Learning Robotic Arm Control From Demonstration: An Exploration and Comparative Study of Imitation-Based Learning Approaches For Robotic Manipulation Task in a Webots Environment

Jelbert C. Zamora
Department of Computer Science, College of Computing and Information Sciences, Caraga State University - Main Campus

Rodrigo M. Raga Jr.
Department of Computer Science, College of Computing and Information Sciences, Caraga State University - Main Campus

Rudolph Joshua U. Candare
Department of Computer Science, College of Computing and Information Sciences, Caraga State University - Main Campus

https://doi.org/10.5109/7395601

出版情報：Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES). 11, pp.775-782, 2025-10-30. International Exchange and Innovation Conference on Engineering & Sciences
バージョン：

Learning Robotic Arm Control From Demonstration: An Exploration and Comparative Study of Imitation-Based Learning Approaches For Robotic Manipulation Task in a Webots Environment

Jelbert C. Zamora[1], Rodrigo M. Raga Jr.[1], Rudolph Joshua U. Candare[1]
[1] Department of Computer Science, Caraga State University - Main Campus, Butuan City, Philippines
rucandare@carsu.edu.ph

Abstract: *This study explores applying Generative Adversarial Imitation Learning (GAIL) to train a 5-DOF robotic arm for grasping and pick-and-place tasks in a Webots simulation. A custom dataset of ~141,000 state-action pairs from 500 expert demonstrations was collected via teleoperation. GAIL's performance was compared to a Behavioral Cloning (BC) baseline. While neither achieved full task success (0%), GAIL demonstrated superior proximity performance, achieving consistently lower median minimum gripper-object distances (0.129m-0.133m) than BC (0.232m-0.269m), a statistically significant improvement (42.7%-50.5%). Counter-intuitively, BC showed higher success at a 0.10m grasping threshold initially but degraded, while GAIL slightly improved. Neither consistently reached the optimal 0.06m threshold. Qualitative analysis revealed GAIL exhibiting emergent manipulation sub-behaviors not seen in BC, highlighting its potential despite quantitative limitations.*

Keywords: Generative Adversarial Imitation Learning (GAIL); Behavior Cloning (BC); Robot arm; Imitation Learning; Webots Simulation

1. INTRODUCTION

This study aims to explore and experiment with the Generative Adversarial Imitation Learning (GAIL) algorithm for training a robotic arm agent to learn robotic manipulation tasks from a demonstration inside a custom simulation. GAIL is an imitation learning framework that combines principles from generative adversarial networks (GANs) and reinforcement learning (RL) to replicate expert behavior without requiring an explicit reward function [1]. The researchers identified a gap in the development of manipulation models for robotic arms, particularly in their ability to perform complex tasks in dynamic environments [10][13]. There's an area to improve the generalizability of robotic learning systems, as existing models struggle to adapt to variations in object properties and unforeseen conditions [2].

The study aims to contribute to the exploration of robotic arm control by leveraging imitation learning approaches and analyzing observed comparisons. The researchers will use teleoperation techniques to produce a custom state-action pairs dataset encompassing robotic arm actions from approaching, grasping, lifting, transporting, and placing the object that serves as the expert policy demonstration. Teleoperation technique enhances the quality of collected data and improves the robot's learning efficiency by allowing seamless human intervention during task execution [3].

Developing sophisticated control strategies for dexterous robotic manipulation often necessitates the use of simulation environments. These environments offer a safe and controlled setting for training agents, allowing for the rapid collection of large datasets and iterative testing of algorithms. Recent research increasingly leverages physics-driven simulations to train robots for complex interaction tasks like Gazebo [4]. The study demonstrates that learning architectures can acquire manipulation skills within simulated settings with reduced reliance in extensive hard-coded behaviors [4].

Several limitations are acknowledged, including the computational demands of training purely standalone GAIL model on high-dimensional data, the risk of overfitting due to the specificity and richness of the collected dataset. The success rate is determined by the potential result of experimenting with the robotic agent to accomplish evaluation methods, specifically measuring the ability to approach objects within the defined threshold (proximity), initiate grasping actions, and complete placement operations within the target location.

2. MATERIALS AND METHODS

The researchers collaborated with the AI and Robotics Laboratory former MineGears at Caraga State University - Main Campus to explore this domain by providing the simulation environment of their robot arm agent.

By utilizing GAIL alone in this study, studies have explored its use in generating human-like movement patterns and efficacy as learning algorithm (e.g. [14][15]).

The study uses a three-stage process: collecting datasets inside the simulation environment using a joystick, training the agent using GAIL, and comparing its performance to Behavior Cloning (BC), which employs offline training through supervised learning.

The simulation environment is consistent across all stages, ensuring that the agent's learned behaviors are tested under the same conditions in which the data was originally collected, and the policy was trained.

