Abstract

Robot technology is one of the pillars of modern society. Advances in information, electronic, and mechanical fields enable us to build and program machines to perform tasks in very different contexts, such as industry, surgery, and space missions.

While in the early day, robot systems were constrained in isolated and known environments. Over the past few decades, robots have been asked to solve tasks in dynamic and unknown/partially known environments, where they must **coexist** and **cooperate** with humans, while solving different **dynamic** tasks [1] (e.g. pick a requested object, whose position is not known a priori).

In this scenario, the desired characteristics of such robotic systems are: **(a) Adaptability to new conditions**, i.e., the system must be able to easily adapt to dynamic changes in system and environmental conditions, performing *"intelligent" behaviors* to handle these new scenarios and solve the desired task; **(b) Adaptability to new tasks**, i.e., the system must be able to easily adapt to both new variations of a known task and completely new tasks by exploiting experience to infer actions and solve them.

Significant advancements have been made by leveraging *learning techniques*, where the control policy is inferred from data. This data can be generated either by **agent experience** [2] or by **expert demonstrations** [3].
In the case of expert demonstrations, the control policy parameters are directly tuned using a dataset containing examples of task execution. Here, the goal is to replicate the tasks observed in the dataset.

Given this background, the thesis is framed in the context of *Learning from Demonstration* (LfD), a learning approach based on expert demonstrations. According to the requirements of adaptability the thesis focus on a specific aspect of LfD, named *Multi-Task LfD*. In this case, the control policy is trained to handle various variations of a specific task (e.g., picking an object from different possible locations) [4] or even entirely different tasks (e.g., a single control policy that solves both picking and placing tasks as

well as assembly tasks) [5, 6]. The goal is to generalize not only with respect to the objects being manipulated and the initial conditions but also with respect to the tasks themselves. This means that it is possible to achieve a system capable of solving new variations by leveraging the knowledge-sharing hypothesis.

In this scenario, the learning procedure is much more challenging because there is the need to include and define the **conditioning signal** (i.e., the signal that informs the policy about the task to execute, the object to manipulate, and the target placing location). Additionally, the environment can contain **multiple distractor objects** (e.g., objects that can potentially be manipulated but are not of interest for a given task variation).

Regarding the conditioning signal, there are at least two intuitive approaches. The first is through a natural language description of the task to be executed [7, 8, 5], and the second is through a video demonstration [4, 6]. The modality of interest is the second one, where another agent (either a robot or a human operator) performs the task in a different environment configuration, records this execution, and provides the video as input to the control policy. The control policy must then infer the intent from the video (i.e., the task to be performed, the object to be manipulated, and the final state) and control the robot to complete the task according to the agent's state, the environment's state, and the commanded task.

Regarding the issue of distractor objects, they are typically defined as items present in the scene but never involved in manipulation operations. Modern deep architectures can handle this scenario effectively, as they can easily learn to ignore these objects since they do not participate in any manipulation tasks. However, in the context proposed in this thesis, the problem is further complicated by the fact that the semantic meaning of an object (i.e., target or distractor) is defined at run-time by the command itself. This means that if the initial configuration consists of four objects (e.g., four boxes of different colors), a specific object may or may not be of interest based on the command given to the robot.

The primary contribution of this thesis is tackling the challenge of distractor objects. A key issue identified in existing literature is **target misidentification**, where the learned control policy generates valid trajectories, enabling the robot to reach, pick, and place objects, but frequently manipulates the wrong object.

To solve this problem, two main considerations were made:

(1) Architectures proposed in the current literature are predominantly **end-to-end**, translating high-dimensional inputs, such as images, into corresponding low-dimensional actions. As a result, the model must learn an implicit representation that encodes both the task objective and the current

state of the environment, including the location of the target object; **(2)** The learning procedure optimizes an **action-centric metric**, meaning that it is not directly linked to task success but instead focuses on mimicking the expert's actions on average. This action-focused optimization can lead to poor encoding of critical information, such as object positions.

These two factors can result in a control policy that fails to effectively guide the robot toward the target object. In particular, it was observed that the early stages of trajectory execution are critical. Even small errors during these initial steps can cause the robot to reach and ultimately pick the wrong object.

Based on these considerations, this thesis explores the development of a **modular** architecture instead of an end-to-end approach. This architecture features modules specifically designed for reasoning about the objects of interest (e.g., target object and placement location). The outputs of these reasoning modules are then integrated to simplify the learning problem for the Control Module. This module is now informed by low-dimensional information, such as the position of the target object, which may be more effectively utilized during the learning process, especially in light of the action-centric cloning loss.

