

Reinforcement Learning-based Switching Controller for a Milliscale Robot in a Constrained Environment

Abbas Tariverdi*, Ulysse Côté-Allard[†], Kim Mathiassen[§], Ole J. Elle^{†‡},
Håvard Kalvøy[¶], Ørjan G. Martinsen*[¶], Jim Tørresen[†]

*Department of Physics, University of Oslo, Oslo, Norway

[†]Department of Informatics, University of Oslo, Oslo, Norway

[‡]The Intervention Centre, Oslo University Hospital, Oslo, Norway

[§]Department of Technology Systems, University of Oslo, Oslo, Norway

[¶]Department of Clinical and Biomedical Engineering, Oslo University Hospital, Oslo, Norway

Abstract—This work presents a reinforcement learning-based switching control mechanism to autonomously move a ferromagnetic object (representing a milliscale robot) around obstacles within a constrained environment in the presence of disturbances. This mechanism can be used to navigate objects (e.g., capsule endoscopy, swarms of drug particles) through complex environments when active control is a necessity but where direct manipulation can be hazardous. The proposed control scheme consists of a switching control architecture implemented by two sub-controllers. The first sub-controller is designed to employ the robot’s inverse kinematic solutions to do an environment search for the to-be-carried ferromagnetic particle while being robust to disturbances. The second sub-controller uses a customized rainbow algorithm to control a robotic arm, i.e., the UR5 robot, to carry a ferromagnetic particle to a desired position through a constrained environment. For the customized Rainbow algorithm, Quantile Huber loss from the Implicit Quantile Networks (IQN) algorithm and ResNet are employed. The proposed controller is first trained and tested in a real-time physics simulation engine (PyBullet). Afterward, the trained controller is transferred to a UR5 robot to remotely transport a ferromagnetic particle in a real-world scenario to demonstrate the applicability of the proposed approach. The experimental results on the UR5 robot show an average success rate of 98.86% over 30 episodes for randomly generated trajectories, demonstrating the viability of the proposed approach for real-life applications. In addition, two classical path finding approaches, Attractor Dynamics and the execution extended Rapidly-Exploring Random Trees (ERRT), are also investigated and compared to the RL-based method. The proposed RL-based algorithm is shown to achieve performance comparable to that of the tested classical path planners whilst being more robust to deploy in dynamical environments.

Note to Practitioners—Deep reinforcement learning methods have been widely applied in computer games and simulations. However, employing these algorithms for practical, real-world applications such as robotics becomes challenging due to the difficulty of obtaining training samples. This paper predominantly focuses on bridging the gap between simulations and the real-world implementation of a reinforcement learning algorithm for a robotic application in the context of miniaturized drug delivery robots and robotic capsule endoscopes. This paper presents the derivation and experimental validation of a reinforcement learning-based algorithm for controlling a magnetically-actuated small-scale robot within a simplified model of the large intestine in the presence of disturbances. We demonstrate the possibility of training a high-fidelity reinforcement learning algorithm fully within a simulated environment before deploying it as-is in

a real-world scenario by carrying out different experiments and simulations. Implementing the presented control framework complements a large body of this work, and the results offer a feasibility study of using reinforcement learning algorithms in practice.

Index Terms—Magnetic Manipulation, Reinforcement Learning, Rapidly-Exploring Random Trees, Attractor Dynamics, ERRT, Switching Control, Targeted Drug Delivery

I. INTRODUCTION

EFFICIENT drug delivery is a difficult task in the medical industry [1], and traditional approaches such as pills and intravenous therapy have been the primary method of drug delivery for decades [2], [3]. One drawback of these traditional administration methods is that a drug cannot be transported directly to targeted tissues/organs to produce a marked effect. On the other hand, carrier-based drug delivery systems (e.g., small-scale delivery robots) could be better suited to accomplish such precisely targeted delivery task [4].

In recent years, there have been substantial efforts to develop soft miniaturized robots capable of navigating the human body and delivering drugs directly to a tumor or other disease sites precisely and non-invasively.

One of the barriers to small-scale drugs moving toward diseased tissues in the bloodstream is the blood vessel’s lining [5], [6]. Therefore, manual delivery processes might still not be as effective as it needs to be [4], [7]. To overcome this barrier, a promising approach is to actively manipulate drugs or swarms of drugs to reach disease sites and push the particles out of the bloodstream and into target sites [6]. One potential solution to address this issue and increase the efficiency of the delivery process is to robotize the process combined with a remote magnetic actuation mechanism [7]. This allows surgeons to have better control over the navigation of drugs even when faced with biophysical barriers, which would otherwise prevent the circulation of drugs.

The miniaturization of conventional robots is limited mainly by the mechanical structures required to allow them to navigate. By contrast, magnetic actuation is a mechanism that can remotely and wirelessly actuate robots without necessitating specialized structures directly in the robots, thus substantially reducing their size and complexity. Reducing the size of

robots enables them to move to difficult-to-reach areas inside biological bodies by applying an external magnetic field that harmlessly penetrates patients' body and applies wrenches on the robots.

Magnetic manipulation of micro/millscale robots with an application such as magnetic drug delivery has been an increasingly popular approach where drug particles in the body are carried and manipulated by external magnetic fields to reach a targeted location [8]–[11]. Moreover, small-scale robots/particles can essentially be seen as magnetic actuation points for continuum manipulators [12], [13]. In other words, devising control and path planning algorithms for micro/millscale robots can eventually contribute to making magnetic continuum manipulators operational in a (semi-)autonomous mode [14].

Remote and robotic-assisted particle or micro/milli-robots manipulation by external magnetic fields has been investigated through traditional control methods, which are reviewed. The work in [15] proposes PID controls and an adaptive control law to remotely operate a microrobot with 3 DOF in an enclosed environment. The designing of classical proportional controllers for controlling two micro-scale magnetic particles in two dimensions while considering interaction wrenches between the particles has been investigated in [16]. Driller et al. in [17] propose a vision-based proportional feedback controller for motion control and path following a set of two microrobots in three dimensions using magnetic gradient pulling. In the work [15]–[17], particles move in an obstacle and disturbance-free space, and the proposed controllers are not capable of dealing with uncertainties in the case of considering a complex environment. Zarrouk et al. [18] address an open-loop control of a particle by using an array of permanent magnets mounted on a robotic arm end-effector. This work does not consider a localization system to recover the actual position of the particle, and therefore, the proposed open-loop method cannot efficiently accomplish a task such as drug delivery within a noisy environment like the human body. The previous work [19] used a configuration of three rotating permanent magnets and employed a vision-based proportional control strategy in which gain coefficients are tuned by multiple trials. Although a vision-based localization sensor is used and the magnitude and direction of the applied magnetic field are controllable in this work, it lacks the ability to reject disturbances. In addition, if an obstacle blocks the particle, the considered task cannot be accomplished. Keller et al. in [20] propose a human-in-the-loop navigation method to control positions and orientations of a magnetically guided capsule endoscopy. The presented method in [20] includes a human in the control loop to manually navigate a capsule endoscope for stomach examinations with multiple functionalities tilting and jumping etc. The work in [21] investigates two open-loop control methods for positioning two and three microrobots in two-dimensional space. The microrobots are geometrically designed to respond uniquely to the same applied magnetic fields. Therefore, individuals and subgroups of microrobots can locomote to desired positions by controlling the magnetic field. The method in [21] is applied to micro-scale particles (all dimensions under 1 mm), and the global positioning of

particles is done in 2-D space using an offline parameter fitting method. Pieters et al. in [22] propose a model-free vision-based kinematic controller for a microrobot constrained to non-holonomic motions. Kim et al. [23] considers a vision-based closed-loop control of an array of magnets on an arbitrary predefined path in three dimensions. The authors in [24] propose and compare both open and closed-loop vision-based controllers for steering a helical microswimmer. The approaches described in [21]–[24] are not equipped to cope with undersecretaries (obstacles and perturbations) within the workspace. Although the modern optimal and reinforcement learning control methods for robots have attracted particular attention of scholars in the field of robotics in recent years [25]–[27], real-world implementation of high-fidelity optimal control and RL strategies have not been the point of focus.

An RL-based model-free control of drug dosing for cancer treatment is presented in [28]. The authors use a Q-learning method to schedule the dosing to maximize immune cells for patients. Nevertheless, the presented method is not categorized among targeted drug delivery approaches. In the paper [29], a Q-learning method with the actor-critic architecture and the deep deterministic policy gradient algorithm is investigated to navigate a particle within an undisturbed 2-D space. The approach uses a linearized model of the system around an equilibrium point to prepare the necessary feedback signals for the controller. However, the approach is not tested in the presence of perturbations. Xu et al. in [30] proposes an RL control of a flexible magnetically actuated joint. The study focuses primarily on the tilt motion control for the tip of the instrument, rather than the navigation of the entire joint, as the primary objective.

Reinforcement learning algorithms have also been used in a variety of applications, including collaborative assembly tasks [31], space capture missions [32], and 3D navigation in cluttered environments [33]. Zhang et al. in [31] employ a deep deterministic policy gradient algorithm for the task allocation problem in human-robot collaborative assembly tasks. Jiang et al. in [32] also modify a deep deterministic policy gradient algorithm to develop an adaptive tracking controller based on a highly nonlinear dynamic model of a three-module cable-driven continuum robot. Furthermore, Hu et al. in [33] propose a reinforcement learning-based point-to-point navigation algorithm for a mobile robot in a 3D constrained terrain environment.

By investigating the existing results on magnetic control of small-scale robots (as were reviewed earlier), the following issues are noteworthy: The existing studies mainly employ open-loop or closed-loop control mechanisms that are incapable of accomplishing a given task in uncertain environments. In other words, in practice, the methods are not robust in the presence of disturbances. In addition, the existing studies do not take into account any obstacles in the design of controllers or trajectory planners, so these methods are not suitable for realistic scenarios.

It is worth mentioning that coping with disturbances and uncertain environments, together with tracking trajectories or set-point regulations, are the primary factors in controller design and trajectory planners. An interesting point in the

magnetic control of milliscale robots is that the mechanism allows teleoperating particles wirelessly. Usually, remote control mechanisms are used when working in an environment that is hazardous or not feasible due to confined or difficult-to-reach areas. This makes the control of magnetic particles more prone to uncertainties.

The general motivation behind the presented method is bridging the gap between simulations and the real-world implementation of a reinforcement learning algorithm for a robotic application. When it comes to obstacle avoidance and perturbation handling scenarios, many approaches have been proposed to solve the challenges. Some methods such as [34]–[36] achieve local optima and therefore can not guarantee a feasible obstacle-free path. Some others [37]–[39] do global searches for a feasible, valid path, yet the heavy computational costs prevent them from reacting properly to pop-up obstacles.

RL-based methods can offer a hybrid behavior by learning a dynamic, time-varying environment. These approaches do re-planning and path deformation on-the-fly to avoid suddenly appearing obstacles and find a valid global path to a determined target. In addition, compared to those methods in which global explorations happen online, the training of RL methods is offline; hence, they are reactive during the execution of a task. In other words, if the deformation of a path does not result in a valid trajectory toward a target, a new valid path is calculated. It is worth mentioning that the quality of the behavior is highly dependent on the samples used for offline training. Also, the system’s performance can only be guaranteed statistically. At the same time, it is feasible to generalize and implement the approach for different environments with varying forms of obstacles and perturbations, provided that more samples of environments are used for offline training and proper tweaking of the controller’s hyperparameters is considered.

To evaluate results of the RL-based mechanism, two classical approaches, Attractor Dynamics [40] and execution extended Rapidly-exploring Random Trees (ERRT) [41] are tested through conducting simulations and experiments in which the results confirm that the RL-based method achieves comparable performance while being a more robust and generic solution to deploy in dynamical, complex environments.

Based on the abovementioned issues, more practical approaches to magnetic control of small-scale robots require investigation. Although designing online path planners with appropriate robust controllers is a problem-specific approach, deep Reinforcement Learning (RL) methods can handle both the trajectory planning and control tasks simultaneously and can be generalized to be used in various workspaces. This paper proposes to employ a deep RL algorithm for magnetic position control of an object in a constrained environment. To the best of the authors’ knowledge, this is the first time that magnetic control of milli-sized (ferro)magnetic particles using an RL algorithm capable of avoiding obstacles and rejecting disturbances is addressed. Therefore, the main contribution of this work are as follows.

- Implementing an RL-based switching framework for dynamic control of a particle in the presence of obstacles in

two dimensions with visual feedback and using a robotic arm in a real-world scenario.

- Through simulations and experiments, the results of the proposed RL-based algorithm and classical approaches such as Attractor Dynamics, and execution extended Rapidly-exploring Random Trees (ERRT) are compared, revealing the practical efficacy of the proposed methodology.

This paper is organized as follows: The problem statement is given in Section II. Preliminaries on the RL and Rainbow algorithms are provided in Section III. Section IV thoroughly describes our proposed RL-based, Attractor dynamics-based, and ERRT-based switching controllers. Section V is devoted to explaining the simulation and experiment setups, and the results from the simulations and real-world scenarios, while their associated discussion is presented in Section VI.

