despite the total time of takeover demonstrations being significantly shorter than that of the initial demonstrations, policies trained on these targeted datasets consistently outperform their counterparts. In a cyclic rice scooping task, policies enhanced with takeover demonstrations achieved superior task performance, highlighting the value of addressing failure cases during policy deployment rather than relying solely on extensive initial datasets.

Beyond showcasing the utility of the RTOT framework, we also explored the use of the Mahalanobis distance as a tool for detecting out-of-distribution (OOD) states during inference. Our analysis reveals that while the Mahalanobis distance can successfully identify undesirable states, such as those that lead to the robot becoming stuck, it is not an exact predictor of final task performance. In particular, although the Mahalanobis distance helps measure deviations from the training distribution, it does not always correlate directly with the quantity of rice scooped in our experiments. Nevertheless, these insights underline the potential of the Mahalanobis distance for enhancing policy robustness by flagging critical states during deployment.

Contributions and Insights.

Real-Time Operator Takeover Paradigm: We demonstrated how the RTOT framework can improve policy performance by addressing failure scenarios iteratively. The targeted nature of takeover demonstrations provides valuable training data, focusing specifically on challenging states that are often missing from initial demonstrations.

Efficiency of Takeover Demonstrations: Our results highlight the efficiency of this approach, with takeover-enhanced policies achieving better performance despite using significantly less total demonstration time. This finding underscores the importance of quality and relevance in training data over sheer quantity.

Out-of-Distribution Detection: We provided a detailed analysis of the Mahalanobis distance as an OOD measure, showing its utility in detecting undesirable states during policy execution. This metric can serve as a foundation for future automated recovery mechanisms in robot control systems.

In conclusion, the RTOT paradigm represents a significant step toward more efficient and adaptive robot learning systems. By enabling real-time human intervention and

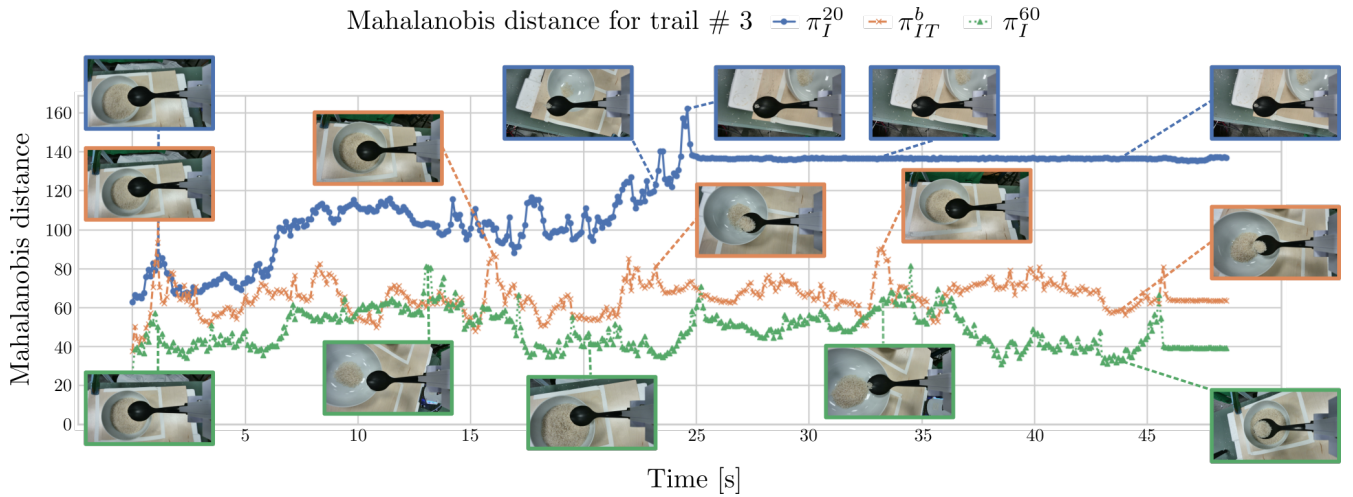


Fig. 7. The Mahalanobis distance of the embedded observations of trail #3 of the models π_I^{20} , π_I^b , and π_I^{60} , with the RGB views shown for specific timesteps.

leveraging targeted demonstrations, this approach provides a practical and scalable solution for addressing the inherent challenges of visuomotor policy training, particularly in complex and unpredictable task environments.

