Enhancing Robotics Education with Reinforcement Learning in Task Planning

Elena Caselli, Michele Polito, Valerio Cornagliotto, Elisa Digo, Laura Gastaldi and Stefano Pastorelli

Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Turin, Italy
elena.caselli@polito.it

Abstract. In addition to conventional industrial tasks, robots are increasingly used in daily service activities for different purposes. Accordingly, in the last decades, robotics-based education has been involved by a rapid development. This paper presents a challenge between a human and a *cobot* consisting in a ball-throwing task aiming to a target. A Reinforcement Learning (RL) algorithm was exploited in virtual simulation to train the robotic arm and optimizing throwing parameters such as direction and release speed. Outcomes demonstrate how RL can be exploited to train robots to perform complex tasks, improving their performance through iterative feedback on throwing accuracy. The educational aim of this activity is to inform participants about various aspects of robotics and artificial intelligence through an interactive experience and hands-on approach. After detailing the physical set-up and the technical implementation of the training process, the paper focuses on the educational impact of the proposed activity during informative events and its potential extensions in various fields.

Keywords: SDG4 · Trajectory Planning · Human-Machine interaction · Reinforcement Learning · Education

1 Introduction

Robotics technology is facing significant new challenges, from executing repetitive tasks in highly structured industrial settings [1] to performing less defined tasks in daily household environments [2]. The rapid advancement of Artificial Intelligence (AI) is transforming the robotic field, expanding its aims and applications in both industry and everyday life [3, 4]. In particular, the integration of AI in collaborative robotics is a powerful method to increase performance and adaptability of human-robot systems. Moreover, in line with the principles of the fourth Sustainable Development Goal (SDG4) of the 2030 Agenda [5], the combination of AI and robotics represents an inclusive and high-quality tool for educational purposes, fostering the development of essential technical skills for employment. Accordingly, AI-based robotics in education is rapidly growing, offering interactive methods to introduce complex concepts in various topics to an academic audience. In this context, practical experiments can reinforce and clarify the theoretical aspects [6–8]. Indeed, hands-on experience and system integration can strengthen the understanding of basic engineering concepts [8], increasing the impact of practical robotics education for both students and young researchers from companies [9].

The educational perspective of this work consists in the development of a challenge between the human and the robot in a ball-throwing task, which serves as a practical demonstration of basic robotic principles allowing the participant to directly experience the complexity of the task. Both the robot and the human are asked to drop a ball into a target cylindrical container placed in a position defined by the opponent. To achieve this goal with the robot, a Reinforcement Learning (RL) algorithm was exploited to train the robotic arm. Indeed, RL stands out for its unique ability in empowering robots to adapt autonomously through a *trial-and-error* approach. Moreover, the combination of robotic and deep learning techniques offers a concrete example of the RL paradigm based on the reward score maximization. The RL training algorithm was implemented and performed in a virtual simulation environment and then results were tested in real-world scenario. This paper provides an insight into the technical tools and methodologies which enabled the robot to achieve high accuracy in performing the throwing-ball task. Moreover, the educational impact of the proposed activity is highlighted by describing its realization during informative events, with a particular focus on its potential applications in various contexts.

2 The technical challenge

2.1 Experimental Set-up

The proposed experimental activity consists in a human-versus-robot challenge where both the human and the robotic arm are asked to score points by throwing the ball into a cylindrical container on the opposite side of a table. In detail, after the human places the target on the nearest half of the table surface in a self-selected random position, the robot is asked to throw the ball to hit the target, and vice versa.

As shown in Fig. 1, the instrumentation adopted for the experimental set-up included a collaborative robotic arm with 6 degrees of freedom (UR3, Universal Robots, Denmark) and a laser sensor (RPLIDAR A2, SLAMTEC, Shanghai, China). The robot and the laser sensor (LIDAR) were placed at the extremity of a 2 meters long table. The surface of the table was divided into two distinct playing fields for the game, one for the *cobot* and one for the human. Each part included three separate rectangular areas at different distances to increase task complexity (blue, orange and yellow). The setup also included two cylindrical containers (11 cm of diameter), which served as the "goals" to be reached by the human and the robot throwing their own ball (4 cm of diameter). A monitor was added to display real-time score updates for both competitors.

