PhD in Information Technology and Electrical Engineering
Università degli Studi di Napoli Federico II

PhD Student: Giuseppe Rauso

Cycle: XXXIX

Training and Research Activities Report

Year: First

Tutor: prof. Albero Finzi

Co-Tutor: prof. Vincenzo Lippiello

Date: November 6, 2024

Training and Research Activities Report

PhD in Information Technology and Electrical Engineering

Cycle: **Author:**

1. Information:

- **PhD student: Giuseppe Rauso**
- **DR number: DR997198**
- **Date of birth: 30/09/1998**
- **Master Science degree: Computer Science** **University: University of Naples Federico II**
- **Doctoral Cycle: XXXIX**
- **Scholarship type: PNRR Partenariato Esteso PE15**
- **Tutor: Alberto Finzi**
- **Co-tutor: Vincenzo Lippiello**

2. Study and training activities:

Activity	Type[1]	Hours	Credits	Dates	Organizer	Certificate[2]
Energy-efficient data science	Seminar	1	0.2	13/12/2023	Prof. Elio Masciari	Y
Robotics meets AI & 5G: il Futuro è adesso!	Seminar	2	0.4	14/12/2023	Prof. Paolo Massarotti	Y
STRATEGIC ORIENTATION FOR STEM RESEARCH & WRITING	Course	24	5	07/12/2023 - 23/02/2024	Dr. Chie Shin Fraser	Y
Hominis	Seminar	5	1	22/02/2024	Prof. Carlo Sansone, Ing. Stefano Marrone	Y
Matrix Analysis for Signal Processing with MATLAB Examples	Course	14	3	30/04/2024 - 28/05/2024	Dr. Massimo Rosamilia	Y
Big Data Architecture and Analytics	Course	20	5	06/05/2024 - 21/05/2024	Dr. Giancarlo Sperlì	Y
Using Deep Learning properly	Course	12	4	23/01/2024 - 08/02/2024	Dr. Andrea Apicella	Y

Introduction to Large Language Models: Evolution and the current state	Seminar	2	0.4	10/06/2024	Dr. Giancarlo Sperlì	Y
SUSTAINABLE IT: STRATEGIES AND BEST PRACTICES FOR A GREEN ENGINEERING FUTURE	Seminar	5	1	27/05/2024	Prof. Antonia Maria Tulino	Y
PERCHE' DIGITAL, IN UN MONDO CHE SEMBRA GIA' ESTENSIVAMENTE DIGITALE E PERCHE' TRASFORMARE – TIM	Seminar	4	0.8	10/05/2024	Prof. Antonia Maria Tulino	Y
INTELLIGENZA ARTIFICIALE E REGOLE DEL MERCATO	Seminar	2	0.4	14/05/2024	Prof. Antonia Maria Tulino	Y
REWIRE THE BRAIN: THE POTENTIAL OF NEUROPLASTICITY	Seminar	2	0.4	09/05/2024	Prof. Antonia Maria Tulino	Y
VERSO UNA GESTIONE SMART DELLA RISORSA IDRICA CON IL SUPPORTO DELLA DIGITAL INNOVATION	Seminar	1	0.2	14/05/2024	Prof. Antonia Maria Tulino	Y
GENERATIVE AI FOR SOFTWARE ENGINEERING: STRATEGIES, IMPACTS AND PRACTICAL APPLICATIONS	Seminar	5	1	29/05/2024	Prof. Antonia Maria Tulino	Y

IEEE Authorship and Open Access Symposium: Tips and Best Practices to Get Published from IEEE Editors	Seminar	2	0.4	07/05/2024	IEEE Xplore	Y
Machine Deception	Seminar	1	0.2	23/05/2024	Prof. Alessandra Rossi	Y
Scienza moderna e disciplina giuridica dell'Intelligenza Artificiale	Course	20	6	03/06/2024 - 03/07/2024	Prof. Lucio Franzese	Y
Topological Signal Processing and Learning	Seminar	1	0.2	17/07/2024	Prof. Antonia Maria Tulino	Y