REFERENCES

- [1] Z. Fu, T. Z. Zhao, and C. Finn, "Mobile aloha: Learning bimanual mobile manipulation with low-cost whole-body teleoperation," in *arXiv*, 2024.
- [2] N. Ingelhart, J. Munkeby, J. van Haastregt, A. Varava, M. C. Welle, and D. Kragic, "A robotic skill learning system built upon diffusion policies and foundation models," *arXiv preprint arXiv:2403.16730*, 2024.
- [3] C. Chi, S. Feng, Y. Du, Z. Xu, E. Cousineau, B. Burchfiel, and S. Song, "Diffusion policy: Visuomotor policy learning via action diffusion," in *Proceedings of Robotics: Science and Systems (RSS)*, 2023.
- [4] T. Z. Zhao, V. Kumar, S. Levine, and C. Finn, *Learning fine-grained bimanual manipulation with low-cost hardware*, 2023. arXiv: 2304.13705 [cs.RO]. [Online]. Available: <https://arxiv.org/abs/2304.13705>.
- [5] J. Pari, N. M. Shafiqullah, S. P. Arunachalam, and L. Pinto, "The surprising effectiveness of representation learning for visual imitation," *arXiv preprint arXiv:2112.01511*, 2021.
- [6] Y. Wang, M. Dong, B. Du, and C. Xu, "Imitation learning from purified demonstration," *arXiv preprint arXiv:2310.07143*, 2023.
- [7] M. Moletta, M. K. Wozniak, M. C. Welle, and D. Kragic, "A virtual reality framework for human-robot collaboration in cloth folding," in *2023 IEEE-RAS 22nd International Conference on Humanoid Robots (Humanoids)*, IEEE, 2023, pp. 1–7.
- [8] M. C. Welle, N. Ingelhart, M. Lippi, M. Wozniak, A. Gasparri, and D. Kragic, "Quest2ros: An app to facilitate teleoperating robots," in *7th International Workshop on Virtual, Augmented, and Mixed-Reality for Human-Robot Interactions*, 2024.
- [9] A. Iyer, Z. Peng, Y. Dai, I. Guzey, S. Haldar, S. Chintala, and L. Pinto, "Open teach: A versatile teleoperation system for robotic manipulation," *arXiv preprint arXiv:2403.07870*, 2024.
- [10] 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.
- [11] J. van Haastregt, M. C. Welle, Y. Zhang, and D. Kragic, "Puppeteer your robot: Augmented reality leader-follower teleoperation," *arXiv preprint arXiv:2407.11741*, 2024.
- [12] S. Chen, C. Wang, K. Nguyen, L. Fei-Fei, and C. K. Liu, "Arcap: Collecting high-quality human demonstrations for robot learning with augmented reality feedback," *arXiv preprint arXiv:2410.08464*, 2024.
- [13] K. Shaw, Y. Li, J. Yang, M. K. Srirama, R. Liu, H. Xiong, R. Mendonca, and D. Pathak, "Bimanual dexterity for complex tasks," in *8th Annual Conference on Robot Learning*, 2024.
- [14] S. Yang, M. Liu, Y. Qin, R. Ding, J. Li, X. Cheng, R. Yang, S. Yi, and X. Wang, "Ace: A cross-platform visual-exoskeletons system for low-cost dexterous teleoperation," *arXiv preprint arXiv:2408.11805*, 2024.
- [15] N. Hansen and X. Wang, "Generalization in reinforcement learning by soft data augmentation," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2021, pp. 13 611–13 617.