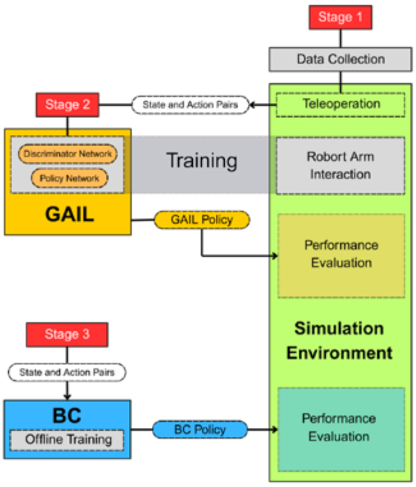

Fig. 1. Methodological flow of the study

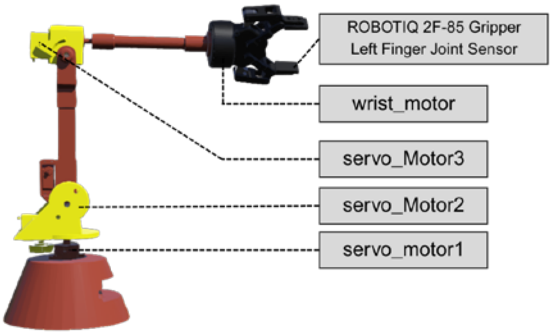

Fig. 2. Mine-Gears Robotic Arm Components

The research utilizes a predefined 5 Degree of Freedom (DOF) robotic manipulator equipped with an end-effector designed specifically for grasping operations. The kinematic chain comprises articulated joints such as Base Rotation Joint, Shoulder Joint, Elbow Joint, Wrist Joint, and End-Effector. Data acquisition and analysis are based on expert demonstrations collected in a robotic arm simulation environment using Webots. Webots as a suitable environment is supported with the study in [8].

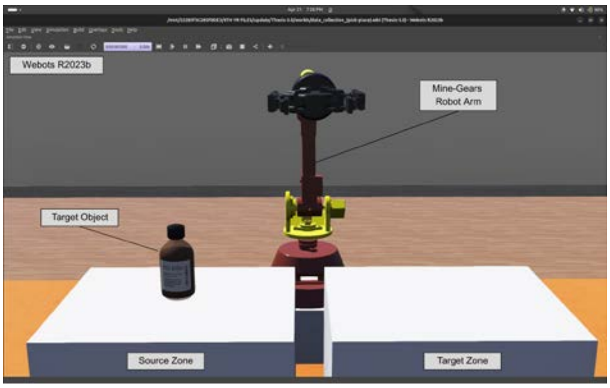

Fig. 3. Simulation Environment of the study (Webots)

The simulation environment was developed using Webots, including a 5-DOF robotic arm, a movable object initialized in randomized positions, a source zone, and a target placement zone with fixed spatial bounds. Success detection logic was built directly into the simulation to validate grasping and placement based on distance thresholds and target zone bounding conditions.

2.2 Custom Data Collection

500 demonstrations were generated using joystick teleoperation by a human operator, recording the actions as control signals to the agent's joints as continuous values and the gripper as discrete values (shown in Fig. 4. (a)(b)). During the process, the trajectory is automatically replayed for validation, allowing experts to verify task success and reject any incomplete or erroneous demonstrations. Success criteria are defined based on grasp success and placement success, with success labels generated during both collection and replay. TensorBoard is utilized to produce graphical results (shown in Fig. 4. (c)) of each trajectory collected if the expected performance has been fully recorded and if the success criteria defined in the simulation are recorded as well, rather than using the eye only for quality check.

Fig. 4. (a) Perform Approach, Grasp, Lift via Joystick

Fig. 4. (b) Perform Transport and Place via Joystick

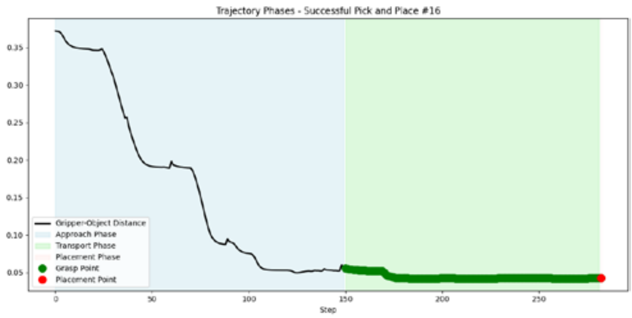

Fig. 4. (c) Recorded trajectory of a confirmed dataset

Corresponding environmental states were captured simultaneously, including the agent's joint angles, object position, and relative distances between the gripper and the object, forming a comprehensive 23-dimensional state vector and 5 action spaces. The trajectories collected are saved in .npz file format, compatible for training with imitation learning algorithms like Generative Adversarial Imitation Learning (GAIL) and Behavior Cloning (BC)

The dataset used in the study is structured and contains multiple arrays representing various aspects of the robot-object interaction during the task execution. These arrays are shown in Table 1.

Table 1. Components of each trajectory across all 500 confirmed high-quality dataset.

Count	Array Name
1	states
2	actions
3	rewards
4	initial_object_position
5	initial_object_orientation
6	object_positions
7	object_orientations
8	ee_positions
9	ee_orientations
10	dones
11	total_rewards
12	is_success
13	is_expert_verified
14	grasp_timestamps
15	placement_timestamps

In this dataset, only the states and actions are used directly for training the model, as they represent the observations inside the simulation environment and the corresponding actions executed. These two are the core components for imitation learning like GAIL. The other data, such as object positions, end-effector movements, success flags, and timestamps, are considered auxiliary metadata. These are helpful for analyzing behavior, evaluating performance, and debugging, but they don't influence the learning process itself.

