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** [9, 10], 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.

[9] P. Foggia, F. Rosa, and M. Vento, "Enhancing robotic demonstration-based learning method with preliminary visual target localization," in *European Robotics Forum*. Springer, 2024, pp. 212–217.

[10] ——, "Improving learning from visual demonstration methods by target localization," in *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*. IEEE, 2024, pp. 740–747.

Abstract

La robotica è una delle tecnologie più importanti della società moderna. I progressi nei campi dell'informazione, dell'elettronica e della meccanica ci permettono di costruire e programmare macchine per svolgere compiti nei contesti più disparati, come l'industria, la chirurgia e il settore aerospaziale. In particolare, nella produzione manifatturiera, i robot vengono utilizzati principalmente per eseguire lavori ripetitivi e potenzialmente dannosi per l'uomo.

In passato, i sistemi robotici venivano utilizzati in contesti con forti limitazioni date dal fatto che l'ambiente era chiuso e noto a priori. Negli ultimi decenni, grazie anche allo sviluppo del paradigma dell'Industria 4.0, i robot sono immersi in contesti più flessibili, rappresentati da ambienti non noti o parzialmente noti a priori, dove devono **coesistere** e **cooperare** con gli esseri umani, risolvendo diversi compiti **dinamici** [1] (ad esempio, prelevare un oggetto richiesto la cui posizione non è nota a priori).

In questo scenario, le caratteristiche desiderate di tali sistemi robotici sono: **(a) Adattabilità a nuove condizioni**, ossia il sistema deve essere in grado di adattarsi facilmente ai cambiamenti delle condizioni di contesto, mostrando comportamenti *"intelligenti"* per affrontare questi nuovi scenari andando a risolvere il problema di interesse; **(b) Adattabilità a nuovi compiti**, ossia il sistema deve essere in grado di adattarsi facilmente sia a nuove varianti di un compito conosciuto che a compiti completamente nuovi, sfruttando l'esperienza per inferire le azioni necessarie a risolverli.

Per raggiungere questi requisiti sono state ampiamente utilizzate *tecniche di apprendimento*, in cui la politica di controllo viene dedotta dai dati. Questi dati possono essere generati sia tramite **l'esperienza dell'agente** [2] che tramite **dimostrazioni di esperti** [3]. Nel caso delle dimostrazioni di esperti, i parametri della politica di controllo vengono direttamente adattati utilizzando un dataset contenente esempi dell'esecuzione del compito. L'obiettivo qui è replicare i compiti osservati nel dataset.

La presente tesi si contestualizza nel *Learning from Demonstration* (LfD), un approccio di apprendimento basato su dimostrazioni di esperti. Rispetto

i requisiti di adattabilità, la tesi si concentra su un aspetto specifico dell'LfD, denominato *Multi-Task LfD*. In questo caso, la politica di controllo viene addestrata per gestire varie varianti di uno specifico compito (ad esempio, prendere un oggetto da diverse posizioni possibili) [4] o persino compiti completamente diversi (ad esempio, una singola politica di controllo che risolve sia compiti di pick-and-place sia compiti di assemblaggio) [5, 6], con l'obiettivo di generalizzare non solo rispetto agli oggetti manipolati e alle condizioni iniziali, ma anche rispetto ai compiti stessi. Questo significa che, sfruttando l'ipotesi del *knowledge-sharing*, possiamo ottenere un sistema in grado di risolvere nuove variazioni.

In questo scenario, bisogna definire un **segnale di condizionamento** (ossia, il segnale che informa la politica sul compito da eseguire, l'oggetto da manipolare e la posizione di destinazione). Inoltre, l'ambiente può contenere diversi **oggetti distrattori** (ad esempio, oggetti che potrebbero essere manipolati ma che non sono di interesse per una determinata variazione del compito).

Per quanto riguarda il segnale di condizionamento, si possono definire almeno due approcci. Il primo rappresentato da una descrizione in linguaggio naturale del compito da eseguire [7, 8, 5], il secondo è rappresentato da una dimostrazione video [4, 6].
Il caso di interesse per la tesi è il secondo, dove un altro agente (sia esso un robot o un operatore umano) esegue il compito in una diversa configurazione ambientale, registra questa esecuzione e fornisce il video come input alla politica di controllo. La politica di controllo deve quindi dedurre l'intento dal video (ossia, il compito da eseguire, l'oggetto da manipolare e lo stato finale) e controllare il robot per completare il compito in base allo stato dell'agente, allo stato dell'ambiente e al compito comandato.

Per quanto riguarda il problema legato alla presenza di oggetti distrattori, in generale questi sono oggetti che non vengono mai considerati in operazioni di manipolazione, semplificando di molto il problema. Tuttavia, nel contesto proposto in questa tesi, il problema è ulteriormente enfatizzato dal fatto che il significato semantico di oggetto di interesse o distrattore è definito a runtime dal comando stesso. Questo significa che se la configurazione iniziale è composta da quattro oggetti (ad esempio, quattro box di colore diverso), sulla base del comando dato al robot un determinato oggetto può diventare o meno di interesse.

Il principale contributo di questa tesi è quello di svillupare un sistema che sia robusto alla presenza di distrattori all'interno della scena. Nello specifico, un problema chiave identificato in letteratura è la **target missidentification**[9, 10], questo significa che la politica di controllo appresa genera traiettorie valide, permettendo al robot di raggiungere, prendere e posizionare

oggetti, ma manipolando l'oggetto sbagliato.