- [16] Z. Zhuang, R. Wang, N. Ingelhart, V. Kyrki, and D. Kragic, "Enhancing visual domain robustness in behaviour cloning via saliency-guided augmentation," in *8th Annual Conference on Robot Learning*, 2024. [Online]. Available: <https://openreview.net/forum?id=CskuWHDBAr>.
- [17] J. Ho, A. Jain, and P. Abbeel, "Denosing diffusion probabilistic models," in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33, Curran Associates, Inc., 2020, pp. 6840–6851. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.
- [18] P. Li, Z. Li, H. Zhang, and J. Bian, "On the generalization properties of diffusion models," in *Advances in Neural Information Processing Systems*, A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, Eds., vol. 36, Curran Associates, Inc., 2023, pp. 2097–2127. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/file/06abed94583030dd50abe6767bd643b1-Paper-Conference.pdf.
- [19] I. Kapelyukh, V. Vosylius, and E. Johns, "Dall-e-bot: Introducing web-scale diffusion models to robotics," *IEEE Robotics and Automation Letters*, vol. 8, no. 7, pp. 3956–3963, 2023. DOI: 10.1109/LRA.2023.3272516.
- [20] U. A. Mishra, S. Xue, Y. Chen, and D. Xu, "Generative skill chaining: Long-horizon skill planning with diffusion models," in *Proceedings of The 7th Conference on Robot Learning*, J. Tan, M. Toussaint, and K. Darvish, Eds., ser. Proceedings of Machine Learning Research, vol. 229, PMLR, Jun. 2023, pp. 2905–2925. [Online]. Available: <https://proceedings.mlr.press/v229/mishra23a.html>.
- [21] A. Sridhar, D. Shah, C. Glossop, and S. Levine, "Nomad: Goal masked diffusion policies for navigation and exploration," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 63–70. DOI: 10.1109/ICRA57147.2024.10610665.
- [22] U. A. Mishra and Y. Chen, "Reorientdiff: Diffusion model based reorientation for object manipulation," in *2024 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2024, pp. 10 867–10 873.
- [23] M. Reuss, M. Li, X. Jia, and R. Lioutikov, "Goal-conditioned imitation learning using score-based diffusion policies," *arXiv preprint arXiv:2304.02532*, 2023.
- [24] P. M. Scheikl, N. Schreiber, C. Haas, N. Freymuth, G. Neumann, R. Lioutikov, and F. Mathis-Ullrich, "Movement primitive diffusion: Learning gentle robotic manipulation of deformable objects," *IEEE Robotics and Automation Letters*, 2024.
- [25] D. A. Pomerleau, "Alvinn: An autonomous land vehicle in a neural network," in *Advances in Neural Information Processing Systems*, D. Touretzky, Ed., vol. 1, Morgan-Kaufmann, 1988. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/1988/file/812b4ba287f5ee0bc9d43bbf5bbe87fb-Paper.pdf.
- [26] A. Mandlkar, 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*, 2021. arXiv: 2108.03298 [cs.RO]. [Online]. Available: <https://arxiv.org/abs/2108.03298>.
- [27] N. M. Shafiqullah, Z. Cui, A. (Altanzaya, and L. Pinto, "Behavior transformers: Cloning k modes with one stone," in *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., vol. 35, Curran Associates, Inc., 2022, pp. 22 955–22 968. [Online]. Available: <https://>