2.3 Generative Adversarial Imitation Learning (GAIL) Training

Before inputting the collected data into the training pipeline, we applied a series of preprocessing steps to ensure that the state and action data were appropriately scaled and normalized. This preprocessing was essential for optimizing the training process and ensuring that the machine learning model could learn efficiently from the expert demonstrations.

To preprocess the state space, we used scikit-learn's StandardScaler to standardize the data. The StandardScaler transforms each feature in the state space such that it has a mean of 0 and a standard deviation of 1. This step is particularly important as it ensures that all features, regardless of their original scales or units, contribute equally to the model.

The study focuses on the exploration of a machine learning model for robotic manipulation tasks using a Generative Adversarial Imitation Learning (GAIL) architecture. The model involves two primary neural network models: the policy network, acting as the generator, which is optimized using Poximal Policy Optimization (PPO) [11] and the discriminator network, providing the learning signal. The policy network generates actions based on the current state of the environment, while the discriminator network functions as a learned reward mechanism. The policy network receives a 23-dimensional state vector, while the discriminator network classifies whether a given state-action pair originates from the expert demonstration dataset or from the policy's current behavior.

The training process follows a back-and-forth loop where the policy and discriminator networks are updated one after the other. This cycle repeats until the training reaches a set number of steps or the model stops improving. Along the way, the system regularly checks how well the model is learning by running evaluations at specific time intervals for debugging and script enhancement purposes. This training takes place inside a simulation environment, where the robot arm interacts with a medicine bottle as the target object. Each step of learning involves the agent trying actions, observing results, and improving based on feedback from the discriminator, making the simulation an essential part of the learning process.

The environment setup is a key part of the GAIL training process, allowing the agent to learn through simulation. It includes three main components: the Robotic Arm Environment, the Environment Wrapper, and the Expert Demonstration Pipeline. The Robotic Arm Environment, built in Webots, simulates a realistic robot arm that can be controlled by adjusting its joints and collecting detailed observations. The Environment Wrapper connects this simulation to reinforcement learning tools such as Stable Baselines3 (SB3) [11] and Gymnasium by handling actions, observations, resets, and when episodes should end.

The training process uses tools like Weights & Biases (WandB) to track learning progress by visualizing graphs of model performance, such as learning curves, loss trends, and how well the policy and discriminator interact over time. This helps clearly see how well the model is improving during training. Final model saving ensures reproducibility and allows for deployment or further fine-tuning.

2.4 Behavior Cloning (BC) Training

The baseline evaluation involves training a similar dataset with Behavioral Cloning (BC), which is the conventional methodology used in imitation learning domain [9]. The process starts with loading the same 500 expert trajectories for training the GAIL policy, normalizing the data using min-max normalization. The BC training process uses supervised learning to map states to actions by mimicking expert demonstrations, with a typical BC training curve showing a decrease in loss over epochs.

Since BC uses supervised learning, it trains directly on the expert demonstration data without interacting with the simulation. This means the training is done offline, running through the terminal, and doesn't require the robot arm to perform actions in real time, unlike GAIL, which depends on continuous interaction with the simulation environment during training.

3. RESULTS AND DISCUSSION

3.1 Data Collection Result

A total of 500 expert trajectories were collected to serve as expert policy demonstrations for the GAIL framework. Each trajectory comprised sequential state-action pairs successfully collected from the teleoperation approach of manually controlling the robot to perform approach, grasp, lift, transport, and place target object within the Webots simulation environment. The resulting dataset consisted of approximately 141,000 samples, with 23 state features and 5 action dimensions across all trajectory data shown in the following tables.

Table 2. State Features Collected

Index	Name	Types	Description
0	state_0	State	Joint 1 position
1	state_1	State	Joint 2 position
2	state_2	State	Joint 3 position
3	state_3	State	Joint 4 position
4	state_4	State	Gripper Status (Open/Close
5	state_5	State	Joint 1 velocity
6	state_6	State	Joint 2 velocity
7	state_7	State	Joint 3 velocity
8	state_8	State	Joint 4 velocity
9	state_9	State	Gripper velocity
10	state_10	State	End-effector x-position
11	state_11	State	End-effector y-position
12	state_12	State	End-effector z-position
13	state_13	State	Object x-position
14	state_14	State	Object y-position
15	state_15	State	Object z-position
16	state_16	State	Distance between gripper and object
17	state_17	State	Gripper contact sensor (left)
18	state_18	State	Gripper contact sensor (right)
19	state_19	State	Gripper force/torque feedback
20	state_20	State	Object linear velocity
21	state_21	State	Object angular velocity
22	state_22	State	Task phase indicator (approaching, grasping, lifting, placing)

Table 3. Action Vectors

Index	Name	Types	Description
0	action_0	Action	Commanded Velocity for Joint 1
1	action_1	Action	Commanded Velocity for Joint 2
2	action_2	Action	Commanded Velocity for Joint 3
3	action_3	Action	Commanded Velocity for Joint 4
4	action_4	Action	Binary Gripper Command (Open/close)

A data quality check was performed across all 500 trajectories to ensure the recorded expert demonstrations were structurally sound and free from technical inconsistencies. The pandas-profiling tool was used to scan the dataset for missing values, feature ranges, and potential outliers.