A gripper tool was mounted on the robot's end effector, which was controlled via the dedicated UR Polyscope software. The tool extensions were 3D-printed for the gripper fingers to improve accuracy in holding the ball securely. The LIDAR was integrated into the setup to identify the correct position of the target. In Fig. 1 the lidar sensor is highlighted with a red shade. This sensor is inserted in a cylindrical box with a thin open slot which allows the emission of the laser to cover the table surface, avoiding other visual disturbances.

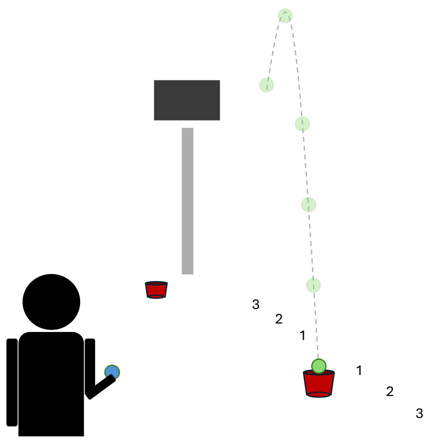


Fig. 1. Experimental set-up.

2.2 RL Algorithm

The *cobot* ability to perform the throwing task demonstrates a parallelism with the human capability. Indeed, the human gaze role in identifying the exact position of the target on the table is played by the LIDAR combined to the *cobot*. Then, the AI process associated to the use of the *cobot* defines the launch parameters as the brain normally does for the human. Finally, the robotic arm physically executes the launch simulating the human gesture. In the specific case of this challenge, before testing results in a real-world environment, the training process of the robot was entirely conducted in simulation exploiting a RL framework. The simulation approach permits to execute many repetitions of the task in a short period of time. The robot was modelled as a rigid body serial link structure considering the nominal standard Denavit-Hartenberg parameters. Even if this approach introduces an error in the trajectory of the real robot, it represents a good approximation. To compensate for this error, it should be necessary to execute a preliminary fitting to obtain a representative model of the robot [10]. However, adopting a nominal robot model provides a more general solution. As shown in Fig. 2, the structure of the learning process consisted of two main components: the RL agent and the simulation environment. More in detail, the RL agent represents the decision-making of the algorithm, implemented using a Twin Delayed Deep Deterministic Policy Gradient – TD3 network. It observes the state of the environment based on the target position, and it generates actions defining the trajectory. The agent actions are the angular velocity of the robot joint 4, and the toss direction, which is imposed by the joint 1 of the robot. Hence, the end effector performs a planned path with adjustable speed and orientation. During the path, the ball is released when the end effector speed vector reaches an angle of 45 deg relative to the table surface. This release angle maximizes the throwing range. Related to each action, the RL agent receives a reward depending on the reward function *f(error)*, which is related to the distance between the current position of the target and the final position of the dropped ball estimated through ideal parabolic trajectory (*error*). The reward function allows the agent to iteratively improve its actions by minimizing the residual error. In the simulation environment, the robot executes the trajectory defined by the actions specified by the RL agent. The environment updates the agent with the current state of the system (*observation*) and provides a reward (*reward*) quantifying the successful or unsuccessful agent outcome of each action. The reward function is designed as a parabola-like function with a maximum at zero *error* and bounded below a negative value. During the training, the observation fed to the agent includes the coordinates of the target placed in a random point of the grid.