II. PROBLEM STATEMENT

Inspired by the literature on small-scale robot control and to address more practical problems (e.g., scenarios in the presence of obstacles and disturbances), this paper aims to propose and implement an RL-based switching framework for dynamic control of a particle in a constrained environment using a robotic arm and visual feedback. In this work, we have employed an RL algorithm in a switching control framework such that the first sub-controller is responsible for finding and tracking the magnetic particle through a camera mounted on the end-effector of a robotic arm. This sub-controller makes the whole system robust against disturbances. A second sub-controller makes the robot arm carry the particle through a constrained environment toward a target position. In addition, a simplified model of the large intestine in two dimensions has been employed as a constrained environment in which a milliscale robot is being carried remotely using an external magnetic field generated by a permanent magnet mounted on a robotic arm.

The strength and direction of a magnetic field produced by an attached permanent magnet on the end-effector are not considered to be adjustable, and due to interference introduced by the environment, a magnetic particle may not respond to the field and be carried within the environment. To guarantee that the particle responds to the external field, we assume that an external magnetic field is effective on a magnetic particle, i.e., a milliscale robot, provided that the particle is within a specific distance from the end-effector of the robotic arm.

As disturbances, we physically draw the magnetic particle away from the magnetic control zone during active control within the constrained environment—a simplified model of the large intestine—. In other words, disturbances are displacements in 2-D space. We assume that disturbances do not occur too quickly and/or with large accelerations. It is worth mentioning that in real-world scenarios such as the drug delivery application, disturbances such as movement of the GI tract or a fluid current running into a vessel or the tract might hold the particle back from a region where the wrenches of the end-effector’s magnet are effective.

III. PRELIMINARIES

RL deals with learning the behavior (policy) of an agent, which in this work is a robot that interacts with its environment to perform a task. It is an approach for solving control problems in which a control policy is learned through repeated trial-and-error interactions between the robot and its environment. In RL, the behavior of the robot is learned to maximize the expected sum of rewards provided by its environment through feedback signals. Assuming discrete time steps t , the robot performs an action $a_t \in \mathbb{R}^n$, $n \in \{1, 2, \dots\}$ that depends on the current state of the system $s_t \in \mathbb{R}^n$ and is also considered as a control action. This action result in a new state s_{t+1} , and the robot receives a reward r_{t+1} . This process is repeated, and the robot's goal is to learn the optimal policy π , which maximizes the expected cumulative rewards.

The traditional Q-Learning [42] is concerned with learning Q-values. Q-values are the expected accumulated rewards the robot receives when following a given policy starting from the state-action pair (s_t, a_t) . In the traditional Q-learning algorithm, we construct a Q-value table containing all the Q-values mapping between the possible states a robot can get by moving between the different states. This basic algorithm can be inefficient in sampling the state space and may get stuck in a local optima [43]. Deep Q-Networks (DQN) algorithm has been an important milestone introduced by [43]. The difference between DQN and its table version is that instead of having all the Q-values stored in a look-up table, they are represented as a multi-layered neural network such that, for a given state s , outputs a vector of action values $Q(s, \theta)$, where θ are the parameters of the network. Also, by approximating these Q-values via a neural network, one can consider action and state space that could not realistically be fitted in memory (mainly for embedded applications). This results in more significant generalization attributes and richer representation.

Despite its usefulness, several limitations of the DQN algorithm are now known, and many extensions have been proposed so far to enhance its speed or stability. Some limitations are as follows: Because of a max operator in the Q-learning equation, DQN may overestimate Q-values and then choose sub-optimal actions in some states. In addition, uniformly sampling from a replay memory in the experience replay mechanism is not optimal [44]. In other words, prioritizing samples in a replay memory based on some criteria, such as the error of the estimated Q-values and the actual Q-values, can help mitigate the estimation deviations [45]. Furthermore, by introducing a function (i.e., the advantage function) to DQN, comparing the goodness of actions can be done [46]. Moreover, unlike the DQN algorithm that finds out about the mean or max of Q-values, learning their distributions has shown a better performance [47]. In noisy networks, weight parameters are modeled as distributions. In other words, weight parameters are sampled from some distributions. However, in noisy networks, predicted actions are covered by distributions, and then finding the best action is more challenging which can be done by reducing the size of sampling distributions or data uncertainty over time [48]. The deep reinforcement learning community has made gradually

several improvements to the DQN algorithm [43] to cope with its shortages, for example, Double Deep Q-learning Networks (DDQN) [44], Prioritized DQN [45], Dueling DQN [46], Distributional DQN [47], Noisy DQN [48]. Finally, Rainbow [49] combines the improvements as mentioned above.

The main characteristics of the Rainbow algorithm are as follows: an experience replay memory is introduced, in which experiences are stored consecutively. Uniformly random sampling from this buffer removes correlation between the observations to some extent. In the employed Rainbow algorithm, an N -step replay buffer is implemented where the most recent N observations are stored. The N -step replay buffer is used in the N -step learning method [50]. Prioritized replay buffer [45] is another crucial feature used in Rainbow, where a criterion is used to measure how unexpected a transition is, and based on this, the most unexpected transition is replayed from memory, and weight and bias parameters are updated accordingly. Noisy networks are an alternative for the ϵ -greedy method used in DQN. In noisy layers, deterministic weight and bias parameters are replaced by a unit Gaussian distribution, and the applied noises to weight and bias parameters are independent. In Dueling architecture [46], there is a possibility to compare the goodness of all actions in a specific state by introducing advantage and value layers. Also, in [47], the authors discuss a method known as Categorical DQN, which learns the distribution of returns instead of the expected return, and using the proposed method, returns distributions satisfy a variant of Bellman's equation. In the implemented Rainbow, a Dueling architecture is integrated with the Categorical DQN method, and the last two layers of advantage and value are replaced by the noisy layers.

IV. METHODS: REINFORCEMENT LEARNING, ATTRACTOR DYNAMICS, AND ERRT-BASED SWITCHING CONTROLLERS

This section primarily discusses the proposed control architecture in which a switching control architecture comprises two controllers, i.e., Sub-controllers 1 and 2, and a switching law. Sub-controller 1 consists of the inverse kinematic of the robot and keeps the magnetic particle within the immediate vicinity of the robotic arm end-effector. Once the particle is far enough from the end-effector, Sub-controller 1 takes the robotic arm to the particle position, which is feedbacked into Sub-controller 1 from an optical tracking system. As soon as the particle is again close enough to the end-effector, an appropriate switching law makes Sub-controller 2 control the robot arm to carry the particle toward an objective inside the environment using a magnetic field produced by an attached permanent magnet on the end-effector.

A. Reinforcement Learning-based Approach

In the RL-based method, Sub-controller 2 is trained beforehand both as a trajectory planner to prevent the robot from crashing into obstacles and also as a controller, which makes the robot globally stable by generating trajectories that always carry the particle toward a target for an arbitrary initial position. Sub-controllers 1 and 2 and the switching law are shown in Figure 1.

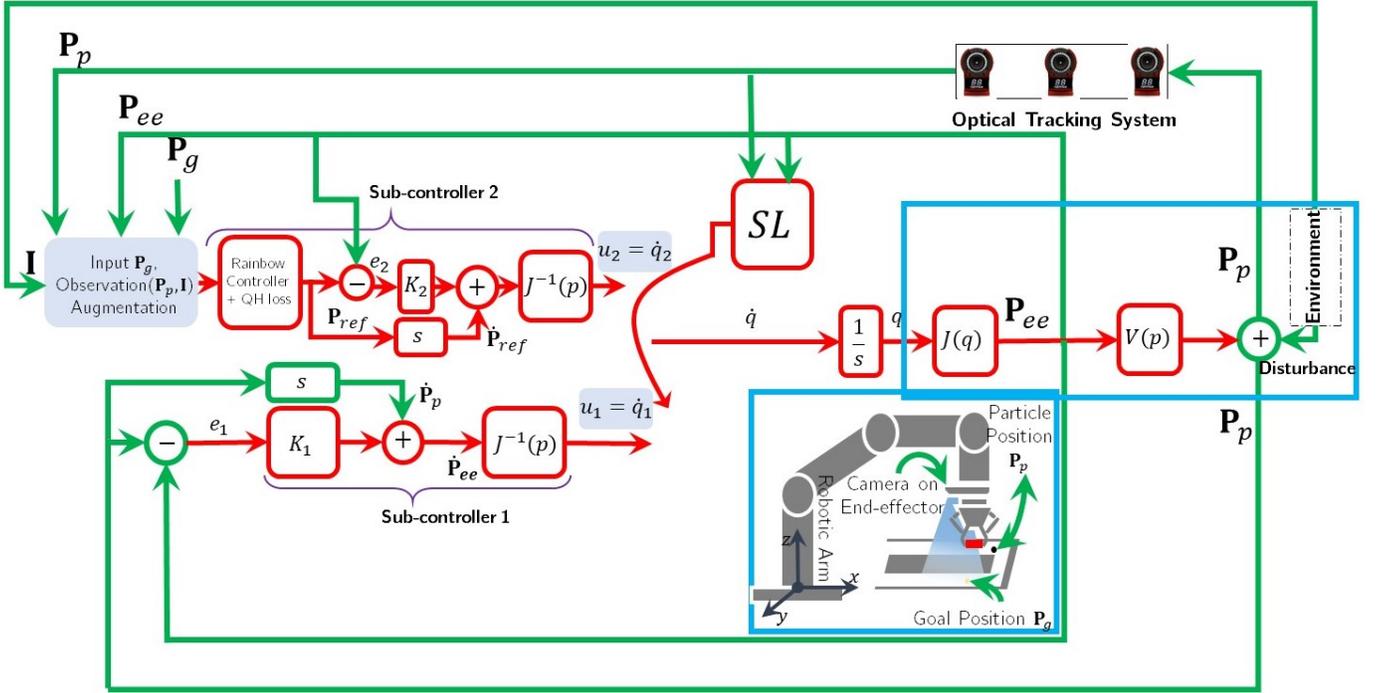


Fig. 1. The switching controller architecture consists of Sub-controllers 1 and 2 and a switching law (SL). SL is a supervisory algorithm that controls the switching between the various controllers. The observations or feedback signals are: the magnetic particle position obtained through the optical tracking system (\mathbf{P}_p), partial information (\mathbf{I}) of the constrained environment captured by an RGB camera on the end-effector, i.e., the RGB camera only covers a portion of the environment for the Sub-controller 2. s represents a derivative filter. In Sub-controller 1, K_1 represents a proportional gain for adjusting the asymptotic convergence of the controller, and in Sub-controller 2, K_2 represents a positive definite 3-by-3 matrix for adjusting the asymptotic convergence of the controller. K_2 is a simple model for the internal control of the robot. Joints are modeled as a single integrator which is represented by $\frac{1}{s}$. Inverse kinematic and kinematic of the robot is denoted by $J^{-1}(p)$ and $J(q)$, respectively. q and \dot{q} represent joint positions and angular velocities. $V(p)$ models the unknown magnetic interaction dynamic between the external magnet on the end-effector and the particle with the input \mathbf{P}_{ee} and output \mathbf{P}_p .

1) *Sub-controller 1*: Since there is no control on the strength and direction of a magnetic field produced by an attached permanent magnet on the end-effector and due to uncertainties and disturbances induced by the environment, the control system should be robust to losing the particle (when the particle is not responding to the field) and able to be automatically retrieved if such an event occurs. Sub-controller 1 is responsible for keeping the magnetic particle within the vicinity of the end-effector through the information it gets from an optical tracking system. In other words, the robotic arm should be capable of finding the magnetic particle or tracking the particle starting from an arbitrary initial position. When Sub-controller 1 is activated, the control-loop mechanism is depicted in Figure 2.

IKFast, The Robot Kinematics Compiler ([51], [52]), which analytically solves the inverse kinematics equations of the robotic arm, UR5 in this work, is employed as Sub-controller 1 to map Cartesian space of the end-effector to the joint space of the UR5. Furthermore, the locus of the end-effector remains parallel to the plane where the particle is manipulated, i.e., the angle between the plane of work and the z-axis should remain unchanged.