To perform this explicit reasoning, a *Conditioned Object Detector* (COD) has been developed. This module, given the video demonstration and the current agent observation as input, predicts the category-agnostic bounding box related to the target object and the final placing location. This low-level positional information is then provided to the control module, which predicts the actions to perform.

The learning procedure is then divided into two steps. The first step involves training the *Conditioned Object Detector* (COD) module, which focuses on explicitly solving cognitive tasks, such as detecting regions of interest represented by the object to be manipulated and its final location. The second step involves training the *Object Conditioned Control Policy* (OCCP), which focuses on solving the control problem using low-level positional information that can be easily mapped into the corresponding actions.

The final system has been extensively tested in simulation environments, then it was also validated on a real-world robotic platform.

Regarding the simulated environment, the system was evaluated on both **multi-variation single-task** scenarios and **multi-variation multi-tasks** scenarios, considering four simulated tasks: Pick-Place, Nut-Assembly, Stack-Block, and Button-Press. Each task had different variations based on the manipulated object and the final state. While the tasks share common properties, they also have specific characteristics. For example, the Nut-Assembly task involves contact-rich, precise manipulation, whereas Pick-Place can be

solved in a much rougher manner.

Overall, the proposed methods demonstrated very promising behaviors and a general improvement over baseline methods that do not include object-related reasoning. Specifically, in the single-task multi-variation scenarios, the proposed method achieved a success rate of **90.13%** on average, which is an improvement of **+28.78%** compared to the baseline method. In the multi-task multi-variation scenarios, the proposed method achieved a success rate of **79.24%** on average, which is an improvement of **+33.23%** compared to the baseline method. This shows that solving manipulation tasks with an object-oriented approach can be an effective paradigm for LfD problems. Additionally, this approach provides interpretable information to the end user, as the predicted bounding boxes can be interpreted as the locations where the robot will move.

In conclusion, the proposed method was tested also in a real-world environment, where the complexity of the problem is heightened by the presence of less and noisy data collected through teleoperation. Even under these challenging conditions, the proposed method demonstrated its effectiveness in addressing both the cognitive and control problems, reaching a success rate of **55.00%**, which is a strong improvement with respect to the **0.00%** of the baseline method in the same training condition. This confirms that the object prior can be successfully applied in real-world scenarios, enabling the development of a reliable system despite limited and noisy trajectory data.

Bibliography

[1] S. Bini, G. Percannella, A. Saggese, and M. Vento, "A multi-task network for speaker and command recognition in industrial environments," *Pattern Recognition Letters*, vol. 176, pp. 62–68, 2023.

[2] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction.* MIT press, 2018.

[3] T. Osa, J. Pajarinen, G. Neumann, J. A. Bagnell, P. Abbeel, J. Peters *et al.*, "An algorithmic perspective on imitation learning," *Foundations and Trends® in Robotics*, vol. 7, no. 1-2, pp. 1–179, 2018.

[4] S. Dasari and A. Gupta, "Transformers for one-shot visual imitation," in *Conference on Robot Learning.* PMLR, 2021, pp. 2071–2084.

[5] 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, I. Leal, K. Lee, S. Levine, Y. Lu, U. Malla, D. Manjunath, I. Mordatch, O. Nachum, C. Parada, J. Peralta, E. Perez, K. Pertsch, J. Quiambao, K. Rao, M. S. Ryoo, G. Salazar, P. R. Sanketi, K. Sayed, J. Singh, S. Sontakke, A. Stone, C. Tan, H. T. Tran, V. Vanhoucke, S. Vega, Q. Vuong, F. Xia, T. Xiao, P. Xu, S. Xu, T. Yu, and B. Zitkovich, "RT-1: robotics transformer for real-world control at scale," in *Robotics: Science and Systems XIX, Daegu, Republic of Korea, July 10-14, 2023*, K. E. Bekris, K. Hauser, S. L. Herbert, and J. Yu, Eds., 2023. [Online]. Available: https://doi.org/10.15607/RSS.2023.XIX.025

[6] Z. Mandi, F. Liu, K. Lee, and P. Abbeel, "Towards more generalizable one-shot visual imitation learning," in *2022 International Conference on Robotics and Automation (ICRA).* IEEE, 2022, pp. 2434–2444.

[7] S. Stepputtis, J. Campbell, M. Phielipp, S. Lee, C. Baral, and H. Ben Amor, "Language-conditioned imitation learning for robot manipulation tasks," *Advances in Neural Information Processing Systems*, vol. 33, pp. 13 139–13 150, 2020.

[8] O. Mees, L. Hermann, E. Rosete-Beas, and W. Burgard, "Calvin: A benchmark for language-conditioned policy learning for long-horizon robot manipulation tasks," *IEEE Robotics and Automation Letters (RA-L)*, vol. 7, no. 3, pp. 7327–7334, 2022.