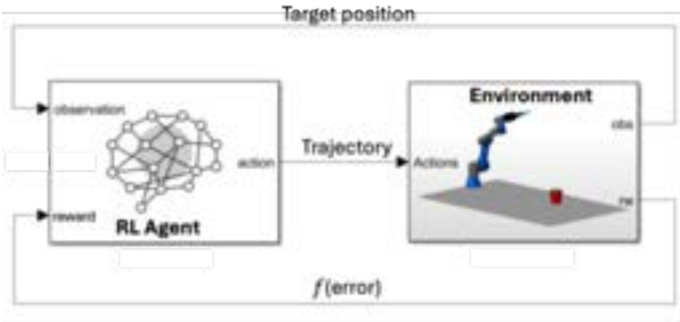


Fig. 2. RL algorithm structure.

The area used for the training process was discretized by a grid with a step of 5 cm, as shown in Fig. 3a. In the real environment, the target coordinates are provided by the LIDAR, which detects points along the external surface of the cylindrical container allowing the identification of the target as the center of the cylinder (Fig. 3b). More in detail, the calculation of the center is based on a least-squares optimization process to achieve a more robust measurement. Additionally, the optimization is constrained to ensure that the resulting circle is consistent with the dimensions of the cylindrical container. This simulated training approach allows the RL agent to refine the robot's control policies in a virtual setting, reducing the training time and the need for physical testing and optimizing performance before real-world implementation.

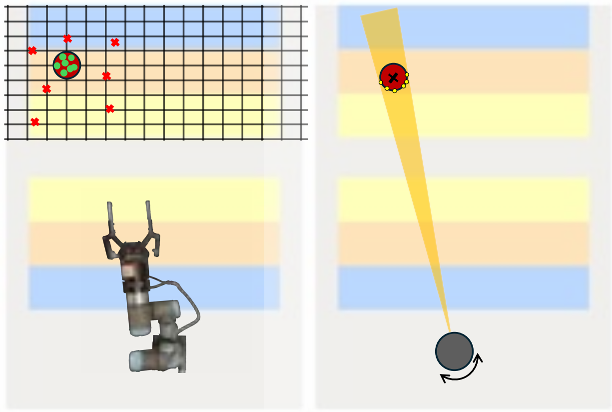


Fig. 3. a) Target position discretization for training; **b)** Target position identification process.

2.3 Robot control and trajectory planning

The control architecture is divided into several parts as shown in Fig 4. The first step, performed in MATLAB, involves calculating the launch parameters. In the MATLAB script, the target coordinates are estimated from the LIDAR data as described before. The target coordinates are then used by the RL agent to compute the optimal launch parameters (i.e. the velocity of joint 4 and the position of joint 1). Additionally, the position of joint 4 corresponding to the direction of end-effector velocity vector of 45 deg relative to the ground is calculated. This information will later be used to command the gripper's opening.

Once the parameters are estimated, they are given as input to a Python script which facilitates communication between the PC and the robot and manages the high-level coordination of the robot actions. Specifically, the script synchronizes the task steps, monitors the robot status, ensures step completion, and sends signals to transition to the next step, along with the necessary data for execution. During the launch phase (toss), the script gives the position of joint 1, and the peak velocity of joint 4 required to plan the trajectory as input to the robot controller. Furthermore, the script monitors the state of joint 4 and triggers the robot controller to open the gripper once the release value is reached. Communication between the PC and the robot occurs at a rate of 125 Hz using the Real Time Data Exchange protocol of Universal Robots.

The joints low level control is handled directly by the robot controller, ensuring the planned trajectories are performed as intended. In the proprietary Universal Robot environment (PolyScope), the various tasks, such as ball picking, container placing, and tossing are programmed. The first two tasks are planned in Cartesian space as point-to-point linear motion, while the tossing motion is planned in joint space using a trapezoidal velocity profile. Specifically, the initial and final joint positions are set, then the velocity profile of the fourth joint is derived, with the angular velocity obtained from the RL and with a predefined constant acceleration. Once the velocity trapezoid for joint 4 is defined, trapezoidal profiles for the other joints are derived by synchronizing the start and end times for all joints, ensuring a coordinated movement.