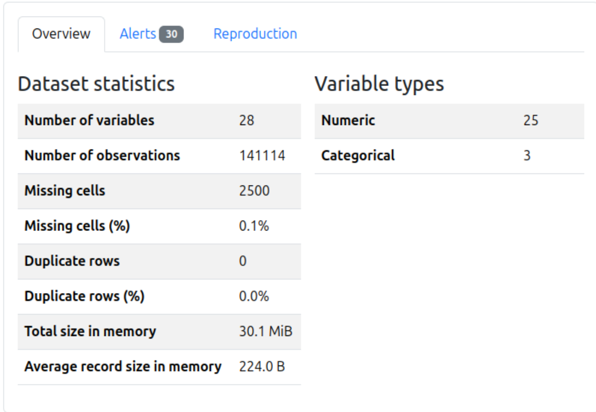

Fig. 5. Data Quality Analysis from Pandas Profiling

The dataset profiling summary (Fig. 5) indicates a robust and clean dataset consisting of 28 variables and over 141,000 observations, with only 0.1% missing cells (2,500 missing values). Notably, there are zero duplicate rows, and the dataset occupies approximately 30.1 MiB in memory. The majority of the variables are numeric (25), with only three categorical features. The small percentage of missing cells is minimal and originates from auxiliary metadata fields mentioned in Table 1 that are not essential for the training.

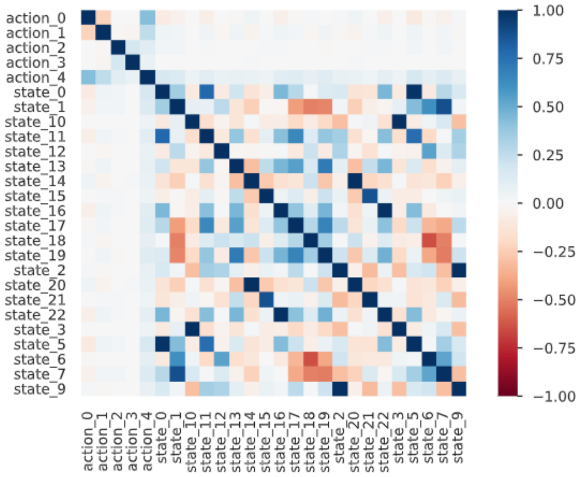

Fig. 6. States and Actions Correlational Heatmap

The correlation heatmap further illustrates the relationships among the numeric variables, particularly the state and action features. As expected, each variable shows perfect correlation with itself along the diagonal. Off-diagonal patterns reveal moderate correlations among some state variables, while action variables appear relatively uncorrelated with each other and most states. This means the dataset has a good mix of useful features that aren't too closely related to each other, which helps the model learn more effectively and reliably in imitation learning methods like GAIL.

3.2 Proximity Result

The study analyzes the gripper-object proximal performance of GAIL over the baseline algorithm (BC) across 50, 100, and 150 evaluation episodes.

The GAIL-trained policy demonstrated exceptional consistency across different evaluation scenarios, maintaining stable median minimum distances between the gripper and target object. This was attributed to its highly concentrated performance around 0.129-0.133 median values, with most attempts clustering tightly together. The interquartile range (IQR) values showed tight grouping of results, indicating that GAIL produced meaningful, predictable behavior.

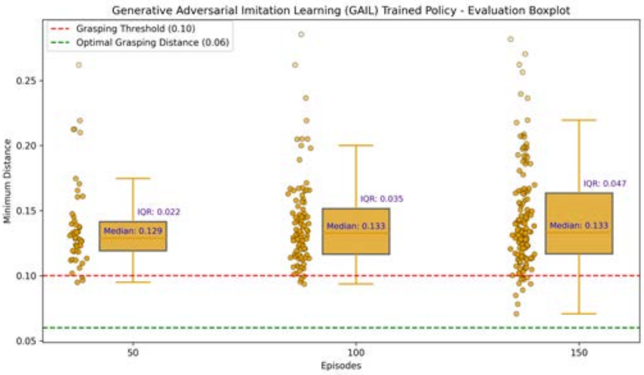

Fig. 7. GAIL produced tighter results near the graspable threshold across 50, 100, and 150 evaluation episodes.

On the other hand, BC-trained policy revealed different performance characteristics. BC generally maintained awful distances, as shown in Figure 8, from the target object, with median minimum distances of 0.241m for 50 episodes, 0.232m for 100 episodes, and 0.269m for 150 episodes. BC produced more scattered results with multiple peaks, suggesting that the BC-trained policy applied to the agent behaved less predictably across trials. The larger IQR values supported this observation.

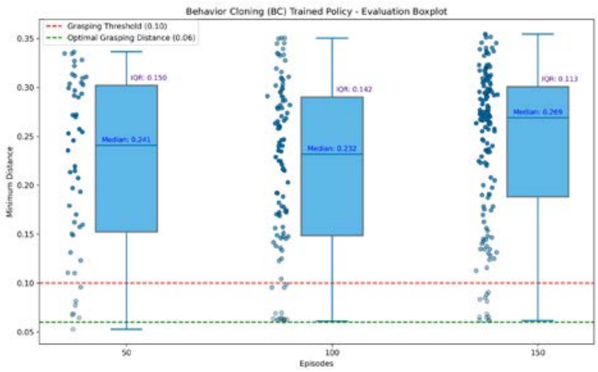

Fig. 8. BC produced more scattered results well above the graspable threshold across 50, 100, and 150 evaluation episodes.