- proceedings.neurips.cc/paper_files/paper/2022/file/90d17e882adbda42349db6f50123817 - Paper - Conference .pdf.
- [28] 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. J. Joshi, R. Julian, D. Kalashnikov, Y. Kuang, *et al.*, *Rt-1: Robotics transformer for real-world control at scale*, 2023. arXiv: 2212.06817 [cs.LG]. [Online]. Available: <https://arxiv.org/abs/2212.06817>.
- [29] 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*, A. Faust, D. Hsu, and G. Neumann, Eds., ser. Proceedings of Machine Learning Research, vol. 164, PMLR, Aug. 2022, pp. 991–1002. [Online]. Available: <https://proceedings.mlr.press/v164/jang22a.html>.
- [30] P. Florence, C. Lynch, A. Zeng, O. A. Ramirez, A. Wahid, L. Downs, A. Wong, J. Lee, I. Mordatch, and J. Tompson, “Implicit behavioral cloning,” in *Proceedings of the 5th Conference on Robot Learning*, A. Faust, D. Hsu, and G. Neumann, Eds., ser. Proceedings of Machine Learning Research, vol. 164, PMLR, Aug. 2022, pp. 158–168. [Online]. Available: <https://proceedings.mlr.press/v164/florence22a.html>.
- [31] Y. Duan, M. Andrychowicz, B. Stadie, O. Jonathan Ho, J. Schneider, I. Sutskever, P. Abbeel, and W. Zaremba, “One-shot imitation learning,” in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30, Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/ba3866600c3540f67c1e9575e213be0a-Paper.pdf.
- [32] S. James, M. Bloesch, and A. J. Davison, “Task-embedded control networks for few-shot imitation learning,” in *Proceedings of The 2nd Conference on Robot Learning*, A. Billard, A. Dragan, J. Peters, and J. Morimoto, Eds., ser. Proceedings of Machine Learning Research, vol. 87, PMLR, 29–31 Oct 2018, pp. 783–795. [Online]. Available: <https://proceedings.mlr.press/v87/james18a.html>.
- [33] S. Dasari and A. Gupta, “Transformers for one-shot visual imitation,” in *Proceedings of the 2020 Conference on Robot Learning*, J. Kober, F. Ramos, and C. Tomlin, Eds., ser. Proceedings of Machine Learning Research, vol. 155, PMLR, 16–18 Nov 2021, pp. 2071–2084. [Online]. Available: <https://proceedings.mlr.press/v155/dasari21a.html>.
- [34] M. Shridhar, L. Manuelli, and D. Fox, “Perceiver-actor: A multi-task transformer for robotic manipulation,” in *Proceedings of The 6th Conference on Robot Learning*, K. Liu, D. Kulic, and J. Ichnowski, Eds., ser. Proceedings of Machine Learning Research, vol. 205, PMLR, 14–18 Dec 2023, pp. 785–799. [Online]. Available: <https://proceedings.mlr.press/v205/shridhar23a.html>.
- [35] H. Kim, Y. Ohmura, and Y. Kuniyoshi, “Goal-conditioned dual-action imitation learning for dexterous dual-arm robot manipulation,” *IEEE Transactions on Robotics*, vol. 40, pp. 2287–2305, 2024. DOI: 10.1109/TRO.2024.3372778.
- [36] S. Ross and D. Bagnell, “Efficient reductions for imitation learning,” in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Y. W. Teh and M. Titterton, Eds., ser. Proceedings of Machine Learning Research, vol. 9, Chia Laguna Resort, Sardinia, Italy: PMLR, 13–15 May 2010, pp. 661–668. [Online]. Available: <https://proceedings.mlr.press/v9/ross10a.html>.
- [37] S. Ross, G. Gordon, and D. Bagnell, “A reduction of imitation learning and structured prediction to no-regret online learning,” in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, G. Gordon, D. Dunson, and M. Dudík, Eds., ser. Proceedings of Machine Learning Research, vol. 15, Fort Lauderdale, FL, USA: PMLR, Nov. 2011, pp. 627–635. [Online]. Available: <https://proceedings.mlr.press/v15/ross11a.html>.
- [38] S. Tu, A. Robey, T. Zhang, and N. Matni, “On the sample complexity of stability constrained imitation learning,” in *Proceedings of The 4th Annual Learning for Dynamics and Control Conference*, R. Firoozi, N. Mehr, E. Yel, R. Antonova, J. Bohg, M. Schwager, and M. Kochenderfer, Eds., ser. Proceedings of Machine Learning Research, vol. 168, PMLR, 23–24 Jun 2022, pp. 180–191. [Online]. Available: <https://proceedings.mlr.press/v168/tu22a.html>.
- [39] M. Kelly, C. Sidrane, K. Driggs-Campbell, and M. J. Kochenderfer, “Hgdagger: Interactive imitation learning with human experts,” in *2019 International Conference on Robotics and Automation (ICRA)*, 2019, pp. 8077–8083. DOI: 10.1109/ICRA.2019.8793698.
- [40] K. Menda, K. Driggs-Campbell, and M. J. Kochenderfer, “Ensembledagger: A bayesian approach to safe imitation learning,” in *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019, pp. 5041–5048. DOI: 10.1109/IROS40897.2019.8968287.