When a disturbance occurs, Sub-controller 1 will be activated and the robot basically employs a kinematic control. Therefore, performance can be very satisfactory, provided that inserted disturbances are not too fast and/or do not require large accelerations. Also it should be noted that one of the

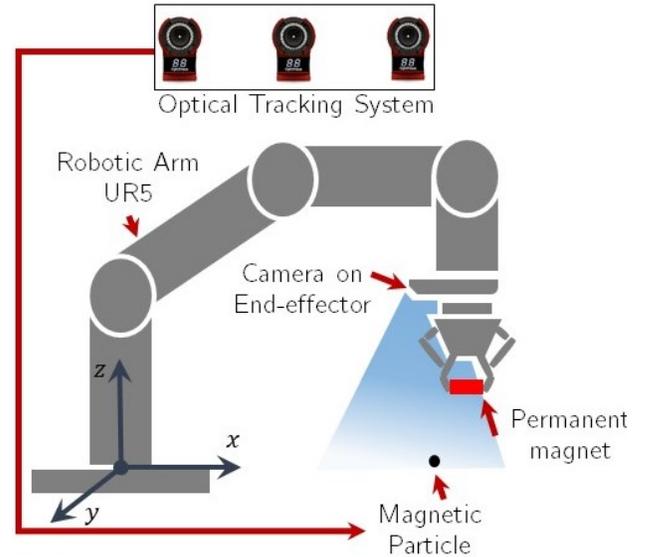


Fig. 2. The control scheme employs analytical inverse kinematic solutions when Sub-controller 1 is activated: The feedback coming from the optical tracking system consists of the particle position, and by employing the analytical inverse kinematic solutions, the robotic arm will locate the particle.

issues for using a kinematic control method is singularities. This implies that the robotic arm needs to be singularity-free within the workspace where the simplified structure of

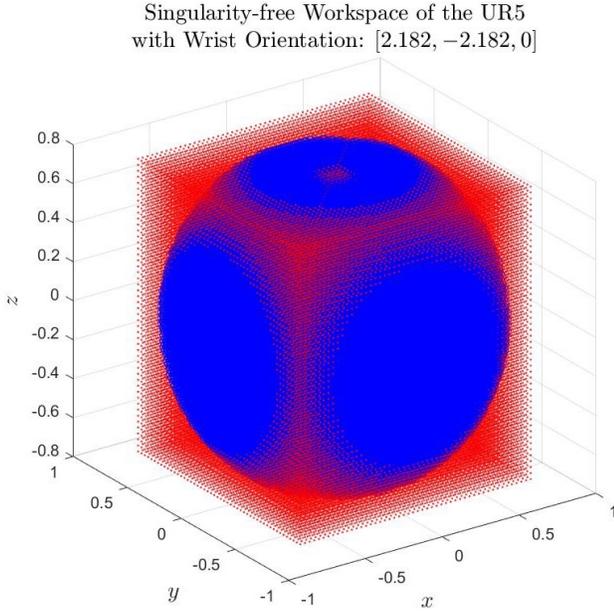


Fig. 3. The blue area shows the singularity-free workspace of the UR5 with the wrist orientation: [2.182, -2.182, 0]. Considering the locus of the end-effector when UR5 manipulates the particle within the working environment, the manipulator will not encounter a singularity (red-colored volume).

the large intestine is located. Outside of the workspace, the RL algorithm will assure that the arm does not get stretched too beyond the region of interest. The singularity analysis of the employed robotic arm (UR5) should be taken into consideration. As the locus of the end-effector remains parallel to the plane where the particle is manipulated, Figure 3 shows that the arm will not encounter singularities.

In the presented scenario, switching between different sub-controllers essentially implies altering desired joints angles—calculated from desired values in Cartesian space—and a closed control architecture within industrial robots or cobots (collaborative robots) such as UR5 has the responsibility of regulating joint angles to desired values (sometimes for simplicity each joint of an industrial arm might be seen as a simple integrator, which is being controlled internally). It should be noted that due to the nature of tasks considered in this work which are set-points regulations—kinematics-based when Sub-controller 1 is active, or based on simplified dynamics (single integrators) when Sub-controller 2 is active—, disturbances do not occur too fast and/or with large accelerations.

2) *Sub-controller 2*: Once the robotic arm locates the magnetic particle, Sub-controller 2 steers the robotic arm to carry the particle to a desired location in a constrained (intestine-like) environment. Importantly, feedback of the optical tracking system is not used in this controller (\mathbf{P}_{ee} is in the vicinity if \mathbf{P}_p), and the robots only use the end-effector RGB camera to perceive the environment partially. When Sub-controller 2 is activated, the control-loop mechanism—depicted in Figure 4—is as follows: The RL algorithm takes the particle toward a user-determined target position within the environment. This sub-controller is active as long as the particle is responsive to the external magnetic field of the

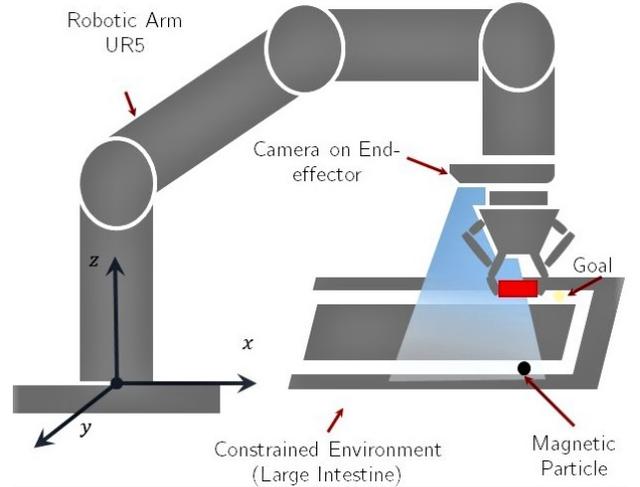


Fig. 4. The control scheme that trains a customized Rainbow algorithm to navigate through a constrained workspace when Sub-controller 2 is activated: Sub-controller 2 steers the robotic arm to carry the particle to a desired location in the environment.

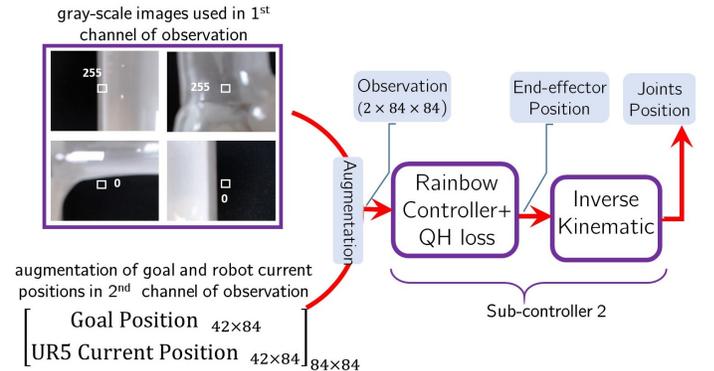


Fig. 5. Observations to train Sub-controller 2: Two-channel images 84×84 , i.e., $2 \times 84 \times 84$. The first channel includes a gray-scale image obtained from the end-effector RGB camera, and the second channel is an augmentation of two matrices with a size of 42×84 which are populated with the goal and robot current positions. The average of the pixel values in the center of each gray-scale image is used for obstacle detection.

magnet attached to the end-effector, in other words, the particle is within a specific threshold from the end-effector.

For Sub-controller 2, inspired by [49], a customized Rainbow algorithm with Quantile Huber loss from Implicit Quantile Networks (IQN) algorithm [53] and ResNet [54] are employed. The Huber loss is traditionally used to train the family of DQN RL algorithms since the introduction of DQN by Mnih et al. in [55]. A comparative study by Obando-Ceron et al. in [56] showed the superior performance of the Huber loss over the more straightforward mean-squared error loss.

Observations which are fed to Sub-controller 2 are two-channels images with dimension 84×84 , i.e., $2 \times 84 \times 84$. It is worth mentioning that images are downsampled from 480×480 to 84×84 . The first channel includes a gray-scale image obtained from the end-effector RGB camera, and the second channel is built by augmentation of two matrices with the size 42×84 , which are populated with the goal and current positions of the robot, respectively (Figure 5). It should be

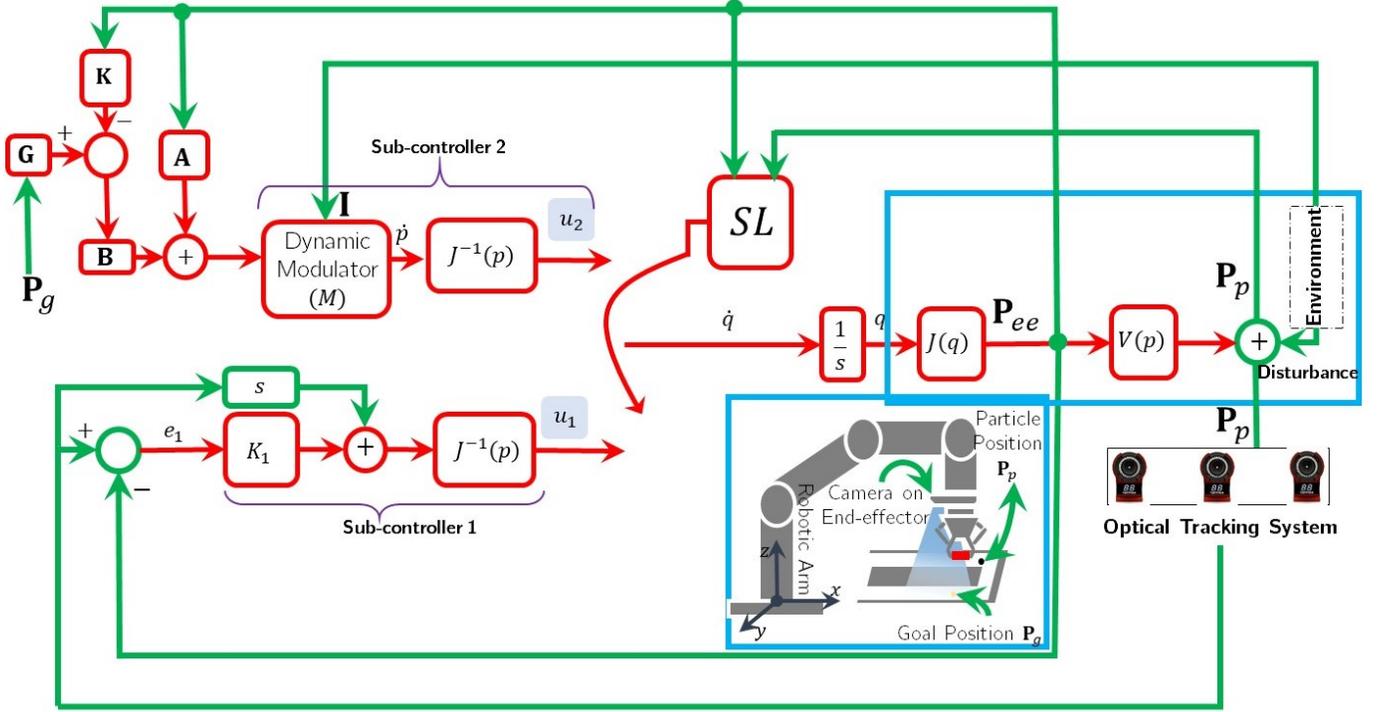


Fig. 6. The switching controller architecture consists of Sub-controllers 1 and 2 and a switching law (SL). SL is a supervisory algorithm that controls the switching between the various controllers. The observations or feedback signals are: the magnetic particle position obtained through the optical tracking system (\mathbf{P}_p), partial information (\mathbf{I}) of the constrained environment captured by an RGB camera on the end-effector, i.e., the RGB camera only covers a portion of the environment for the Sub-controller 2. \mathbf{K} is the gain of the state feedback controller, $\mathbf{B}_{2 \times 2}$, and $\mathbf{A}_{2 \times 2}$ are the identity matrices ($Id_{2 \times 2}$). Furthermore, $\mathbf{G}_{2 \times 2} = [\mathbf{K} \quad Id] \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ Id & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ Id \end{bmatrix}$. The symbol s represents a derivative filter. In Sub-controller 1, K_1 represents a proportional gain for adjusting the asymptotic convergence of the controller. Joints are modeled as a single integrator which is represented by $\frac{1}{s}$. Inverse kinematic and kinematic of the robot is denoted by $J^{-1}(p)$ and $J(q)$, respectively. q and \dot{q} represent joint positions and angular velocities. $V(p)$ models the unknown magnetic interaction dynamic between the external magnet on the end-effector and the particle with the input \mathbf{P}_{ee} and output \mathbf{P}_p .

noted that a simple segmentation process is done on each RGB camera to detect obstacles. A square of pixels of size 15-by-15 in the center of each image is considered, and then the mean of the pixel values in that square is calculated and used to detect obstacles. An obstacle is detected when the pixels' average is < 150 . Four discrete actions are considered for controlling the robotic arm end-effector in the xy -plane in two dimensions by setting a new position for the end-effector with a small change with respect to the end-effector current position. Figure 5 shows details of the training.

Furthermore, as a reward function employed in the training of Sub-controller 2, an episode is terminated with a negative reward if there is no answer for the inverse kinematic or number of steps in each episode exceeds 150 or it is detected that the end-effector would force the particle inside an obstacle. In addition, an episode is terminated with a positive reward if a goal is detected by the end-effector-camera within 2 cm in the xy -plane. In precise terms, if the Norm-2 of the error between a current position of the robotic arm and a goal position is ≤ 2 cm, that target is reached. It should be added that during training, goals are spawns randomly at different locations in each episode.

3) *Switching Signal*: In the switching control architecture, the updating law for activating the sub-controllers is as follows:

$$SL = \frac{1}{2} \left(3 + \text{sgn}(T - e) \right) \quad (1)$$

where SL is the index of to-be-active Sub-controller, $e = \|\mathbf{P}_{ee} - \mathbf{P}_p\|$ and \mathbf{P}_{ee} is the current position of the robotic arm, \mathbf{P}_p is the position of the particle, and T is an arbitrary threshold set to 10 cm for the experiment. Although T is an arbitrary threshold, it is dependent on the step size of the manipulator, transporting the particle and also the particle's size. In other words, the reduced threshold is allowed by using the smaller step or particle sizes.

It should be noted that $\text{sgn}(\arg) = 1$, if $\arg \geq 0$, otherwise $\text{sgn}(\arg) = -1$. In other words, Sub-controller 2 is activated unless the euclidean distance between end-effector and particle positions is more than or equal T .