1) Courses, Seminar, Doctoral School, Research, Tutorship
2) Choose: Y or N

2.1. Study and training activities - credits earned

	Courses	**Seminars**	**Research**	**Tutorship**	**Total**
Bimonth 1	**0**	**0.6**	**7**	**0**	**7.6**
Bimonth 2	**5**	**1**	**10**	**0**	**16**
Bimonth 3	**0**	**0**	**6**	**0**	**6**
Bimonth 4	**12**	**4.8**	**4**	**0**	**20.8**
Bimonth 5	**6**	**0.2**	**4**	**0**	**10.2**
Bimonth 6	**0**	**0**	**4**	**0**	**4**
Total	**23**	**6.6**	**35**	**0**	**64.6**
Expected	**20 - 40**	**5 - 10**	**10 - 35**	**0 – 1.6**	

3. Research activity:

During my first year of the PhD, I focused on the incremental learning of structured robotic tasks in simulated environments. This approach allows for decomposing the main task into subtasks and training specialized modules. In my research, particular attention was given to the modularity and reusability of the components. Specifically, I primarily addressed two topics: the first concerns incremental learning from demonstration of robotic manipulation tasks in a virtual environment, building upon the work carried out in my master's thesis; the second concerns structured task learning and execution using combined attentional mechanisms, both text-visual and task-oriented.

Learning from Demonstration (LfD) is a valuable approach for robotic task training, providing a natural alternative to traditional programming and accelerating learning through human guidance. Demonstrations in virtual environments are particularly advantageous, as they remove the complexities of physical interactions. However, VR demonstrations come with their own challenges, including the gap between virtual and real-world dynamics and the difficulty in reproducing realistic conditions for effective training. Research in this area explores various methods for delivering expert demonstrations in VR, such as teleoperation, kinesthetic guidance, and motion capture. In my research, I focused on VR-based demonstrations for training robotic manipulators with complex, multi-fingered end-effectors, using techniques like Behavioral Cloning and Generative Adversarial Imitation Learning (GAIL) to enhance learning. One of the main objectives is to provide a generalizable and modular methodology that can be applied to various robotic manipulators. Demonstrations obtained in a simulated environment by teleoperating the robot in VR allow for immersion in the same conditions as the robotic agent, thereby increasing their quality and fidelity. I combined imitation learning and reinforcement learning techniques to enable a robotic manipulator to incrementally learn tasks, leveraging a limited number of demonstrations within the chosen framework. In the works that I presented respectively at the AIRO 2023 workshop and the IROS 2024 conference, I applied this methodology to learning the grasp-and-lift task, considering two incremental phases: in the first phase, the learning focuses solely on the robotic hand and the proximity grasp, without considering its movement in space; in the second phase, the training involves positioning the robotic hand near the object to be lifted, taking into account two specific subtasks: top grasping and side grasping. In this phase, the policy learned in the first phase is used for the robotic hand. The use of demonstrations enabled effective learning of positioning based on the communicated task to the agent (from above for top grasp and from the side for side grasp), utilizing a very simple reward function (1 for success, 0 for failure), thereby leveraging the expert's movements as the primary guide, especially in the early training phases. The results demonstrated the effectiveness of the method, yielding very satisfactory outcomes both in the same conditions as the training and, particularly in the first phase, in previously unseen conditions.

The second research area, on the other hand, concerns attentional mechanisms and their models used in machine learning. Specifically, I have explored in the literature the application of these models in reinforcement learning, particularly in the presence of multi-modal observations. Over the years, various attentional mechanism models have emerged in machine learning, drawing inspiration from neuroscience. These models have been applied across diverse areas, including image classification, translation, and question answering. In reinforcement learning, attention mechanisms have been integrated into systems to enhance focus on relevant visual features based on rewards. Different approaches, such as self-attention and multi-attention mechanisms, have been explored to improve learning efficiency. Within the scope of language-conditioned reinforcement learning, researchers have investigated how to align natural language instructions with the agent's observations, aiming for multimodal attention that enhances both performance and interpretability in task execution. In my research, the interaction between text and visual attention models is investigated within a reinforcement learning framework, where robotic agents are trained to accomplish tasks specified by natural language sentences. In the work accepted at AIxIA 2024, I studied the interaction between tasks specified in natural language and the visual features observed in environments defined in BabyAI, a MiniGrid-based platform featuring grid-simulated scenarios and tasks expressed in a synthetic language called Baby Language. Despite its simplicity, this language allows for a combinatorially rich set of 2.48×10^{19} possible instructions. The main objective of the research was to enable the learning of tasks specified in