The Mann-Whitney U tests comparing the distance distributions between BC and GAIL yielded p-values of 0.0000 across all episode counts, confirming that GAIL's performance improvements were statistically significant ($p < 0.05$). This result allowed the researchers to confidently reject the null hypothesis (BC Median = GAIL Median), confirming that the performance differences between BC and GAIL were not due to random variation but represent genuine improvements in gripper approaching the target object capability. A distribution analysis illustrating this claim is shown in Figure 9 and Table 3.

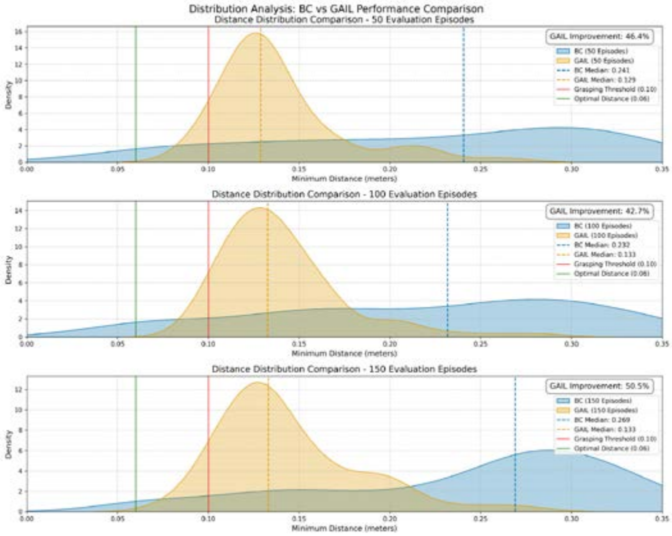

Fig. 9. Illustration of GAIL's consistent performance over BC's multimodality behavior with distance closer to the object's grasping thresholds.

Notable in these plots is how the overlap between BC and GAIL distributions decreases over time. At 50 episodes, there's moderate overlap between the blue and yellow areas, but by 150 episodes, they've become much more separate. This reinforces that BC and GAIL aren't just slight variations of the same approach; they're developing fundamentally different strategies for the grasping task.

As GAIL maintains its performance while BC degrades, the separation between them grows more obvious. This implies that these two trained policies behave in deployment and task evaluation very differently, with GAIL showing much better stability and meaningful actions.

Table 3. (a) Mann-Whitney U test result, rejecting the null hypothesis

Evaluation Episodes	BC Median	GAIL Median	Absolute Improvement
50	0.241	0.129	0.112
100	0.232	0.133	0.099
150	0.269	0.133	0.136

Table 3. (b) cont.

Evaluation Episodes	Improvement (%)	p-value	Statistically Significant
50	46.4%	0.0000	Yes
100	42.7%	0.0000	Yes
150	50.5%	0.0000	Yes

Despite GAIL's superior proximity performance, BC achieved higher success rates at meeting the 0.10m grasping threshold in most evaluations. As shown in Figure 10, BC succeeded 14.0% of the time in 50 and 100 episode evaluations and 8.0% in 150 episodes, while GAIL achieved 6.0%, 5.0%, and 6.7% respectively. This seeming contradiction is attributed to BC's broader distance distribution, where outlier episodes occasionally reached the threshold, whereas GAIL's consistent approaches often fell just short. Interestingly, BC's success rate decreased with more episodes, while GAIL's slightly improved between 100 and 150 episodes, suggesting BC's policy degraded over extended testing.

At the more challenging optimal distance threshold of 0.06m, both algorithms struggled, as illustrated in Figure 11, mostly showing 0% success. BC shortly reached a 2.0% success rate in 50 episodes but couldn't keep it up in longer tests. This shows that it's hard for either method alone to achieve higher precision of the gripper going close to the object for optimal grasping position.

Fig. 10. GAIL vs BC Grasping Threshold (≤0.10m) Success Rate

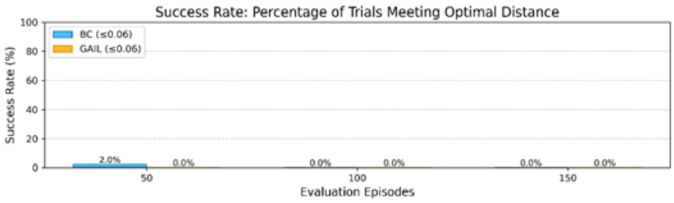

Fig. 11. GAIL vs BC Grasping Threshold (≤0.06m) Success Rate

3.3 Grasping and Pick-and-Place Result

The evaluation of grasping and pick-place performance revealed significant limitations in both approaches. Across the majority of evaluation episodes for both BC and GAIL, success rates were predominantly zero percent for complete task execution as shown in the table below.

Table 4. Grasping and Pick-and-Place Success Rate for both GAIL and BC

Algo	Eval. Eps.	Proximity (%)		Grasp (%)	Pick-Place (%)
		(≤0.10m)	(≤0.06m)		
GAIL	50	6.0%	0.0%	0.0%	0.0%
	100	5.0%	0.0%	0.0%	0.0%
	150	6.7%	0.0%	0.0%	0.0%
BC	50	14.0%	2.0%	0.0%	0.0%
	100	14.0%	0.0%	0.0%	0.0%
	150	8.0%	0.0%	0.0%	0.0%

Despite the generally poor quantitative results, qualitative observations we recorded inside the simulation during the 150 evaluation episodes of GAIL revealed a notable instance where the agent successfully grasped, lift, transport, and even place the target object just inside the source zone and not towards the experiment's target zone. This occurred during Episode 107, providing valuable insight into the agent's emerging capabilities, which is not captured by binary success metrics (Yes or No Basis).