B. Attractor Dynamics-based Approach

In the Attractor Dynamics-based method, a state feedback controller together with a dynamic modulator [40, Section 3.2] are designed to ensure that the robot will not hit convex obstacles while carrying the particle toward a goal. The block diagram of the algorithm is shown in Figure 6. Sub-controller 1 and Switching law are the same as what is discussed in Section IV-A1. For Sub-controller 2, each joint of the manipulator is modeled as a continuous single integrator:

$$\begin{aligned}\dot{p}(t) &= \mathbf{A}p(t) + \mathbf{B}u(t), \\ o(t) &= p(t)\end{aligned}\quad (2)$$

where $p(t) = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^{2 \times 1}$ is Cartesian position of the end-effector in a 2-D space (xy -plane), $\mathbf{B}_{2 \times 2}$, state matrix and $\mathbf{A}_{2 \times 2}$, input matrix, are the identity matrices ($Id_{2 \times 2}$). Furthermore, $u(t) \in \mathbb{R}^{2 \times 1}$ is a 2-by-1 control input. Let \mathbf{P}_g denotes the desired constant 2-by-1 vector for the output $o(t)$ to track asymptotically.

The control goal of Sub-controller 2 is to design a state feedback controller in which $u(t)$ depends on $p(t)$ and \mathbf{P}_g so that the regulation error $e(t) = \mathbf{P}_g - o(t)$ goes to zero when $t \rightarrow \infty$. It can be shown ([57], [58]) that the state feedback control law:

$$u(t) = \mathbf{G}\mathbf{P}_g - \mathbf{K}p(t) \quad (3)$$

makes System 2 a globally asymptotically stable system,

$$\text{where } \mathbf{G}_{2 \times 2} = [K \quad Id]_{2 \times 4} \begin{bmatrix} \mathbf{A} & \mathbf{B} \\ Id & \mathbf{0} \end{bmatrix}_{4 \times 4}^{-1} \begin{bmatrix} \mathbf{0} \\ Id \end{bmatrix}_{4 \times 2}.$$

Although the control law 3 makes the manipulator stable, it cannot prevent the robot from colliding with obstacles. Therefore, a real-time obstacle avoidance strategy should be considered together with the law 3. To design the obstacle avoidance protocol, we follow the same line of ideas as in [40]. In this work, we consider rectangular 2-D obstacles based on superellipse curves, where $\Gamma(p) : \left\| \frac{x-x_o}{a} \right\|^n + \left\| \frac{y-y_o}{b} \right\|^n = 1$ represents the boundary points of an obstacle with the center point $[x_o, y_o]$, furthermore a and b are called the semi-diameters of the curve. In other words, the curve $\Gamma(p)$ is confined in the rectangle $\|x-x_o\| \leq a$ and $\|y-y_o\| \leq b$. The similarity of the curve to a rectangle is adjustable with the parameter n , and with $n > 2$, the curve looks like a rectangle with rounded corners.

The dynamic modulator matrix M (designed in [40, Section 3.2]) propagates the influence of the obstacle on the motion flow with the maximum effect at the boundaries of the obstacle, and vanishes for points far from it. Applying the dynamic modulator, on the linear system 2 yields:

$$\dot{p}(t) = M \left(\mathbf{A}p(t) + \mathbf{B}u(t) \right) \quad (4)$$

where $u(t)$ is given in Equation 3. Details of design, stability and convergence analysis of the system with a modulation matrix can be found in [40]. Here we skip the analysis for brevity.

For the computer implementation of 4, a simple way is to replace the derivative by a difference

$$p(t_{k+1}) = M \left(\mathbf{A}p(t_k) + \mathbf{B}\mathbf{G}\mathbf{P}_g - \mathbf{K}p(t_k) \right) h + p(t_k)h$$

where t_k is the sampling instant and $h = t_{k+1} - t_k$ is the sampling period.

C. ERRT-based Approach

To better contextualize the proposed RL-based control algorithm, this work additionally proposes employing an ERRT-based algorithm, the specifics of which is detailed in [41] as a path planner within the Sub-controller 2.

ERRT is a sampling-based planning method for a continuous domain. Its two extensions compared to RRT method [59]—the waypoint cache and adaptive cost penalty search—enable the method to explore the environment and plan an obstacle-free path on the fly. However, the quality of the generated path and efficient replanning heavily depends on the processing of the received obstacles images from the environment. Therefore, to be able to use this algorithm to solve the presented problem in this paper, we develop a technique to optimally allocate obstacles in rectangular shapes to be used in the ERRT algorithm. In addition, the applicability of algorithm in real-time heavily depends on how much the environment is polluted by obstacles, the complexity of obstacle shapes, and also whether obstacles are static or dynamic and how fast obstacles are changing.

The block diagram of the switching controller based on this approach is presented in Figure 7. Sub-controller 1 and Switching law are the same as what was mentioned in Section IV-A1. However, in Sub-controller 2, ERRT, instead of the RL-based algorithm, is used as a path planner in Cartesian space and the output is fed to the inverse kinematic function to generate joint angles.

In the obstacle processor, a quadtree [60] as a tree data structure is used for spatial 2-D searching to optimally partition obstacles into minimum number of rectangular obstacles within a grayscale image respecting a user-defined threshold for height and width of acceptable smallest rectangle, as depicted in Figure 8. As it is shown, there exist regions with different shades of gray and the contrast ranges from black—pixel value 0—at the weakest intensity to white—pixel value 255—at the strongest. To reduce the cont of computation in the ERRT algorithm, only regions with color intensity of 50 or less. These rectangular obstacles can be used efficiently by the ERRT method.

D. A Discussion on Stability

For the stability of this Sub-controller 1 in RL, Attractor dynamics, and ERRT-based methods, it can be shown that

$$\begin{aligned}e_1 &= \mathbf{P}_p - \mathbf{P}_{ee}, \\ \dot{e}_1 &= \dot{\mathbf{P}}_p - \dot{\mathbf{P}}_{ee}, \\ \dot{e}_1 &= \dot{\mathbf{P}}_p - J(q)\dot{q}, \\ \dot{e}_1 &= \dot{\mathbf{P}}_p - J(q)J^{-1}(q) \left(\dot{\mathbf{P}}_p + K_1(\mathbf{P}_p - \mathbf{P}_{ee}) \right), \\ \dot{e}_1 &= -K_1 e_1\end{aligned}\quad (5)$$

where K_1 is a 3-by-3 diagonal positive definite matrix, then it is guaranteed that Cartesian vector-valued e_1 converges to 0 asymptotically for any initial values.

As it is discussed in Section IV-A1, a closed control architecture within industrial robots or cobots (collaborative robots) such as UR5 has the responsibility of regulating joint angles to desired values (sometimes for simplicity each joint of an industrial arm might be seen as a simple integrator, which is being controlled internally). For the stability of this Sub-

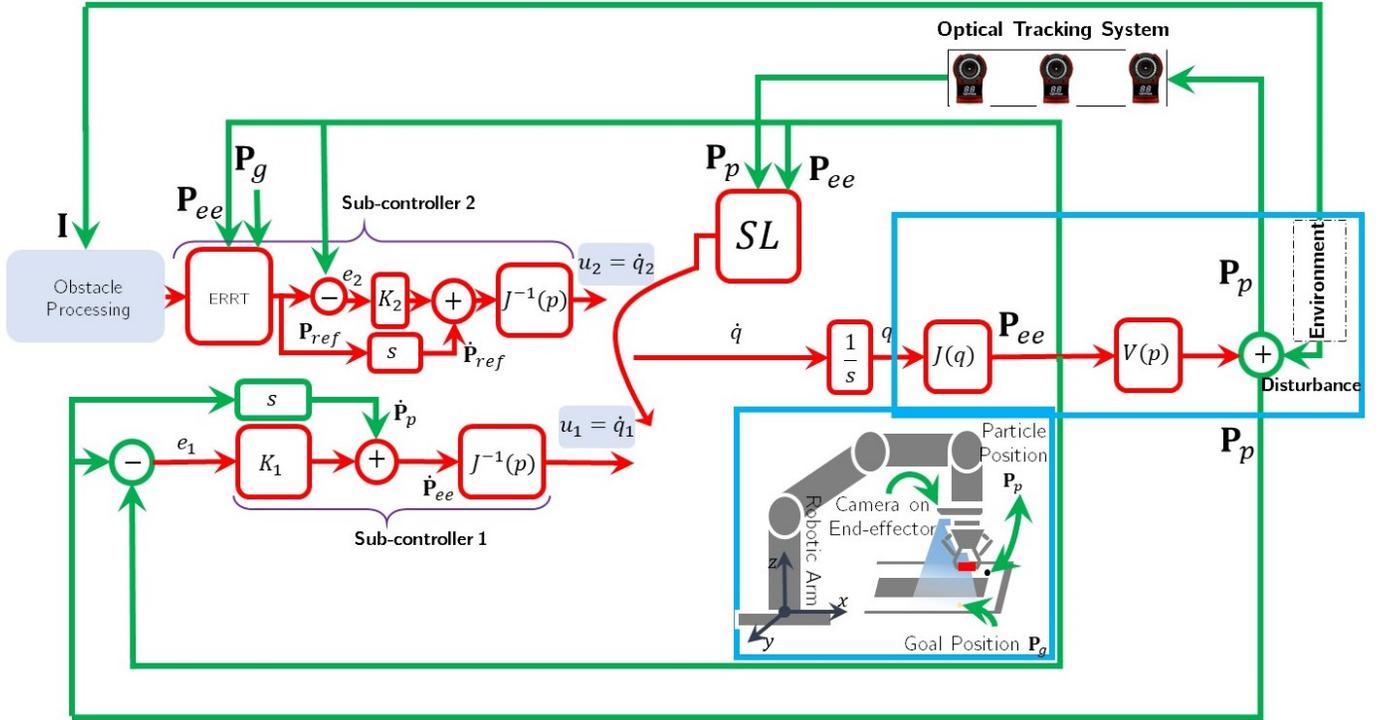


Fig. 7. The switching controller architecture consists of Sub-controllers 1 and 2 and a switching law (SL). SL is a supervisory algorithm that controls the switching between the various controllers. The observations or feedback signals are: the magnetic particle position obtained through the optical tracking system (\mathbf{P}_p), partial information (\mathbf{I}) of the constrained environment captured by an RGB camera on the end-effector, i.e., The RGB camera only covers a portion of the environment for the Sub-controller 2. \mathbf{I} is fed to the obstacle processing unit which optimally partitions obstacles into minimum number of rectangular obstacles. The ERRT unit generates an obstacle-free path based on the given feedback information. The symbol s represents a derivative filter. In Sub-controller 1, K_1 represents a proportional gain for adjusting the asymptotic convergence of the controller, and in Sub-controller 2, K_2 represents a positive definite 3-by-3 matrix for adjusting the asymptotic convergence of the controller. K_2 is a simple model for the internal control of the robot. Joints are modeled as a single integrator which is represented by $\frac{1}{s}$. Inverse kinematic and kinematic of the robot is denoted by $J^{-1}(p)$ and $J(q)$, respectively. q and \dot{q} represent joint positions and angular velocities. $V(p)$ models the unknown magnetic interaction dynamic between the external magnet on the end-effector and the particle with the input \mathbf{P}_{ee} and output \mathbf{P}_p .

controller 2 in RL and ERRT-based methods, it can be shown that

$$\begin{aligned}
 e_2 &= \mathbf{P}_{ref} - \mathbf{P}_{ee}, \\
 \dot{e}_2 &= \dot{\mathbf{P}}_{ref} - \dot{\mathbf{P}}_{ee}, \\
 \dot{e}_2 &= \dot{\mathbf{P}}_{ref} - J(q)\dot{q}, \\
 \dot{e}_2 &= \dot{\mathbf{P}}_{ref} - J(q)J^{-1}(q)\left(\dot{\mathbf{P}}_p + K_1(\mathbf{P}_{ref} - \mathbf{P}_{ee})\right), \\
 \dot{e}_2 &= -K_2 e_2
 \end{aligned} \tag{6}$$

where K_2 is a 3-by-3 diagonal positive definite matrix, then it is guaranteed that Cartesian vector-valued e_2 converges to 0 asymptotically for any initial values.

Both RL and ERRT algorithms act as path planners and it can be seen that the robot's internal controller stabilizes the system, as long as an obstacle-free path within the workspace of the robot is generated. This is determined by how Sub-controller 2 is well trained in the RL algorithm or how good the quality of a path is in the ERRT method. Furthermore, in the attractor dynamic-based method, the global asymptotic stability of Sub-controller 2 is guaranteed as it is discussed earlier in Section IV-B.

It should be noted that each control loop has different set-points and the loops are entirely separated in that sense. In other words, the switching between those two stable control

loops does not introduce insatiability to the system and the loops accomplish separated sub-tasks with different set-points. Therefore, this question may arise that whether or not there is a possibility of having only one controller to accomplish the task. Since there is no need for a collision avoidance strategy when the manipulator is moving to the location of a disturbed particle, two different controllers are considered.