natural language through reinforcement learning, utilizing a form of multimodal attention. To this end, the proposed attention mechanism allows the agent to learn per-word attention maps and define a word attention. In the experiments conducted, I examined the effectiveness of the proposed mechanism for the 'go to' and 'pick up' tasks, achieving satisfactory results and good generalization capability even on larger maps than those used during training. In the paper accepted at the AIRO 2024 workshop, I extended this idea by introducing a form of task attention, working with tasks structured with 'before' and 'after', and enabling the agent to learn a masking of the words related to tasks that have already been completed, allowing it to focus solely on those related to tasks yet to be completed. Once again, the learning process is incremental, starting from the policy obtained after training on a single task using the method described previously, and then undergoing further training with structured tasks by adding the module responsible for task attention. The filtering of words achieved with the proposed mechanism allows the agent to achieve better performance in shorter time frames.

During my literature review, I collected studies on the applications of transformers and graph neural networks in the context of reinforcement learning, with the aim of integrating these models into my future research. The application of transformers, for instance, could be particularly interesting for incorporating natural language to specify tasks, while graph neural networks may serve as a powerful tool for representing internal states or task graphs. Additionally, I also explored methodologies related to multi-agent reinforcement learning and multi-agent learning from demonstration. This knowledge will be valuable for future research efforts within the SpaceItUp project, particularly concerning the task of multi-robot space exploration.

My current research continues to focus on these two areas. I am working on enhancing the simulation environment developed for obtaining demonstrations in virtual reality for robotic manipulation tasks, aiming to achieve maximum generalization for the use of robotic manipulators. Additionally, I am transitioning the environments I developed in Unity to the Gym format, which will provide greater flexibility for developing reinforcement learning and imitation learning algorithms. As for the other area of research, the immediate next step involves applying the policies used on a real robot for exploration and object retrieval in a realistic environment, mapping low-level information from the robot to an abstraction compatible with the grid-based environment used during training. Furthermore, I will explore and adapt these methodologies to multi-agent scenarios in my future research.

4. Research products:

Workshop paper: G. Rauso, R. Caccavale, A. Finzi, “Learning Robotic Manipulation Tasks based on Incremental Demonstrations in a Virtual Environment”. In: Proceedings of the 10th Italian Workshop on Artificial Intelligence and Robotics co-located with the the 22nd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2023). AIRO 2023, Rome. CEUR Workshop Proceedings, vol 3686; Published 2024.

Conference paper: G. Rauso, R. Caccavale, A. Finzi, “Incremental Learning of Robotic Manipulation Tasks through Virtual Reality Demonstrations”. In: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS 2024, Abu Dhabi. Accepted and presented.

Conference paper: G. Rauso, R. Caccavale, A. Finzi, "Combined Text-Visual Attention Models for Robot Task Learning and Execution". In: 23rd International Conference of the Italian Association for Artificial Intelligence. AIxIA 2024, Bolzano. Accepted.

Workshop paper: G. Rauso, R. Caccavale, V. Lippiello, A. Finzi, "Integrating Text-Visual and Task Attention for Language-Guided Robot Learning". In: 11th Italian Workshop on Artificial Intelligence and Robotics co-located with the the 23rd International Conference of the Italian Association for Artificial Intelligence (AIxIA 2024). AIRO 2024, Bolzano. Accepted.

Prototype software: I developed a simulated environment in Unity to control the joints of a robotic hand in virtual reality through hand tracking, as well as to control the spatial positioning of the robotic hand. In this environment, I recorded demonstrations used for imitation learning.

5. Conferences and seminars attended

Conference: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems - IROS 2024, Abu Dhabi, 14-18 October 2024; presented the paper "Incremental Learning of Robotic Manipulation Tasks through Virtual Reality Demonstrations".

6. Activity abroad:

None.

7. Activity in partner companies:

None.

8. Tutorship

None.