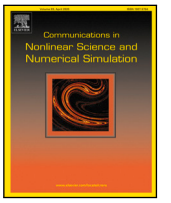
Commun Nonlinear Sci Numer Simulat 128 (2024) 107666

Contents lists available at ScienceDirect

Communications in Nonlinear Science and Numerical Simulation

journal homepage: www.elsevier.com/locate/cnsns



Research paper

Adaptive micro-locomotion in a dynamically changing environment via context detection

Zonghao Zou [a], Yuexin Liu [b], Alan C.H. Tsang [c], Y.-N. Young [d], On Shun Pak [e,*]

[a] Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, NY 14850, USA
[b] Department of Engineering Technology and Industrial Distribution, Texas A&M University, College Station, TX 77843, USA
[c] Department of Mechanical Engineering, The University of Hong Kong, Hong Kong, China
[d] Department of Mathematical Sciences, New Jersey Institute of Technology, Newark, NJ 07102, USA
[e] Department of Applied Mathematics and Department of Mechanical Engineering, Santa Clara University, Santa Clara, CA 95053, USA

ABSTRACT

Substantial efforts have exploited reinforcement learning (RL) in the development of micro-robotic locomotion. These RL-powered micro-robots are capable of learning a locomotory policy based on their experience interacting with the surroundings, without requiring prior knowledge on the physics of locomotion in that environment. However, in their applications, micro-robots often encounter changes in the environment and need to adapt their locomotory gaits like living organisms in order to achieve robust locomotion performance. In standard RL methods, such a non-stationary environment can cause the micro-robots to continuously relearn the policy from scratch, degrading their locomotion performance. In this work, we explore a first use of a recently developed context detection method combined with deep RL to facilitate micro-robotic locomotion in a dynamically changing environment. As a proof-of-principle, we consider a simple micro-robot immersed in non-stationary environments switching between a viscous fluid environment and a dry frictional environment. We show that the RL with context detection approach enables the micro-robot to effectively detect changes in the environment and deploy specialized locomotory gaits for different environments accordingly to achieve significantly improved locomotion. Our results suggest the integration of deep RL with context detection as a potential tool for robust micro-robotic locomotion across different environments.

1. Introduction

In the famous 1959 seminar "There's Plenty of Room at the Bottom" on nanotechnology, Feynman outlined his vision of swallowing a micro-robot that can roam the human body to perform micro-surgery. Building such micro-robots not only require micro-/nanofabrication techniques but also knowledge of their physics of locomotion at the microscopic scale. When entering the microscopic world, various physical forces come out in different proportions; simply scaling down macroscopic locomotion strategies therefore would not work effectively. In particular, swimming at the microscopic scale becomes very challenging due to the dominance of the viscous force over the inertial force [1,2]. The Reynolds number (Re), which compares the magnitude of the viscous force to the inertial force, is negligibly small for micron-sized objects (*e.g.*, 10^{-6} for flagellated bacteria to 10^{-2} for spermatozoa) [3,4]. At such low Re, Purcell's scallop theorem [1] rules out the possibility of generating self-propulsion using any reciprocal motion – a sequence of body configurations that possess time-reversal symmetry. Common macroscopic swimming strategies at high Re, such as flapping motions of wings, therefore become ineffective at low Re. To swim at the microscale, microorganisms have evolved different propulsion strategies with their sophisticated biological molecular machinery. However, without similar molecular machinery as microorganisms do, designing simple synthetic mechanisms that enable self-propulsion at low Re represents a fundamental challenge in developing swimming micro-robots. To this end, Purcell [1] presented a pioneering example with a three-link micro-swimmer to demonstrate how a swimmer can generate self-propulsion in the absence of inertia by

* Corresponding author.
E-mail address: opak@scu.edu (O.S. Pak).

https://doi.org/10.1016/j.cnsns.2023.107666
Received 6 March 2023; Received in revised form 31 October 2023; Accepted 2 November 2023
Available online 10 November 2023

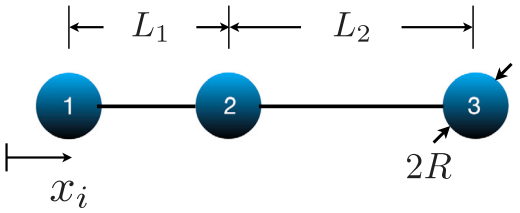


Fig. 1. A model micro-robot consisting of three identical spheres of radius R connected by two extensible arms with lengths L_1 and L_2. The position of the spheres are denoted by the coordinates x_i. The goal of the micro-robot is to perform locomotory gaits by adjusting the arm lengths L_1 and L_2 to generate net displacement in the positive x-direction.