V. SIMULATIONS, EXPERIMENTS, AND RESULTS

For simulations and real-world experiments, we use the UR5 robotic arm from Universal Robots. The UR5 robot is a collaborative robot with six joints. Built-in safety mechanisms remove the need for safety guards between humans and the robot, and this opens possibilities for using the robot in medical applications where the robot is in direct contact with a patient [61]. It has a low-level controller called URControl, and it can be programmed by communicating over a TCP/IP connection. To control the robot at the script level, a programming language called URScript can be used. URScript programs can be sent from a computer to the robot controller (URControl) as strings over the socket once the connection with the robot is established. The URControl runs programs in real-time and sends information at a maximum frequency of 125Hz. The output of the URControl contains all the

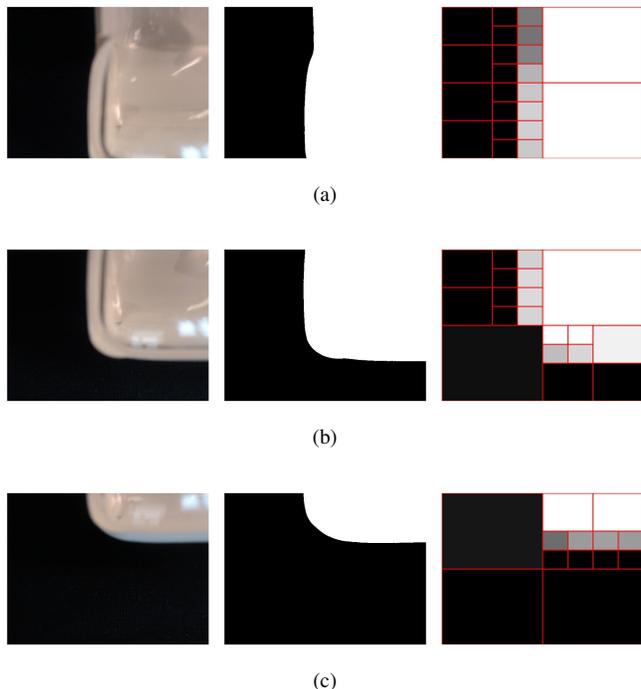


Fig. 8. Examples of inputs and outputs for the obstacle processing unit. (a)-(c): the most left images are the inputs to the unit. The input images are converted to grayscale images first, then a quadtree algorithm is used to optimally partition obstacles within a grayscale images into minimum number of rectangular obstacles respecting an user-defined threshold for height and width of acceptable smallest rectangle. In these figures, the threshold is 50 for both height and width.

information the robot needs to perform a movement, including angular positions, velocities, target accelerations, and currents for all joints. In this work, the *moveJ* command, which offers a position control interface in tool-space, is used to control the UR5 [62]. It should be noted that achieving high-frequency real-time performance for a control loop is limited by the bandwidths of employed sensors and also the computational time required to calculate the control input.

The UR5 manipulator is equipped with the Robotiq 2F-85 adaptive gripper, holding a neodymium block magnet in a fixed position. The experiment setup is shown in Figure 9, where images from the end-effector camera are used to detect the magnet and identify obstacles in the environment.

We use a sphere NdFeB magnet with grade N42 and diameter 5 mm as a magnetic particle. Besides, a neodymium block magnet with dimensions $50 \times 25 \times 10$ mm with grade N35 is attached to the UR5’s end-effector to produce enough magnetic field strength to carry the magnetic particle. The magnets are shown in Figure 10. To avoid a collision between two magnets, the UR5 moves in an imaginary plane with a 10 cm distance from the surface of the table.

A. Simulations, Experiments, and Results: Reinforcement Learning-based Approach

Training a policy in the real world can be costly. Simulations can speed up the learning process and help avoid potentially unwanted actions that can damage the robot or the surrounding environment. OpenAI Gym and PyBullet [63] are used for this

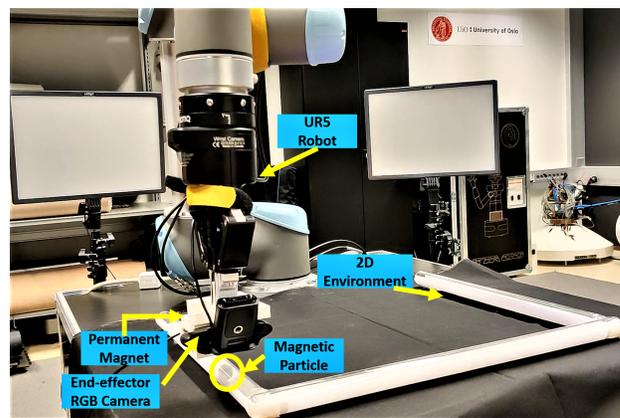


Fig. 9. The experimental setup: a UR5, a permanent magnet, and an RGB camera mounted on the end-effector; the constrained 2D workspace; and the object (a sphere magnet).

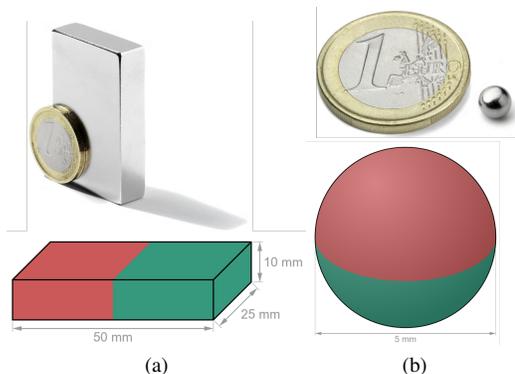


Fig. 10. (a): A neodymium block magnet with dimensions of $50 \times 25 \times 10$ mm and grade N35. (b): A sphere NdFeB magnet with grade N42 and diameter 5 mm as a magnetic particle.

purpose. However, modeling a complex environment or robots can be challenging, and this may introduce a simulation-reality gap [64]. To minimize this gap, the UR5 swept and imaged the whole real-world environment using the mounted RGB camera (with step = 1 cm in both x and y -axes). As discussed, the UR5 end-effector was kept in an imaginary plane with a 10 cm distance from the surface of the table during the imaging process. Samples of these images are shown in Figure 5. The images are fed to the algorithm in the training phase.

An overview of the environment used for training Sub-controller 2 is shown in Figure 11. The customized Rainbow algorithm, as previously explained, is employed, where we add images from the end-effector RGB camera together with goal and end-effector positions to the observation space. Image-based observations are used to avoid collisions with obstacles in the environment. Figure 5 shows details of the training.

As mentioned, OpenAI Gym and PyBullet are employed to train the customized Rainbow algorithm for Sub-controller 2, and Figure 12 represents the learning performance of the employed RL algorithm over the training episodes. It is worth mentioning that in the simulation, the magnetic particle is not considered, and therefore the whole switching control architecture is not implemented. After training, the average

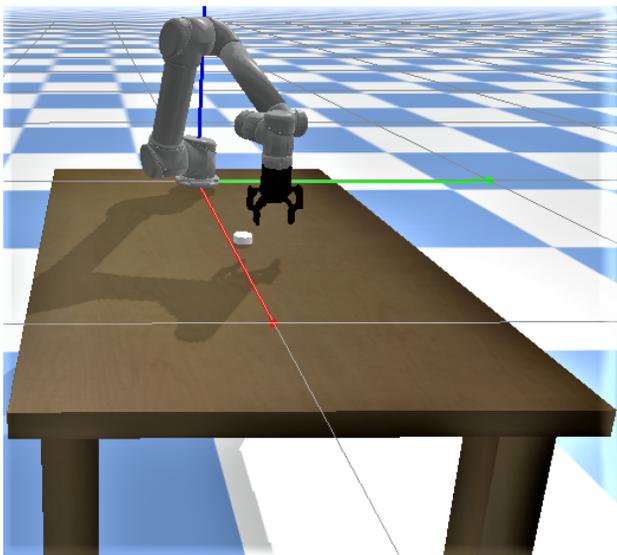


Fig. 11. The simulation environment used to train the Rainbow controller: Images taken from the real-world environment are used as observations. The robot gets a negative reward for hitting obstacles and not finding an inverse kinematic solution. It gets positive rewards for reaching the target.

success rate for the learned policy is 100% for randomly generated initial and goal positions (calculated over 30 random restarts).

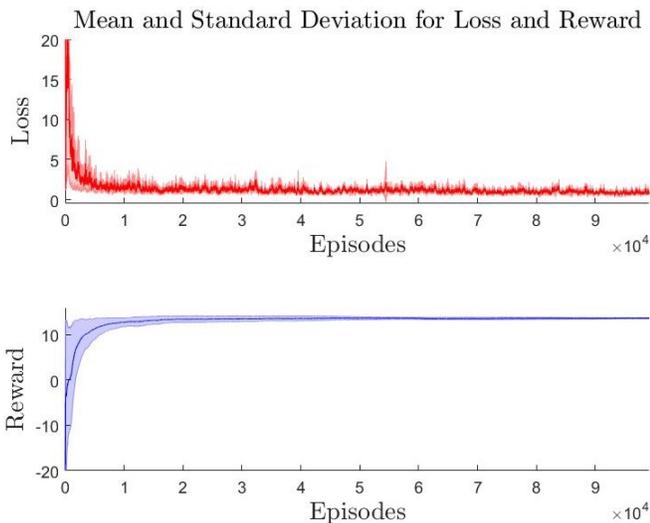


Fig. 12. The mean and standard deviation of loss and rewards over each training episode for 5 rounds of training. At the beginning of training, the system's performance is poor, as there is a high standard deviation for the losses and rewards. Eventually, the average and standard deviation of the rewards and losses will improve, meaning that the system will become more reliable and, statistically speaking, will accomplish the task.

Afterward, we consider the magnetic particle in the loop, and the full controller structure is implemented. In the real world, the magnetic particle is consistently carried to a randomly placed target starting from a random position. We repeated undisturbed scenarios for arbitrary initial and goal positions for 30 episodes, and the average success rate is 100% with the mean tracking error 0.02 (m) in both axes, the average maximum error 0.08 (m) and 0.09 (m) in x ,

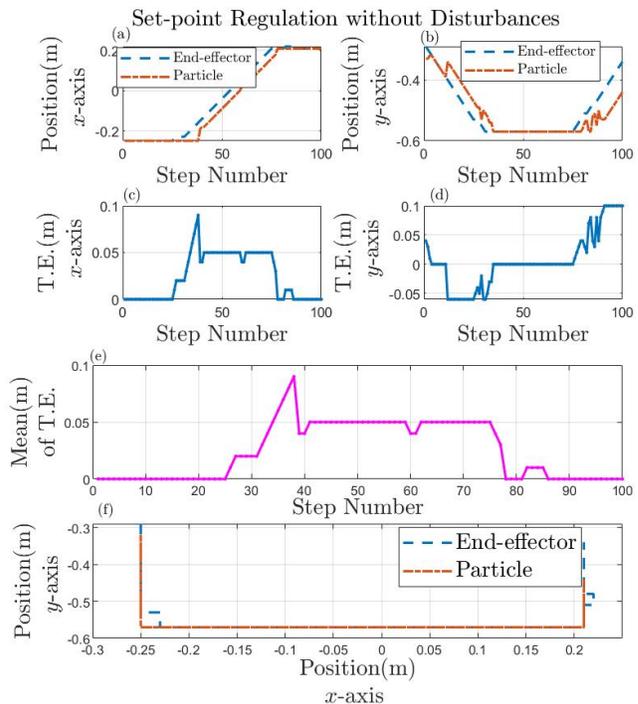


Fig. 13. A scenario of set-point regulation without disturbances in one episode. (a,b): Cartesian positions of the end-effector and the magnetic particle are depicted in both x and y -axes. (c,d): Tracking Errors (T.E.) in each axis are shown. Errors in each axis increase when there is motion in that axis. (e): The Euclidean norm of errors in both axes at each step number is shown. (f): The locus of the particle and end-effector in the 2-D workspace is depicted.

and y -axes, respectively. Figure 13 represents a set-point regulation scenario without any disturbances. In this specific scenario, the particle is supposed to reach the goal position $[0.21, -0.40, 0.10]$ within the constrained environment.

Finally, the robustness of the system to deal with external disturbances which could result from real-world uncertainties is studied. As disturbances, we physically draw the magnetic particle away from the magnetic control zone during active control. As a demonstration case, Figure 14 depicts the particle and end-effector trajectories through the environment in the presence of disturbances where the particle is supposed to reach the goal position $[0.21, -0.40, 0.10]$. We repeated the disturbed scenarios for random disturbances, initial, and target positions for 30 episodes, and the average success rate was 96.6%.

In Figure 12, we run 5 rounds of training, and each training consists of multiple episodes (more than 90 thousand). In each episode, initial positions and goals are set randomly. In the initial episodes (for all 5 rounds of training), the robot most likely fails to accomplish the task, which is why the average of the rewards for the initial episodes considering 5 rounds of training is a negative number and there is a relatively high loss as well. In addition, it can be seen that at the beginning of training, the system's performance is poor as there is a high standard deviation for the losses and rewards considering the 5 rounds of training. After some more episodes, the average and standard deviation of the rewards and losses, considering

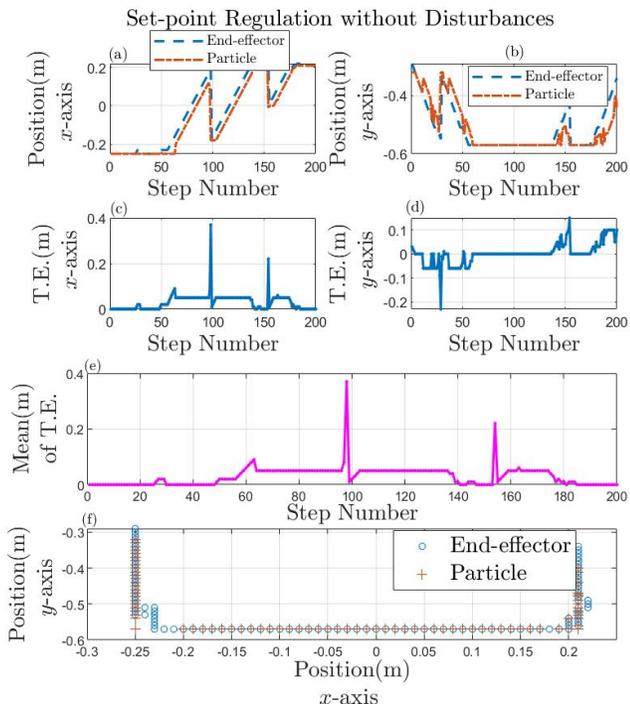


Fig. 14. A scenario of set-point regulation with three disturbances in one episode. Disturbances happen at step numbers 29, 98, and 154 (a,b): Cartesian positions of the end-effector and the magnetic particle are depicted in both x and y -axes. In (a), disturbances at step numbers 98 and 154, and in (b), a disturbance at step number 98 is visible. (c,d): Tracking Errors (T.E.) in each axis are shown. Errors in each axis increase when there is motion in that axis or when a disturbance happens. The magnitude of disturbances at step numbers 98 and 154 and step number 29 can be seen in (c) and (d), respectively. (e): The Euclidean norm of errors in both axes at each step number is shown. (f): The locus of the particle and end-effector in the 2-D workspace is depicted.

all rounds of training, get better, and this means that the system gets more reliable and, statistically speaking, it will accomplish the task. After 90 thousand episodes, the time to reach a random target position from a random initial position will differ. Figures 13 and 14 only show two episodes after the controller is trained (after 90 thousand episodes).