Per risolvere questo problema, sono state fatte due considerazioni principali: **(1)** Le architetture proposte in letteratura sono prevalentemente **end-to-end**, questo significa che traducono input ad alta dimensionalità (immagini) nelle corrispondenti azioni (posa del gripper rispetto ad un frame di riferimento). Con questo approccio, il modello deve imparare una rappresentazione implicita che codifica sia l'obiettivo del compito che lo stato corrente dell'ambiente, compresa la posizione dell'oggetto target; **(2)** La procedura di apprendimento ottimizza una funzione di errore che si focalizza esclusivamente sull'azione, questo significa che il sistema durante l'addestramento ha come obiettivo quello di generare in media le stesse azioni presenti nel dataset. Questa procedura può portare il sistema al non focalizzarsi sulla codifica di informazioni di interesse come la posizione degli oggetti.

Questi due fattori possono portare a una politica di controllo che non riesce a guidare efficacemente il robot verso l'oggetto target. In particolare, è stato osservato che le fasi iniziali dell'esecuzione della traiettoria sono cruciali. Infatti, anche piccoli errori durante questi primi passi possono portare il robot al raggiungimento e la presa dell'oggetto sbagliato.

Basandosi su queste considerazioni, questa tesi esplora lo sviluppo di un'architettura **modulare**, in contrapposizione agli approcci end-to-end proposti in letteratura. Questa architettura prevede moduli specificamente progettati per ragionare sulle zone di interesse (ad esempio, la posizione dell'oggetto target e la posizione finale di posizionamento). Quindi, una volta individuate queste zone di interesse, attraverso la generazione di bounding-box, questi possono essere integrati nell'input del modulo di controllo, che ora riceve anche informazioni a bassa dimensionalità, come la posizione dell'oggetto target, che possono essere più facilmente utilizzate durante il processo di apprendimento, soprattutto alla luce della perdita centrata sull'imitazione dell'azione.

Per eseguire questo ragionamento esplicito, è stato sviluppato un *Conditioned Object Detector* (COD). Questo modulo, dato in input il video della dimostrazione e l'osservazione corrente dell'agente, predice il bounding-box relativo all'oggetto target e alla posizione finale.

La procedura di apprendimento viene quindi suddivisa in due fasi. La prima fase prevede l'addestramento del modulo COD, focalizzandosi sulla risoluzione dei problemi cognitivi di detection. La seconda fase prevede l'addestramento della *Object-Conditioned Control Policy* (OCCP), che si concentra sulla risoluzione del problema di controllo sfruttando le informazioni posizionali generati dal COD.

Il sistema finale è stato ampiamente testato in simulato, dove è possibile generare scenari e collezionare traiettorie per il dataset. La validazione

del sistema si è conclusa attraverso il testing dei metodi proposti su una piattaforma robotica reale.

Per quanto riguarda l'ambiente simulato, il sistema è stato valutato sia in scenari definiti **multi-variation single-task** che in scenari definiti **multi-variation multi-task**, considerando quattro compiti: Pick-Place, Nut-Assembly, Stack-Block e Press-Button. Ogni task è caratterizzato dalla presenza di diverse varianti definite sulla base dell'oggetto manipolato e del suo stato finale. Per esempio, nel task di Pick-Place sono presenti 4 box e 4 bin, le variazioni sono rappresentate dalle possibili combinazioni di box da prelevare e bin dove eseguire il placing.

I task selezionati sono caratterizzate dalla presenza sia di caratteristiche comuni, ma anche di caratteristiche specifiche. Ad esempio, il compito di Nut-Assembly comporta una manipolazione precisa, mentre il Pick-Place può essere risolto in modo molto più approssimativo.

Nel complesso, i metodi proposti hanno dimostrato comportamenti molto promettenti e un miglioramento generale rispetto ai metodi di base che non includono il ragionamento sugli oggetti. Nello specifico, nello scenario multi-variation single-task, il miglior metodo proposto ha raggiunto un tasso di successo medio del **90.13%** che rappresenta un miglioramento del **+28.78%** rispetto alla baseline utilzzata. Invece, nello scenario multi-variation multi-task, il miglio metodo proposto ha raggiunto un tasso di successo medio del **79.24%** che rappresenta un miglioramento del **+33.23%** rispetto alla baseline utilizzata.

Ciò dimostra che risolvere compiti di manipolazione con un approccio orientato agli oggetti può essere un paradigma efficace per i problemi di LfD. Inoltre, questo approccio fornisce informazioni interpretabili all'utente finale, poiché i bounding-box predetti possono essere interpretati come le posizioni verso cui si muoverà il robot.

In conclusione, il metodo proposto è stato testato anche in un ambiente reale, dove la complessità del problema è aumentata dalla presenza di meno dati e rumorosi raccolti tramite teleoperazione. Anche in queste condizioni, il metodo proposto ha dimostrato la sua efficacia nell'affrontare sia i problemi cognitivi che quelli di controllo, raggiungento un tasso di successo del **55.00%**, che risulta un netto miglioramento rispetto allo **0.00%** della baseline a parità di condizioni di addestramento. Ciò conferma che l'informazione legata all'oggetto di interesse può essere applicata con successo in scenari reali, consentendo lo sviluppo di un sistema affidabile nonostante dati limitati e rumorosi.

Bibliografia

[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.

[9] P. Foggia, F. Rosa, and M. Vento, "Enhancing robotic demonstration-based learning method with preliminary visual target localization," in *European Robotics Forum*. Springer, 2024, pp. 212–217.

[10] ——, "Improving learning from visual demonstration methods by target localization," in *2024 33rd IEEE International Conference on Robot and Human Interactive Communication (ROMAN)*. IEEE, 2024, pp. 740–747.