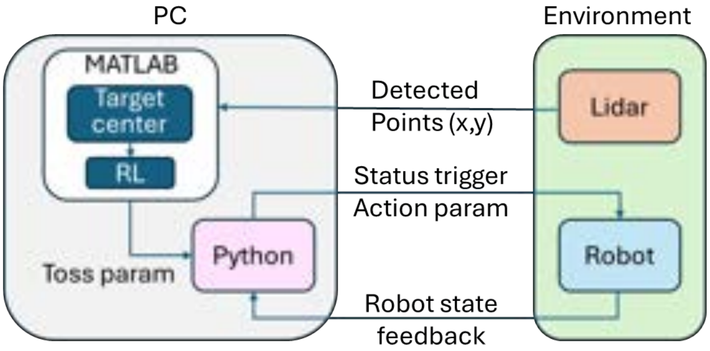


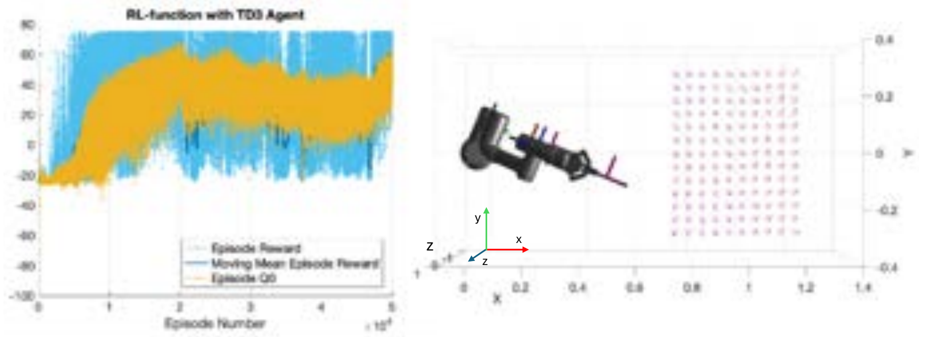
Fig. 4. High level Control architecture.

3 Experimental results and practical applications

During the training, the RL agent completed 50,000 simulated trials, allowing it to explore a wide range of actions and refine its strategy to optimize performance (Fig. 4a). As a result, thanks to the training process, the agent achieved a launch error minor than 2 cm (Fig. 4b). The iterative process enables the agent to develop an optimal strategy to minimize the error and to improve the accuracy over time learning by rewards. After the design and validation of the set-up, this activity was tested on many occasions related to the academic context such as an educational demonstration for students of Mechanical Engineering at Politecnico di Torino (Turin, Italy).

Moreover, the same activity was also proposed in three off-campus educational events offering a unique space to share and discuss advancements in science and technology on contemporary topics:

- UNIGHT – European Night of Researchers [11] in September 2023 in Turin, an annual appointment for researchers in which the main objective is to share knowledge of different fields with the community (Fig. 5a).
- Biennale Tecnologia [12] in April 2024 in Turin, biennial event with the goal of exploring the relationship between technology and society (Fig. 5b).
- Settimana Arcobaleno at Liceo Statale Edoardi Amaldi in February 2024 in Novi Ligure (Alessandria), an orienteering activity for high school student (Fig. 5c).



a) b)

Fig. 4. a) RL agent training process, where the episode reward is represented in light blue, its moving mean in dark blue, and the learning score Q0 in yellow; **b)** Launch error during training

Participation in the mentioned events provided a valuable opportunity for dialogue and exchange with the public and other researchers, with a focus on innovation. Moreover, this work represents an example of how academic research can contribute to addressing future challenges, combining innovative approaches and advanced techniques in different fields. The test bench also served as a platform to explain fundamental concepts related to robotics and artificial intelligence. Specifically, it includes: (i) the robotic spatial awareness through target localization, (ii) the robotic control and synchronization, (iii) the RL paradigm illustrated by the challenge, and (iv) insights into the differences between simulation and real-world applications.