B. Simulations and Results: Attractor Dynamics-based Approach

The following example illustrates the Attractor Dynamic-based method without any disturbance applied.

Example 1. Consider a 2-D shape of the constrained environment with the manipulator's initial Cartesian position $[-0.78, 0.73, 0.10]$ and target position $[0.77, 0.19, 0.10]$. An overview of the simulation is shown in Figure 15. It should be noted that the manipulator can sense an obstacle partially with a maximum height and width of 2 cm. The system employs the state feedback control law given in (3). Since there is no perturbation is assumed, Sub-controller 1 (as shown in Figure 6) is not active.

An advantage of the considered conventional method is that there is no need for offline training or parameter tuning; however, the method cannot handle concave or connected

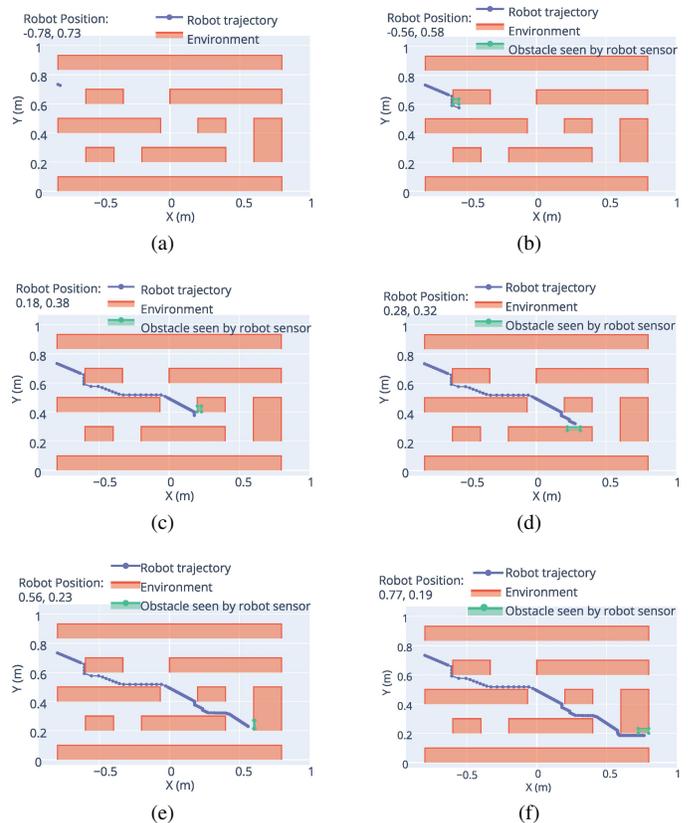


Fig. 15. Online obstacle avoidance of UR5 carrying an undisturbed particle through a 2-D environment as discussed in Example 1. Information from a part of the obstacle seen by the robot is used to make the particle go around the obstacle. Since there is no perturbation, Sub-controller 2 carries the particle toward the target position.

obstacles. Since the method requires analytical modeling of the obstacles boundary, we consider a rectangle around the part of the obstacle detected by the robot's sensor — we fit the point cloud with a rectangle—. However, the actual shape of obstacle(s) may not be necessarily a rectangle. In other words, the approach might be conservative in some scenarios.

In the following examples, a disturbance is introduced to the motion of the particle.

Example 2. Suppose the same 2-D space as Example 1 is employed. The manipulator's initial Cartesian position $[0.72, 0.76, 0.10]$ and target position $[-0.77, 0.12, 0.10]$. An overview of the simulation is shown in Figure 16. Furthermore, a perturbation in x -axis with magnitude 0.63 m is applied to the particle when the particle is at the position $[-0.21, 0.55, 0]$ or at step number 50 where Sub-controller 1 (as shown Figure 6) is active and takes the manipulator to the new position of particle. When the manipulators in the place, Sub-controller 2 has the responsibility of carrying the particle toward the target position. Also in this example, the manipulator can partially sense an obstacle with a maximum height and width of 2 cm. As shown in Figure 16, the particle passes the obstacles based on received information from the obstacles (part of obstacle seen by the robot), its direction, and relative position to the obstacles.

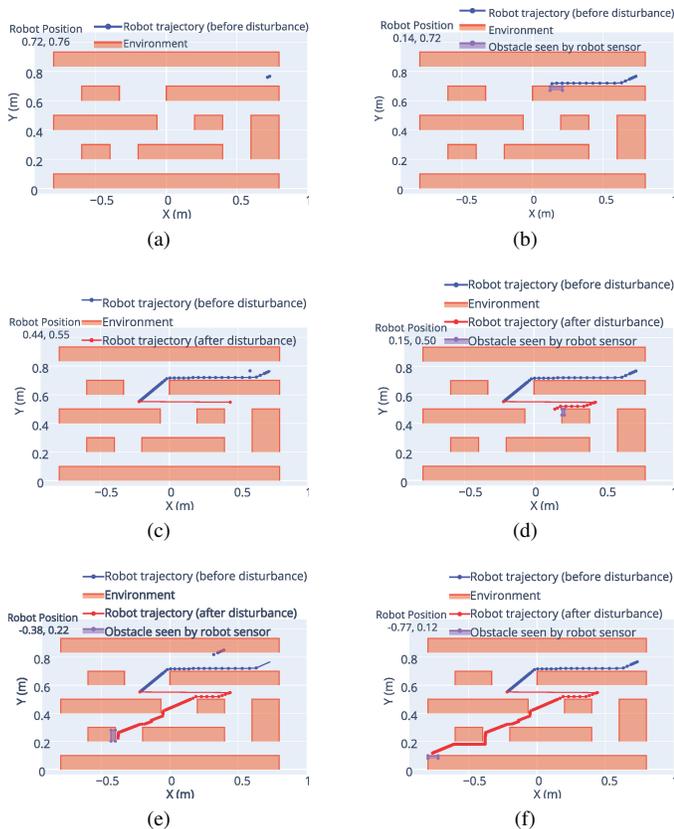


Fig. 16. Online obstacle avoidance of UR5 carrying a disturbed particle through a 2-D environment as discussed in Example 2. A disturbance in x -axis with magnitude 0.63 m is applied to the particle when the particle is at the position $[-0.21, 0.55, 0]$, as shown in Figure 16d. When the disturbance occurs, Sub-controller 1 directs the manipulator to the new position of the particle. When the manipulator reaches the position, Sub-controller 2 has the responsibility of carrying the particle toward the target position (the path shown in red). Information from a part of the obstacle seen by the robot is used to make the particle go around the obstacle.

C. Simulations and Results: ERRT-based Approach

The following example illustrates the performance of ERRT-based method with disturbances applied to the particle.

Example 3. Consider a 2-D environment as depicted in Figure 17 with the manipulator's initial Cartesian position $[-0.24, -0.27, 0.10]$ and target position $[0.19, -0.28, 0.10]$.

In the next scenario, two disturbances happen at locations $[0.21, -0.34, 0.10]$ and $[0.20, -0.47, 0.10]$ where the particle is displaced to new positions $[-0.19, -0.48, 0.10]$ and $[-0.24, -0.28, 0.10]$, respectively as shown in Figure 18 where initial, target, and disturbance positions, explored area and selected path are depicted. In addition, it should be noted that the ERRT algorithm senses obstacles through received images from the environment depending on the robot's end-effector current position (obstacles are partially seen by the camera mounted on robot).

Figure 19 depicts the particle and end-effector trajectories together with Tracking Error (T.E.) in both axes x and y and mean of T.E., through the environment in the presence of disturbances where the particle is supposed to reach the goal position $[0.21, -0.40, 0.10]$.

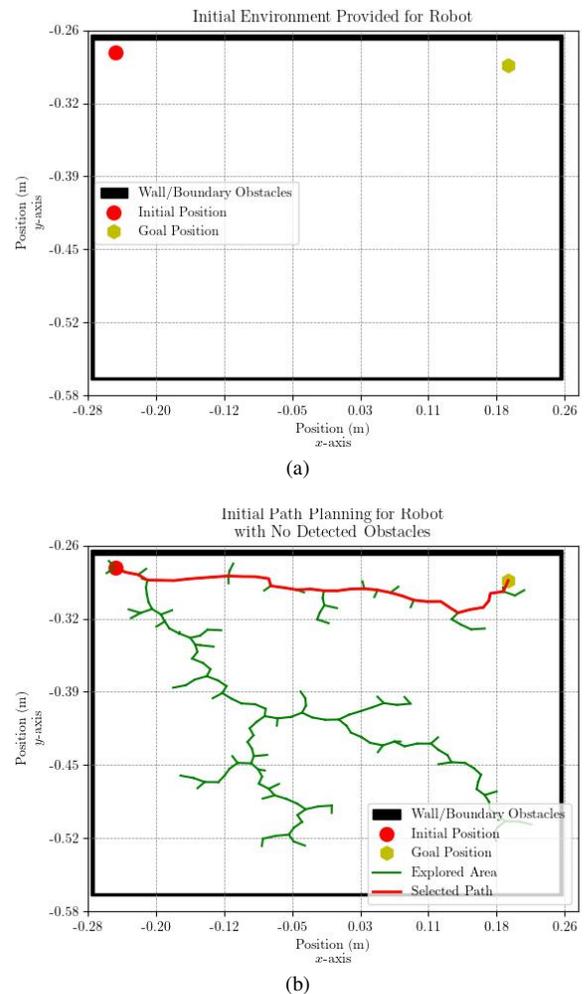


Fig. 17. (a): The initial environment provided for the robot with boundary obstacles to limit the search space. (b): The initial planned path by ERRT-based path planner from the initial position to the goal position together with the explored and selected paths and boundary obstacles to limit the search space. No obstacle is detected by the camera.

VI. DISCUSSIONS

In this paper, we examine the feasibility of implementing the Rainbow algorithm [49] customized with Quantile Huber loss from Implicit Quantile Networks (IQN) algorithm [53] and ResNet [54] to accomplish set-point regulation tasks within a constrained environment in the presence of disturbances. Our results suggest that deep RL methods as a trajectory planner prevent the robot from encountering singularities and guide the particle toward a goal while avoiding obstacles. Furthermore, the results show that it is possible to train the RL algorithm fully within a simulation environment and deploy it as-is in a real-world scenario for remote magnetic control with reliable behavior in the presence of disturbances and uncertainties. Although traditional control approaches may ensure some performance levels under restricted assumptions, their design is challenging, especially when only partial information from the environment is available at each time step. Also, in many traditional path planning and control design approaches, an exact analytical model of the robot and its interaction with the surrounding environment is a necessity; contrastingly, RL

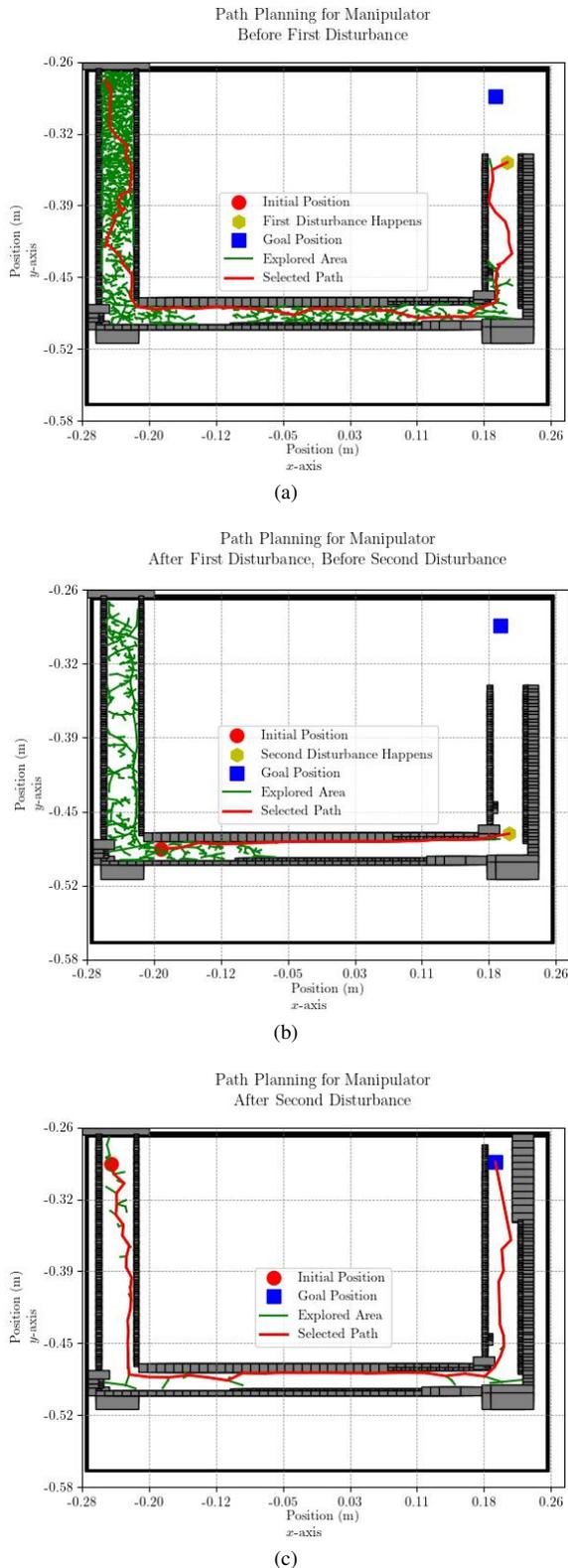


Fig. 18. The initial, target, and disturbance locations, explored area and selected path. (a): The initial position before the first disturbance is at $[-0.24, -0.27, 0.10]$, target position is at $[0.19, -0.28, 0.1]$, and the first disturbance occurs at $[0.21, -0.34, 0.1]$. (b): The initial position after the first disturbance is at $[-0.19, -0.48, 0.1]$, the target position is the same as before, and the second disturbance occurs at $[0.20, -0.47, 0.1]$. (c): The initial position after the second disturbance is at $[-0.24, -0.28, 0.1]$, the target position is kept the same as before, and no disturbance occurs.