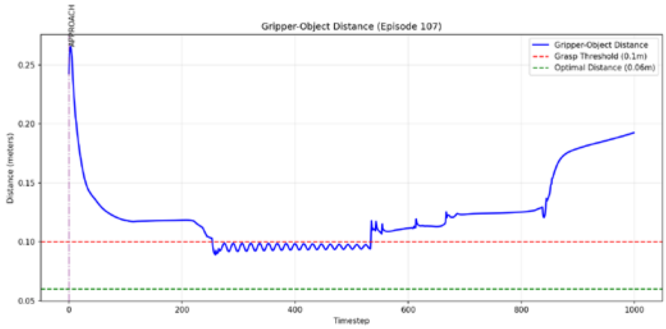

Fig. 12. Episode 107 of 150 GAIL's evaluation episode

At the start of the episode, the agent exhibited a clear approach behavior. The distance between the gripper and the object decreased rapidly from approximately 0.26 meters to just above the 0.1-meter grasp threshold. This behavior indicates a learned intent to approach towards the object, which is a foundational requirement for successful grasping. Around timestep 250 onward, the gripper began to shift around the 0.1-meter mark, occasionally moving slightly below. This pattern shows that the agent was attempting to position itself precisely for a grasp. The subtle swings during this interval imply agent and object transport performance from what we captured via recording (sample clips in Figure13). These brief moments of proximity are promising, as they fall within the acceptable threshold for triggering a grasp.

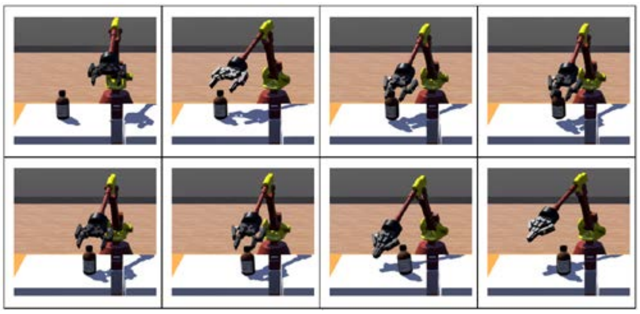

Fig. 13. Screenshots of recorded GAIL-trained Policy Performing Remarkably alike to expert demonstration

Beyond timestep 500, the gripper gradually increased its distance from the object, with no return to the graspable zone. This trend illustrates that the grasp was done triggered since the agent's gripper released the object after a short transport, followed by the agent drifting away.

This distance profile illustrates that while the agent did not successfully complete the full task, it did engage in structured interaction phases. Most notably, it showed that training a multivariable robot for manipulation tasks using GAIL is capable of performing most specially approaches and aligning with the object to a reasonable degree with a notable likelihood of replicating our expert demonstration behavior.

Supporting this interpretation, the 3D trajectory plot in Figure 14 shows a curved descent of the gripper and a spiral-like motion near an object, suggesting the agent attempted to align with the object before initiating a grasp. The trajectory does not progress towards the target zone, indicating an incomplete pick-and-place sequence. The object exhibits subtle movement, suggesting contact occurred. The agent did not successfully complete the entire pick-and-place task, but the data shows signs of learned sub-behaviors. However, this success was limited to gripping, lifting a bit, transporting shortly, and placing the object within the source zone without transporting it to the target zone. This observation highlights the need for meticulous evaluation criteria and expanded testing to recognize and notice partial task completion in complex robotic manipulation tasks.

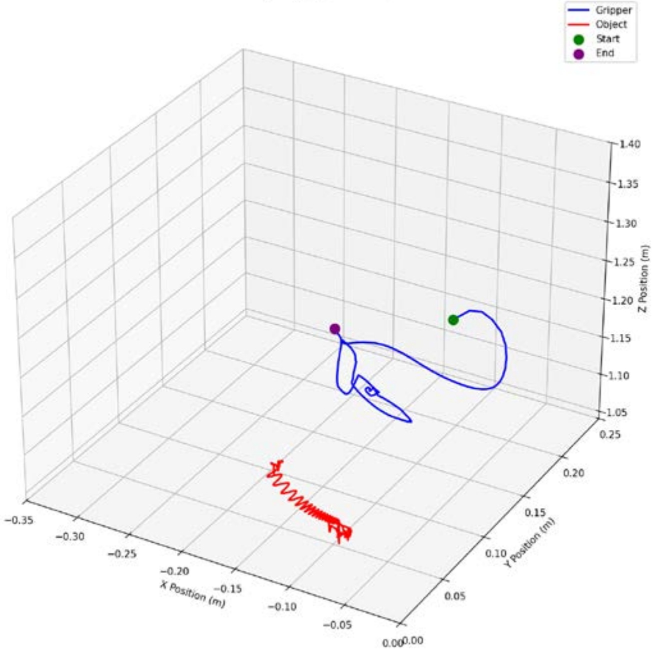

Fig. 14. Episode 107 of 150 GAIL evaluation episodes 3D Plot

Compared to the single notable grasp and pick-place attempt observed in GAIL's evaluation, the agent's testing using the trained policy of Behavior Cloning (BC) exhibited even more limited capabilities across all 150 recorded episodes. While GAIL showed some signs of structured manipulation behavior, such as alignment attempts and brief object contact, BC's behavior remained constrained to object approach alone, without progressing to any form of grasp or manipulation.