performing cyclic motions in a non-reciprocal manner, overcoming the constraints due to kinematic reversibility. In a latter effort, Najafi and Golestanian [5] presented another elegant example of a micro-swimmer consisting of three linked spheres (Fig. 1), which can self-propel by varying the distances between the spheres. In addition to these ingenious designs, growing cross-disciplinary interest has contributed to the development of swimming micro-robots for potential biomedical applications, such as drug delivery and microsurgery [6,7].

More recent efforts have explored the applications of machine learning techniques in studying locomotion [8–16] and navigation [17–22] problems in fluid environments. In particular, reinforcement learning (RL) has been employed to equip micro-swimmers with the capability to learn effective swimming gaits without prior knowledge of low-Re locomotion [23,24]. For instance, recent studies have shown that a micro-swimmer consisting of linked spheres can successfully acquire swimming gaits [11] previously invented by Najafi and Golestanian [5] and adjust its gaits to navigate in targeted swimming directions [13]. Furthermore, this versatile approach can empower the cooperative swimming of a pair of microswimmers [15] and is adaptable for application with other reconfigurable microswimmers [12,16]. Similar RL techniques can also be applied to learn locomotion strategies in a drastically different environment, such as coordinated crawling movements on frictional surfaces [25,26]. This is particularly important for biomedical applications of micro-robots, which may encounter complex and heterogeneous biological environments consisting of both fluid and solid terrains. Successful biomedical applications of micro-robots therefore rely on their ability to traverse vastly different environments [27–29]. However, micro-robots to date are typically designed for operation in a specific environment. When there is a change in the environment (e.g., from a fluid terrain to a solid terrain), the optimal locomotory gaits in the original environment may become largely ineffective in the new environment. Such a dynamically changing environment is an example of a non-stationary environment in RL, where the dynamics of the environment may change in unknown or unpredictable manners.

A challenge associated with learning in non-stationary environments is "catastrophic forgetting" [30,31], which is the tendency of an agent to quickly forget previously learned policy upon learning new policy when encountering a new environment. Consequently, a micro-robot would need to learn from scratch whenever it transitions from one environment to another, even for an environment that has already been experienced and learned by the micro-robot. The time required for the relearning process degrades the locomotion performance. Moreover, the epochs at which these changes in environment occur are unknown to the micro-robot, adding uncertainty to its overall locomotion performance. In contrast, natural microorganisms can encode memory patterns of previously experienced environmental stimuli, enabling them to robustly adapt to changes in environment [32–35]. Many previous works have focused on improving the RL algorithms to better adapt to the newly occurring information, but few have provided a method to maintain separate data structure suited for a non-stationary environment [36–38]. To address RL in non-stationary environments, it is desirable to have a RL agent that can retain its previously learnt policy in various environments and quickly adapt its policy according to the change in a non-stationary environment.

In this work, we demonstrate a first use of a recently developed context detection method [38] combined with deep RL [39] to facilitate the locomotion of a micro-robot in a dynamically changing environment. Here, the term "context" indicates a type of dynamics in a specific environment; hence, a context change in this work corresponds to a change in the environment in which the micro-robot is immersed. The context detection method computes high-confidence change-point detection statistics, in real time, to detect changes in the environment (or context) and inform the decision-making strategy. The algorithm makes use of the change-point detection statistics to decide whether new policies need to be created and deployed when a change in environment is encountered or previously-optimized policies might be reused. Similar context detection methods have been implemented in applications such as traffic lights control and maze escaping in non-stationary environments [36,38]. Here we introduce its first use in micro-robotic locomotion. Specifically, here we consider the locomotion of a simple model micro-robot consisting of three linked spheres (Fig. 1), which can encounter changes between two environments with drastically different dynamics: a viscous fluid medium and a dry frictional medium. We show that the RL context detection approach can enable the micro-robot to effectively detect changes in the environment and adapt its locomotory gaits accordingly, realizing an "amphibious" micro-robot for both aquatic (viscous fluid medium) and terrestrial (dry frictional medium) locomotion. We compare the locomotion performance of the micro-robot against RL without context detection in various scenarios. Taken together, the results represent a proof-of-principle demonstration of the context detection approach in micro-robotic locomotion and suggest its potential use for locomotion across more complex non-stationary environments in future studies.