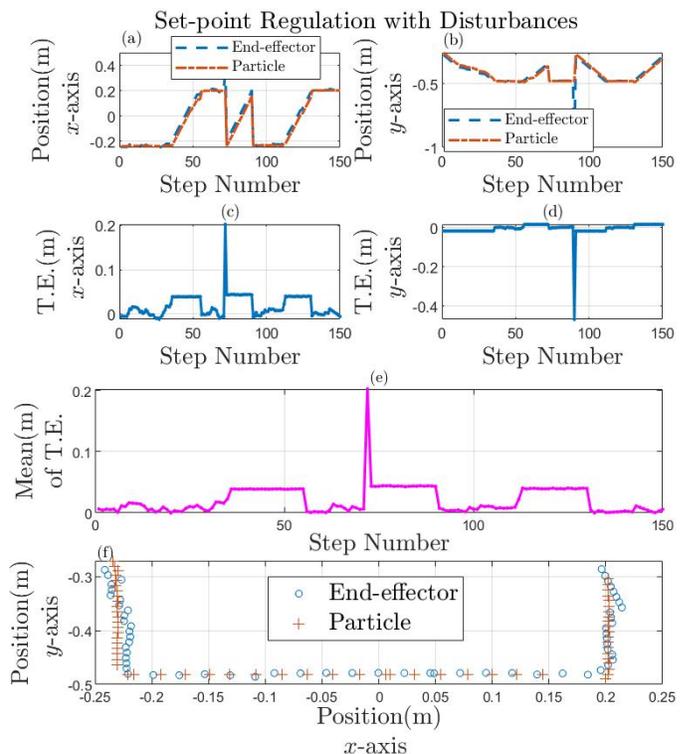


Fig. 19. A scenario of set-point regulation with three disturbances in one episode. Disturbances happen at step numbers 72 and 90 (a,b): Cartesian positions of the end-effector and the magnetic particle are depicted in both x and y -axes. In (a), a disturbance at step numbers 72, and in (b), a disturbance at step number 90 are visible. (c,d): Tracking Errors (T.E.) in each axis are shown. Errors in each axis increase when there is motion in that axis or when a disturbance happens. The magnitude of disturbances at step numbers 72 and step number 90 can be seen in (c) and (d), respectively. (e): The Euclidean norm of errors in both axes at each step number is shown. (f): The locus of the particle and end-effector in the 2-D workspace is depicted.

methods do not rely on pre-existing knowledge or models.

The drug delivery example is a futuristic, potential application in which the human intervascular system is a 3-D, time-varying, and dynamic environment. However, compared to traditional methods, the strength of RL-based methods is that they are applicable to the more complex environment without a need to re-design the controller structure—which is based on neural networks—. Yet, there might be a need for more samples in the offline training and proper tweaking of the controller’s hyperparameters.

A simple image segmentation algorithm was employed in this work; however, equipping the RL method with a high-fidelity segmentation algorithm would help in better understanding the environment, thereby reducing the simulation-reality gap and augmenting the generalization of the learning algorithm to be used in a different environment. We experienced that learning performance is also sensitive to the choice and format of observation space; therefore, learning parameters must be chosen carefully.

It should be noted that as a first step to evaluate the feasibility of the proposed approach, we considered an optical tracking system to detect and guarantee the presence of a magnet in the immediate vicinity of the UR5 end-effector and an RGB camera to detect the constrained workspace. However,

for more practical uses, all the imaging modalities are better integrated at the end-effector and replaced by more relevant radiography methods, which are considered safe procedures, and can be used to see through the body.

It is worth mentioning that the presented Attractor Dynamics-based conventional method can only be used where obstacle(s) are convex and are not connected. The method requires analytical modeling of the obstacle's boundary. To this end, we consider a rectangle around the part of the obstacle detected by the robot's sensor—we fit the point cloud with a rectangle—. However, the actual shape of the obstacle(s) may not necessarily be a rectangle. In other words, the approach might be conservative in some scenarios. Also, integration errors can cause issues in the discrete implementation of the approach in a way that the next computed point falls inside an obstacle boundary due to the integration error, and trajectories tend to stay inside the obstacle (as no trajectory can enter or leave obstacles).

To highlight the benefits of the proposed RL-based algorithm, an ERRT-based approach is also employed to accomplish the same task. ERRT is a sampling-based planning method that explores the environment and plans an obstacle-free path on the fly.

In order to handle feeding the algorithm with varying obstacle shapes, a technique is developed to optimally partition obstacles into rectangular shapes that can be processed by the ERRT algorithm. This is the first difference between RL and ERRT-based methods. In the RL-based approach, there is no need to be concerned about the shape of obstacles as the whole obstacle image will be taken into account. Therefore, by approximating obstacles, the available workspace for generating an obstacle-free path might be reduced. At the same time, this approximation will also increase the computation loads to some extent, which eventually will lead to decreased bandwidth for the control loop.

In the proposed RL-based algorithm, training or exploring the environment takes place offline, yet in the ERRT method, environment exploration happens on the fly, and this directly affects whether the method can be used in real-time. If the environment is filled with complex and/or dynamic obstacles, the ERRT algorithm might not be responsive to the changes due to computational loads. However, this is not the case in the RL-based algorithm. In the presented scenario in this paper, the environment is static so it was feasible to employ ERRT method.

In the ERRT method, there is a need for calibration of the environment with respect to the robot to keep track of the obstacles' positions. Yet, in the RL-based approach, the current pose of the robot will be assigned to the received obstacle image, and this will be preserved in the neural network, which will enable employing the RL-based method in dynamic 3-D environments without being concerned about the calibration. The calibration constraint prevents ERRT from being implemented in 3-D environments.

VII. CONCLUSION

This article derives and validates a customized Rainbow RL method for online trajectory planning and remote control of

a ferromagnetic particle. Using magnetic actuation, the robot learned to robustly carry a small ferromagnetic object in a constrained environment. Furthermore, the trained network is integrated with the two-module controller, which is deployed as-is in a validation experiment in the real world, where the experiment showed the robustness of the approach against disturbances. Also, a conventional controller based on the Attractor Dynamics-based approach is designed. Afterward, simulations were carried out using more complex environments, and the shortcomings of the proposed method were also discussed. Finally, the experiment results from the ERRT-based method highlight the improved robustness to dynamic environments offered by the proposed RL-based algorithm in this work. For future work, a three-dimensional environment can be investigated. Also, in this work, the constrained environment is static, meaning it does not change over time. However, by training the RL algorithm in dynamic environments, there is a possibility of considering time-varying workspaces.

ACKNOWLEDGMENT

This work was partially supported by the Research Council of Norway (RCN) as a part of the Vulnerability in the Robot Society (VIROS) project under Grant Agreement No. 288285, Predictive and Intuitive Robot Companion (PIRC) under Grant Agreement No. 312333; and through its Centres of Excellence scheme, Project No. 262762.

REFERENCES

- [1] H. Lu, M. Zhang, Y. Yang, Q. Huang, T. Fukuda, Z. Wang, and Y. Shen, "A bioinspired multilegged soft millirobot that functions in both dry and wet conditions," *Nature communications*, vol. 9, no. 1, pp. 1–7, 2018.
- [2] G. Tiwari, R. Tiwari, B. Sriwastawa, L. Bhati, S. Pandey, P. Pandey, and S. K. Bannerjee, "Drug delivery systems: An updated review," *International journal of pharmaceutical investigation*, vol. 2, no. 1, p. 2, 2012.
- [3] H. Wen, H. Jung, and X. Li, "Drug delivery approaches in addressing clinical pharmacology-related issues: opportunities and challenges," *The AAPS journal*, vol. 17, no. 6, pp. 1327–1340, 2015.
- [4] M. Hu, X. Ge, X. Chen, W. Mao, X. Qian, and W.-E. Yuan, "Micro/nanorobot: A promising targeted drug delivery system," *Pharmaceutics*, vol. 12, no. 7, p. 665, 2020.
- [5] K. Bourzac, "Nanotechnology: carrying drugs," *Nature*, vol. 491, no. 7425, pp. S58–S60, 2012.
- [6] S. Schuerle, A. P. Soleimany, T. Yeh, G. Anand, M. Häberli, H. Fleming, N. Mirkhani, F. Qiu, S. Hauert, X. Wang, *et al.*, "Synthetic and living micropropellers for convection-enhanced nanoparticle transport," *Science advances*, vol. 5, no. 4, p. eaav4803, 2019.
- [7] T. Sun, Y. Zhang, C. Power, P. M. Alexander, J. T. Sutton, M. Aryal, N. Vykhodtseva, E. L. Miller, and N. J. McDannold, "Closed-loop control of targeted ultrasound drug delivery across the blood–brain/tumor barriers in a rat glioma model," *Proceedings of the National Academy of Sciences*, vol. 114, no. 48, pp. E10281–E10290, 2017.
- [8] A. K. Hoshiar, T.-A. Le, F. U. Amin, M. O. Kim, and J. Yoon, "Studies of aggregated nanoparticles steering during magnetic-guided drug delivery in the blood vessels," *Journal of Magnetism and Magnetic Materials*, vol. 427, pp. 181–187, 2017.
- [9] A. K. Hoshiar, T.-A. Le, F. U. Amin, M. O. Kim, and J. Yoon, "A novel magnetic actuation scheme to disaggregate nanoparticles and enhance passage across the blood–brain barrier," *Nanomaterials*, vol. 8, no. 1, p. 3, 2018.
- [10] F. U. Amin, A. K. Hoshiar, T. D. Do, Y. Noh, S. A. Shah, M. S. Khan, J. Yoon, and M. O. Kim, "Osmotin-loaded magnetic nanoparticles with electromagnetic guidance for the treatment of alzheimer's disease," *Nanoscale*, vol. 9, no. 30, pp. 10619–10632, 2017.
- [11] A. K. Hoshiar, T.-A. Le, F. U. Amin, M. O. Kim, and J. Yoon, "Studies of aggregated nanoparticles steering during magnetic-guided drug delivery in the blood vessels," *Journal of Magnetism and Magnetic Materials*, vol. 427, pp. 181–187, 2017.