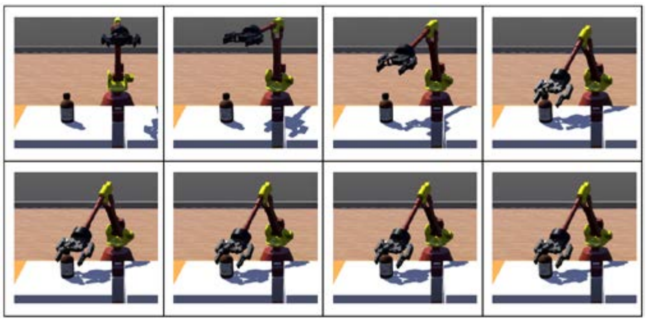

Fig. 15. Episode 92 of 150 BC-trained Policy Evaluation Performing Object Approach Only

In Episode 92, the closest gripper-object distance among 150 BC evaluation episodes, the agent successfully reduced the distance below the grasp threshold. However, no further actions were taken, and the agent remained static until the episode concluded. The 3D trajectory in Figure 16 showed a straightforward descent toward the object followed by minimal spatial activity. The agent trained using the BC algorithm learned to approach the object but lacked the capacity to transition into meaningful manipulation behaviors like grasping or pick-an-place.

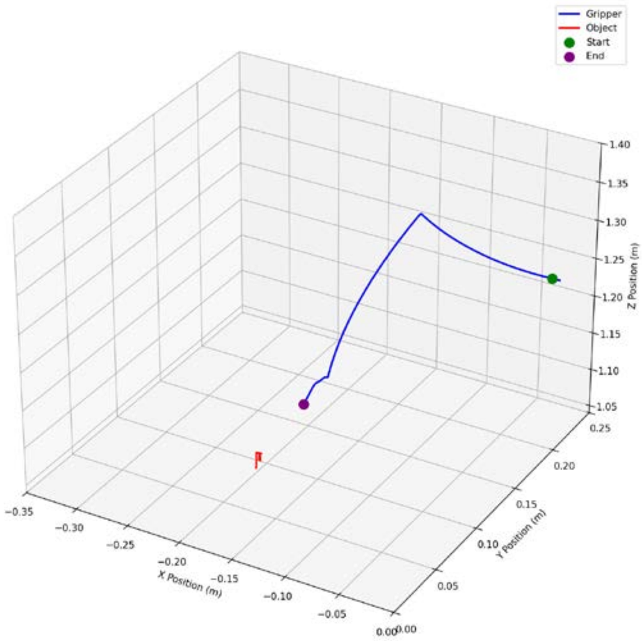

Fig. 16 Episode 92 of 150 BC Evaluation Episode 3D Plot

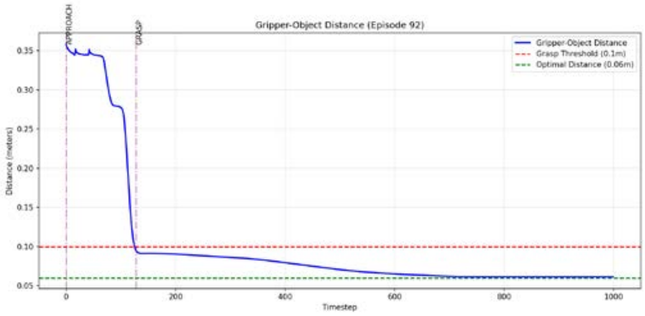

Fig. 17. BC's Closest Gripper-Object Distance across 150 episodes

This contrast with GAIL's behavior highlights the potential of adversarial imitation methods to capture richer task dynamics beyond mere proximity. However, behavior cloning is limited when trained on finite demonstrations without corrective feedback, and its likelihood of replicating expert demonstrations is limited due to the algorithm's offline nature and no simulation interaction needed.

4. CONCLUSION

This study demonstrated that GAIL can effectively learn complex robotic manipulation behaviors in simulation, showing consistent and precise object approach even if full task completion was not achieved. Compared to Behavioral Cloning (BC), GAIL produced more stable and adaptable policies due to its adversarial training loop. BC, while easier to train, struggled with longer episodes and showed less consistent behavior.

GAIL also performed better in maintaining close proximity to the target object and showed potential for generalization under randomized conditions in the simulation environment. Notably, domain randomization, in the form of randomized object positions, was applied throughout the entire pipeline: during expert data collection, model training, and evaluation. This approach exposed the models to a variety of conditions, enhancing robustness and better simulating real-world variability.

Overall, the research achieved its goals, successfully collecting expert data, training a GAIL model, observing generalizability performance from randomized object positions, and comparing it with BC to highlight each method's strengths and limitations in robotic manipulation tasks inside a sophisticated simulation software with custom environment.