- [41] A. Kumar, A. Zhou, G. Tucker, and S. Levine, “Conservative q-learning for offline reinforcement learning,” in *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds., vol. 33, Curran Associates, Inc., 2020, pp. 1179–1191. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2020/file/0d2b2061826a5df3221116a5085a6052 - Paper .pdf.
- [42] T. Yu, A. Kumar, R. Rafailov, A. Rajeswaran, S. Levine, and C. Finn, “Combo: Conservative offline model-based policy optimization,” in *Advances in Neural Information Processing Systems*, M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, Eds., vol. 34, Curran Associates, Inc., 2021, pp. 28954–28967. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2021/file/f29a179746902e331572c483c45e5086-Paper.pdf.
- [43] P. Florence, L. Manuelli, and R. Tedrake, “Self-supervised correspondence in visuomotor policy learning,” *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 492–499, 2020. DOI: 10.1109/LRA.2019.2956365.
- [44] L. Ke, J. Wang, T. Bhattacharjee, B. Boots, and S. Srinivasa, “Grasping with chopsticks: Combating covariate shift in model-free imitation learning for fine manipulation,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 2021, pp. 6185–6191. DOI: 10.1109/ICRA48506.2021.9561662.
- [45] A. Zhou, M. J. Kim, L. Wang, P. Florence, and C. Finn, “Nerf in the palm of your hand: Corrective augmentation for robotics via novel-view synthesis,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2023, pp. 17907–17917.
- [46] A. Reichlin, G. L. Marchetti, H. Yin, A. Ghadirzadeh, and D. Kragic, *Back to the manifold: Recovering from out-of-distribution states*, 2022. arXiv: 2207.08673 [cs.LG]. [Online]. Available: <https://arxiv.org/abs/2207.08673>.
- [47] J. Yang, K. Zhou, Y. Li, and Z. Liu, “Generalized out-of-distribution detection: A survey,” *International Journal of Computer Vision*, vol. 132, no. 12, pp. 5635–5662, 2024.
- [48] R. Sinha, A. Sharma, S. Banerjee, T. Lew, R. Luo, S. M. Richards, Y. Sun, E. Schmerling, and M. Pavone, “A system-level view on out-of-distribution data in robotics,” *arXiv preprint arXiv:2212.14020*, 2022.
- [49] A. Obaigbena, O. A. Lottu, E. D. Ugwuanyi, B. S. Jacks, E. O. Sodiya, and O. D. Daraojimba, “Ai and human-robot interaction: A review of recent advances and challenges,” *GSC Advanced Research and Reviews*, vol. 18, no. 2, pp. 321–330, 2024.
- [50] T. Wang, P. Zheng, S. Li, and L. Wang, “Multimodal human–robot interaction for human-centric smart manufacturing: A survey,” *Advanced Intelligent Systems*, vol. 6, no. 3, p. 2300359, 2024.
- [51] K. Darvish, L. Penco, J. Ramos, R. Cisneros, J. Pratt, E. Yoshida, S. Ivaldi, and D. Pucci, “Teleoperation of humanoid robots: A survey,” *IEEE Transactions on Robotics*, 2023.
- [52] Z. Makhataeva and H. A. Varol, “Augmented reality for robotics: A review,” *Robotics*, vol. 9, no. 2, p. 21, 2020.
- [53] C. Barentine, A. McNay, R. Pfaffenbichler, A. Smith, E. Rosen, and E. Phillips, “A vr teleoperation suite with manipulation assist,” in *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 2021, pp. 442–446.
- [54] S. Xu, S. Moore, and A. Cosgun, “Shared-control robotic manipulation in virtual reality,” *arXiv preprint arXiv:2205.10564*, 2022.
- [55] V. Ortenzi, M. Filipovica, D. Abdulkarim, T. Pardi, C. Takahashi, A. Wing, M. Di Luca, and K. J. Kuchenbecker, “Robot, pass me the tool: Handle visibility facilitates task-oriented handovers,” in *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2022, pp. 1–9.
- [56] K. Chandan, V. Kudalkar, X. Li, and S. Zhang, “Arroch: Augmented reality for robots collaborating with a human,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2021, pp. 3787–3793.
- [57] P. Lindemann, N. Müller, and G. Rigoll, “Exploring the use of augmented reality interfaces for driver assistance in short-notice takeovers,” in *2019 IEEE Intelligent Vehicles Symposium (IV)*, 2019, pp. 804–809. DOI: 10.1109/IVS.2019.8814237.
- [58] W. Ma, A. Duan, H.-Y. Lee, P. Zheng, and D. Navarro-Alarcon, “Human-aware reactive task planning of sequential robotic manipulation tasks,” *IEEE Transactions on Industrial Informatics*, pp. 1–10, 2025. DOI: 10.1109/TII.2024.3514130.
- [59] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Courneau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.