- [12] A. Tariverdi, V. K. Venkiteswaran, Ø. G. Martinsen, O. J. Elle, J. Tørresen, and S. Misra, “Dynamic modeling of soft continuum manipulators using lie group variational integration,” *Plos one*, vol. 15, no. 7, p. e0236121, 2020.
- [13] A. Tariverdi, V. K. Venkiteswaran, M. Richter, O. J. Elle, J. Tørresen, K. Mathiassen, S. Misra, and Ø. G. Martinsen, “A recurrent neural-network-based real-time dynamic model for soft continuum manipulators,” *Frontiers in Robotics and AI*, vol. 8, p. 631303, 2021.
- [14] A. Tariverdi, K. Mathiassen, V. Søyseth, H. Kalvøy, O. J. Elle, J. Tørresen, Ø. G. Martinsen, and M. E. Høvin, “Physics-based simulation and control framework for steering a magnetically-actuated guidewire,” in *2022 8th International Conference on Control, Decision and Information Technologies (CoDIT)*, vol. 1, pp. 770–774, IEEE, 2022.
- [15] M. B. Khamesee, N. Kato, Y. Nomura, and T. Nakamura, “Design and control of a microrobotic system using magnetic levitation,” *IEEE/ASME transactions on mechatronics*, vol. 7, no. 1, pp. 1–14, 2002.
- [16] M. Salehizadeh and E. Diller, “Two-agent formation control of magnetic microrobots in two dimensions,” *Journal of Micro-Bio Robotics*, vol. 12, no. 1, pp. 9–19, 2017.
- [17] E. Diller, J. Giltinan, and M. Sitti, “Independent control of multiple magnetic microrobots in three dimensions,” *The International Journal of Robotics Research*, vol. 32, no. 5, pp. 614–631, 2013.
- [18] A. Zarrouk, K. Belharet, O. Tahri, and A. Ferreira, “A four-magnet system for 2d wireless open-loop control of microrobots,” in *2019 International Conference on Robotics and Automation (ICRA)*, pp. 883–888, IEEE, 2019.
- [19] P. Ryan and E. Diller, “Magnetic actuation for full dexterity microrobotic control using rotating permanent magnets,” *IEEE Transactions on Robotics*, vol. 33, no. 6, pp. 1398–1409, 2017.
- [20] H. Keller, A. Juloski, H. Kawano, M. Bechtold, A. Kimura, H. Takizawa, and R. Kuth, “Method for navigation and control of a magnetically guided capsule endoscope in the human stomach,” in *2012 4th IEEE RAS & EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pp. 859–865, IEEE, 2012.
- [21] E. Diller, S. Floyd, C. Pawashe, and M. Sitti, “Control of multiple heterogeneous magnetic microrobots in two dimensions on nonspecialized surfaces,” *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 172–182, 2011.
- [22] R. Pieters, H.-W. Tung, S. Charreyron, D. F. Sargent, and B. J. Nelson, “Rodbot: A rolling microrobot for micromanipulation,” in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 4042–4047, IEEE, 2015.
- [23] S. J. Kim, S. M. Jeon, J. K. Nam, and G. H. Jang, “Closed-loop control of a self-positioning and rolling magnetic microrobot on 3d thin surfaces using biplane imaging,” *IEEE Transactions on Magnetics*, vol. 50, no. 11, pp. 1–4, 2014.
- [24] T. Xu, G. Hwang, N. Andreff, and S. Régnier, “Planar path following of 3-d steering scaled-up helical microswimmers,” *IEEE Transactions on Robotics*, vol. 31, no. 1, pp. 117–127, 2015.
- [25] X. Wang, J. Liu, Y. Zhang, B. Shi, D. Jiang, and H. Peng, “A unified symplectic pseudospectral method for motion planning and tracking control of 3d underactuated overhead cranes,” *International Journal of Robust and Nonlinear Control*, vol. 29, no. 7, pp. 2236–2253, 2019.
- [26] H. Peng, F. Li, J. Liu, and Z. Ju, “A symplectic instantaneous optimal control for robot trajectory tracking with differential-algebraic equation models,” *IEEE Transactions on Industrial Electronics*, vol. 67, no. 5, pp. 3819–3829, 2019.
- [27] D. Jiang, Z. Cai, H. Peng, and Z. Wu, “Coordinated control based on reinforcement learning for dual-arm continuum manipulators in space capture missions,” *Journal of Aerospace Engineering*, vol. 34, no. 6, p. 04021087, 2021.
- [28] R. Padmanabhan, N. Meskin, and W. M. Haddad, “Reinforcement learning-based control of drug dosing for cancer chemotherapy treatment,” *Mathematical biosciences*, vol. 293, pp. 11–20, 2017.
- [29] E. Bejar and A. Moran, “Deep reinforcement learning based neuro-control for a two-dimensional magnetic positioning system,” in *2018 4th International Conference on Control, Automation and Robotics (ICCAR)*, pp. 268–273, IEEE, 2018.
- [30] D. Xu, Y. Zhang, W. Tan, and H. Wei, “Reinforcement learning control of a novel magnetic actuated flexible-joint robotic camera system for single incision laparoscopic surgery,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 1236–1241, IEEE, 2021.
- [31] R. Zhang, Q. Lv, J. Li, J. Bao, T. Liu, and S. Liu, “A reinforcement learning method for human-robot collaboration in assembly tasks,” *Robotics and Computer-Integrated Manufacturing*, vol. 73, p. 102227, 2022.
- [32] D. Jiang, Z. Cai, Z. Liu, H. Peng, and Z. Wu, “An integrated tracking control approach based on reinforcement learning for a continuum robot in space capture missions,” *Journal of Aerospace Engineering*, vol. 35, no. 5, p. 04022065, 2022.
- [33] H. Hu, K. Zhang, A. H. Tan, M. Ruan, C. Agia, and G. Nejat, “A sim-to-real pipeline for deep reinforcement learning for autonomous robot navigation in cluttered rough terrain,” *IEEE Robotics and Automation Letters*, vol. 6, no. 4, pp. 6569–6576, 2021.
- [34] V. J. Lumelsky and T. Skewis, “Incorporating range sensing in the robot navigation function,” *IEEE transactions on Systems, Man, and Cybernetics*, vol. 20, no. 5, pp. 1058–1069, 1990.
- [35] J. Borenstein, Y. Koren, et al., “The vector field histogram-fast obstacle avoidance for mobile robots,” *IEEE transactions on robotics and automation*, vol. 7, no. 3, pp. 278–288, 1991.
- [36] R. Simmons, “The curvature-velocity method for local obstacle avoidance,” in *Proceedings of IEEE international conference on robotics and automation*, vol. 4, pp. 3375–3382, IEEE, 1996.
- [37] R. Diankov and J. Kuffner, “Randomized statistical path planning,” in *2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1–6, IEEE, 2007.
- [38] B. Burns and O. Brock, “Toward optimal configuration space sampling,” in *Robotics: Science and Systems*, pp. 105–112, Citeseer, 2005.
- [39] C. Sprunk, B. Lau, P. Pfaffz, and W. Burgard, “Online generation of kinodynamic trajectories for non-circular omnidirectional robots,” in *2011 IEEE International Conference on Robotics and Automation*, pp. 72–77, IEEE, 2011.
- [40] S. M. Khansari-Zadeh and A. Billard, “A dynamical system approach to realtime obstacle avoidance,” *Autonomous Robots*, vol. 32, no. 4, pp. 433–454, 2012.
- [41] J. Bruce and M. M. Veloso, “Real-time randomized path planning for robot navigation,” in *Robot soccer world cup*, pp. 288–295, Springer, 2002.
- [42] C. J. Watkins and P. Dayan, “Q-learning,” *Machine learning*, vol. 8, no. 3–4, pp. 279–292, 1992.
- [43] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, “Playing atari with deep reinforcement learning,” *arXiv preprint arXiv:1312.5602*, 2013.
- [44] H. Van Hasselt, A. Guez, and D. Silver, “Deep reinforcement learning with double q-learning,” *arXiv preprint arXiv:1509.06461*, 2015.
- [45] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, “Prioritized experience replay,” *arXiv preprint arXiv:1511.05952*, 2015.
- [46] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, and N. Freitas, “Dueling network architectures for deep reinforcement learning,” in *International conference on machine learning*, pp. 1995–2003, 2016.
- [47] M. G. Bellemare, W. Dabney, and R. Munos, “A distributional perspective on reinforcement learning,” *arXiv preprint arXiv:1707.06887*, 2017.
- [48] M. Fortunato, M. G. Azar, B. Piot, J. Menick, I. Osband, A. Graves, V. Mnih, R. Munos, D. Hassabis, O. Pietquin, et al., “Noisy networks for exploration,” *arXiv preprint arXiv:1706.10295*, 2017.
- [49] M. Hessel, J. Modayil, H. Van Hasselt, T. Schaul, G. Ostrovski, W. Dabney, D. Horgan, B. Piot, M. Azar, and D. Silver, “Rainbow: Combining improvements in deep reinforcement learning,” *arXiv preprint arXiv:1710.02298*, 2017.
- [50] R. S. Sutton, “Learning to predict by the methods of temporal differences,” *Machine learning*, vol. 3, no. 1, pp. 9–44, 1988.
- [51] R. Diankov, “Automated construction of robotic manipulation programs,” 2010.
- [52] A. Zeng, S. Song, S. Welker, J. Lee, A. Rodriguez, and T. Funkhouser, “Learning synergies between pushing and grasping with self-supervised deep reinforcement learning,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4238–4245, IEEE, 2018.
- [53] W. Dabney, G. Ostrovski, D. Silver, and R. Munos, “Implicit quantile networks for distributional reinforcement learning,” in *International conference on machine learning*, pp. 1096–1105, PMLR, 2018.
- [54] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.
- [55] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al., “Human-level control through deep reinforcement learning,” *nature*, vol. 518, no. 7540, pp. 529–533, 2015.
- [56] J. S. Obando-Ceron and P. S. Castro, “Revisiting rainbow: Promoting more insightful and inclusive deep reinforcement learning research,” *arXiv preprint arXiv:2011.14826*, 2020.
- [57] W. L. Brogan, *Modern control theory*. Pearson education india, 1991.
- [58] B. M. Chen, Z. Lin, and Y. Shamash, *Linear Systems Theory*. 2004.

- [59] S. M. LaValle *et al.*, “Rapidly-exploring random trees: A new tool for path planning,” 1998.
- [60] R. A. Finkel and J. L. Bentley, “Quad trees a data structure for retrieval on composite keys,” *Acta informatica*, vol. 4, no. 1, pp. 1–9, 1974.
- [61] K. Mathiassen, J. E. Fjellin, K. Glette, P. K. Hol, and O. J. Elle, “An ultrasound robotic system using the commercial robot ur5,” *Frontiers in Robotics and AI*, vol. 3, p. 1, 2016.
- [62] A. R. Mahmood, D. Korenkevych, B. J. Komer, and J. Bergstra, “Setting up a reinforcement learning task with a real-world robot,” in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 4635–4640, IEEE, 2018.
- [63] E. Coumans and Y. Bai, “Pybullet, a python module for physics simulation for games, robotics and machine learning.” pybullet.org, 2016–2019.
- [64] K. Bousmalis, A. Irpan, P. Wohlhart, Y. Bai, M. Kelcey, M. Kalakrishnan, L. Downs, J. Ibarz, P. Pastor, K. Konolige, *et al.*, “Using simulation and domain adaptation to improve efficiency of deep robotic grasping,” in *2018 IEEE international conference on robotics and automation (ICRA)*, pp. 4243–4250, IEEE, 2018.



Abbas Tariverdi received the B.Sc. and M.Sc. degree in electrical and control engineering from Amirkabir University of Technology (Tehran Polytechnic), Tehran, Iran, in 2014 and 2016, respectively. He is currently pursuing the Ph.D. degree with the Department of Physics, University of Oslo, Norway. The topic of his Ph.D. thesis is modeling and control of a continuum manipulator for clinical automation. His main research interests include (non-)linear control, medical robotics, human-robot interaction, and real-time systems.



Ulysse Côté-Allard received the Ph.D. degree in electrical engineering from Université Laval, Québec, QC, Canada, in 2020. He is currently a researcher fellow with the Robotics and Intelligent Systems Research Group, University of Oslo, Norway. His main research interests include rehabilitation engineering, biosignal-based control, and human-robot interaction. He is a recipient of the Best Paper Award from the IEEE Systems, Man, and Cybernetics Conference and the 2022 IEEE Engineering in Medicine and Biology first Prize

Paper Award.



Kim Mathiassen received a Masters degree in engineering cybernetics from the Norwegian University of Science and Technology in 2010 and a PhD in robotics from the University of Oslo in 2017. He currently works at the Norwegian Defence Research Establishment with unmanned ground vehicles and autonomous systems, and has is an associate professor at the University of Oslo teaching robotics. His current research interests span from low level robot control to high level planning and reasoning.



Ole J. Elle received the master’s degree from the Department of Production and Quality Engineering, Norwegian Institute of Technology, Trondheim, Norway, in 1990, and the Ph.D. degree in robotic surgery from the Norwegian University of Science and Technology, Trondheim, in 2004. He is currently the Head of the Section for Technology Research with The Intervention Center, Oslo University Hospital, and also an Adjunct Associate Professor with the Department of Informatics, University of Oslo, Norway.



Håvard Kalvøy received the Ph.D. degree from the Department of Physics, University of Oslo, Norway, in 2010. He is the head of R&D at the Department of Clinical and Biomedical engineering at Oslo University Hospital, Norway. His research is mainly focused on medical sensor technology and the passive, electrical properties of biomaterials.



Ørjan G. Martinsen received the Ph.D. degree from the Department of Physics, University of Oslo, Norway in 1995 on a thesis on the electrical properties of human skin. Since then he has worked at the Department of Physics, University of Oslo, Norway from 2002 as a full professor. He also has a part time position as a senior researcher at the Department of Clinical and Biomedical Engineering at Oslo University Hospital, Norway. His research has mainly been focused on the passive, electrical properties of biomaterials.



Jim Tørresen received the M.Sc. degree in computer architecture and design from the Norwegian University of Science and Technology, in 1991, and the Dr.Ing. (Ph.D.) degree in computer architecture and design from the University of Trondheim, in 1996. He is a full Professor with the Department of Informatics, University of Oslo and a Principal Investigator with the RITMO Centre for Interdisciplinary Studies in Rhythm, Time and Motion. His research interests include artificial intelligence, ethical aspects of AI and robotics, machine learning, robotics, and applying this to complex real-world applications. He is a member of the Norwegian Academy of Technological Sciences (NTVA) and the National Committee for Research Ethics in Science and Technology (NENT). More details can be found at: <http://jimotoer.no/>.