Finally, the possibility to challenge the robot is a way to increase the interest of students and the community towards the potentiality of *cobots* and the concrete exploitation of AI techniques.

a) b) c)

Fig. 5. a) UNIGHT – European Night of Researchers (2023); **b)** Biennale Tecnologia (2024); **c)** Settimana Arcobaleno at Liceo Statale Edoardi Amaldi (2024)

In conclusion, the proposed activity is an example of how a robotic arm can accomplish a complex task through training based on RL techniques. One of the strengths of *cobots* is their compactness, though this can be a limitation for certain applications. In this study, the goal was to extend the workspace of the robotic arm, offering an engaging way to explore the potentiality of *cobots* in some real-world applications:

- leisure activities raising community awareness of the use of technology in everyday life contexts;
- sorting of waste in a urban environment, automating the process and facilitating human operations with a reduced propensity for workload musculoskeletal diseases;
- sorting of parcels and other non-fragile materials in an industrial scenario, reducing production times and alleviating the physical workload of human operators;
- rehabilitative activities with the *cobot* directly launching the ball to the human, improving the gripping and coordination capabilities in pathologies such as hemiplegia and cerebral palsy [13].

References

1. Digo E, Caselli E, Polito M, Antonelli M, Gastaldi L, Pastorelli S (2023) Test–Retest Repeatability of Human Gestures in Manipulation Tasks. Applied Sciences (Switzerland) 13:. https://doi.org/10.3390/app13137808
2. Wu M, Kandra A, Shen X (2014) An education-oriented robotic anthropomorphic hand system. In: Proceedings of the ASME 2014 Dynamic Systems and Control Conference. ASME
3. Dogmus Z, Erdem E, Patoglu V (2015) ReAct!: An interactive educational tool for AI planning for robotics. IEEE Transactions on Education 58:15–24. https://doi.org/10.1109/TE.2014.2318678
4. Gomes de Souza G, Soares de Lima BL, Vieira dos Santos R, Monteira de Almeida FJ (2021) Development of a Robotic System with Movements Based on Computer Vision Detection. In: Smart Innovation, Systems and Technologies. Springer Science and Business Media Deutschland GmbH, pp 363–372
5. United Nations HOMEPAGE: Goals. https://sdgs.un.org/goals. Accessed 15 Nov 2024
6. Ferri BH, Ahmed S, Michaels J, Dean E, Garyet C, Shearman S (2009) Signal Processing Experiments with the LEGO MINDSTORMS NXT Kit for Use in Signals and Systems Courses. In: 2009 American Control Conference. IEEE Xplore
7. Nagai K (2001) Learning while doing: Practical robotics education. IEEE Robot Autom Mag 8:39–43. https://doi.org/10.1109/100.932756
8. Otaran A, Tokatli O, Patoglu V (2016) Hands-On Learning with a Series Elastic Educational Robot. In: 10th International Conference, EuroHaptics 2016. Springer, pp 3–16
9. Nagai K (2001) Enabling Students to Acquire the Knowledge and Experience Necessary to Produce Advanced Technologies. IEEE Robot Autom Mag 39–43

10. Cornagliotto V, Digo E, Pastorelli S (2021) Using a robot calibration approach toward fitting a human arm model. In: Mechanisms and Machine Science. Springer Science and Business Media B.V., pp 199–207
11. UNIGHT HOMEPAGE. https://unightproject.eu/it. Accessed 12 Nov 2024
12. Biennale Tecnologia HOMEPAGE. https://www.biennaletecnologia.it/. Accessed 12 Nov 2024
13. Raharjo S, Wibowo A, Andiana O, Ibrahim, Pelana R, Susiono R, Sujarwo, Antoni R, Amalia EF (2021) Throw and catch the ball games for children with dyskinesia-type cerebral palsy to improve eye-hand coordination movements. Journal of Physical Education and Sport 21:2419–2424. https://doi.org/10.7752/jpes.2021.s4325