5. RECOMMENDATIONS

To enhance future investigations into robotic manipulation learning, it is recommended that researchers consider the impact of robotic hand design on learned task performance. As noted by Lotti and Vassura, the mechanical design of articulated fingers and the use of mechanisms like tendon pulleys are fundamental to achieving effective grasping and manipulation capabilities in physical robotic hands [5]. Exploring how variations in the simulated robot's hand or gripper design might affect the learning efficiency and success rates of algorithms such as GAIL and BC could provide valuable insights into the interplay between hardware capabilities and learning strategies in simulation.

While the application of Generative Adversarial Imitation Learning (GAIL) demonstrated promising results, including the emergence of manipulation sub-behaviors not present in baseline method (e.g. BC), achieving full task completion for complex manipulation sequences remains a challenge. Future work should therefore explore alternative and hybrid learning approaches to enhance performance and robustness. This could include integrating Behavior Cloning (BC) pre-training followed by GAIL fine-tuning or investigating recent algorithms like Diffusion-Reward GAIL or other hybrid methods [6][7][12]. These might help balance training stability and behavioral richness. Furthermore, approaches such as "attention-based robot learning of haptic interaction", as explored by Moringen et al. [4], which leverage physics-driven simulations to learn complex interaction skills with minimal reliance on prior knowledge, may offer valuable insights. Investigating such methods within the existing Webots simulation environment, which is also a physics-driven simulation, could provide alternative strategies for training robotic arms for manipulation tasks.

Also, While GAIL algorithm has shown the ability to approach objects and align the gripper effectively, future work should focus as well on achieving full manipulation behaviors, specifically triggering the gripper to grasp, lift, transport, and even place the object into a defined target zone as this study’s ultimate goal. Finally, extend evaluation episodes and enhance evaluation success metrics design.

6. REFERENCES

Reference to a journal publication:

[1] Ho, J., & Ermon, S. (2016). Generative adversarial imitation learning. Advances in neural information processing systems, 29. https://scholar.google.com/scholar_lookup?arxiv_id=1606.03476

[2] Xie, A., Lee, L., Xiao, T., Finn, C. (2023). Decomposing the generalization gap in imitation learning for visual robotic manipulation. arXiv preprint arXiv:2307.03659

[3] Xu, Z., Zhao, Y., Wu, K., Liu, N., Ji, J., Che, Z., Liu, C. H., & Tang, J. (2025). HACTS: A Human-As-Copilot Teleoperation System for Robot Learning. arXiv preprint arXiv:2503.24070.

[4]Nisky, I., Hartcher-O’Brien, J., Wiertlewski, M., & Smeets, J. (2020). Haptics: science, technology, applications. In Lecture notes in computer science. https://hdl.handle.net/2324/6937662

[5]Manocha, N., Agnihotri, H., Khan, N. M. A., & Neha, E. (2024). Conceptual Design of 17 DOF Dexterous Robotic Hand for grasping and Manipulation Task. Evergreen, 11(2), 1254–1262. https://hdl.handle.net/2324/7183432

[6]Lai, C.-M., Wang, H.-C., Hsieh, P.-C., Wang, Y.-C. F., Chen, M.-H., & Sun, S.-H. (2024). Diffusion-Reward Adversarial Imitation Learning. *arXiv:2405.16194 [cs.LG]*.

[7]Belkhale, S., Cui, Y., & Sadigh, D. (2023). HYDRA: Hybrid Robot Actions for Imitation Learning. PMLR. https://proceedings.mlr.press/v229/belkhale23a.html

[8]Khalil, A. B., Sun, Z., & Tian, Y. (2020). Dexterous grasping in clutter using deep reinforcement learning with prioritized experience replay. Robotics and Autonomous Systems, 133, 103620. https://doi.org/10.1016/j.robot.2020.103620.

[9]Torabi, F., Warnell, G., & Stone, P. (2018). Behavioral cloning from observation. arXiv preprint arXiv:1805.01954.

[10]Morales A, León B, Chinellato E and Suárez R (2022) Editorial: Current Challenges and Future Developments in Robot Grasping. Front. Robot. AI 9:973208. doi: 10.3389/frobt.2022.973208

[11]Schulman, J., Wolski, F., Dhariwal, P., Radford, A., & Klimov, O. (2017). Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347.

[12]Fei, C., Wang, B., Zhuang, Y., Zhang, Z., Hao, J., Zhang, H., ... & Liu, W. (2020). Triple-GAIL: a multi-modal imitation learning framework with generative adversarial nets. arXiv preprint arXiv:2005.10622.

[13]Surati, S., Hedaoo, S., Rotti, T., Ahuja, V., & Patel, N. (2021). Pick and Place Robotic Arm: a review paper. (311). https://www.irjet.net/volume8-issue02

[14]Merel, J., Tassa, Y., TB, D., Srinivasan, S., Lemmon, J., Wang, Z., ... & Heess, N. (2017). Learning human behaviors from motion capture by adversarial imitation. arXiv preprint arXiv:1707.02201.

[15]Ou, Y., & Tavakoli, M. (2023). Towards safe and efficient reinforcement learning for surgical robots using real-time human supervision and demonstration. In 2023 International Symposium on Medical Robotics (ISMR) (pp. 1–7). IEEE. https://doi.org/10.1109/ISMR57123.2023.10130214