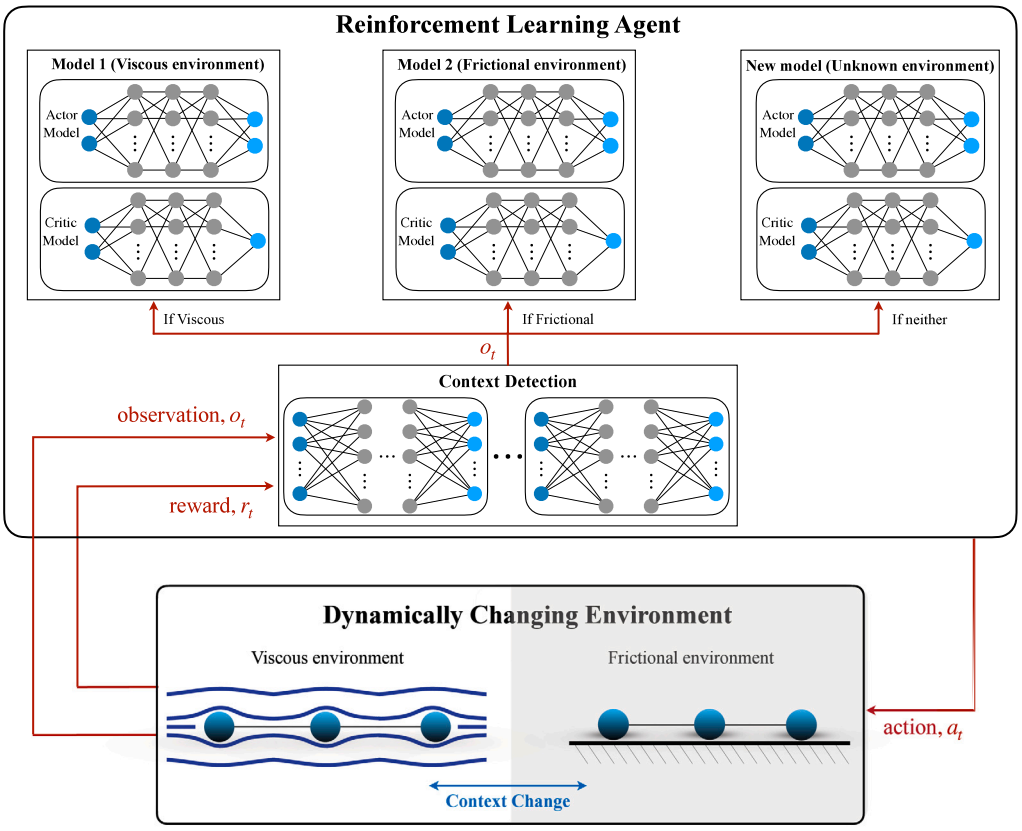


Fig. 2. Schematic of the reinforcement learning with context detection algorithm. The micro-robot performs an action a_t in a dynamically changing environment (bottom), where the current environment (viscous or frictional) is unknown to the micro-robot. After taking the action, the micro-robot reaches a new geometric configuration, which is the next observation o_t. The reward r_t is calculated based on the displacement of the centroid. Receiving the reward and the next observation, the reinforcement learning agent utilizes its context detection method to determine the model that best fits the current environment – a model for the viscous or frictional environment or a new model to be created (Top). Based on the selected model, the agent advises the next action to the micro-robot, and the next iteration begins. During the training process, the agent progressively updates its policy models and context detection method.

2. A model micro-robot for "amphibious" locomotion

We consider a simple reconfigurable system composed of three identical spheres of radius R connected by two arms with variable lengths, L_1 and L_2, as illustrated in Fig. 1. This system generalizes the micro-swimmer first studied by Najafi and Golestanian [5], which permits only discrete actions of a single arm at a given time, by allowing continuous and simultaneous actuation of the two arms. This allows the emergence of more complex gaits for locomotion in environments other than the purely viscous fluid medium considered previously [5]. By symmetry, the micro-robot can only generate motion in the x-direction, and the position of sphere i is denoted by the coordinate x_i. The centroid of the micro-robot is therefore given by $x_c = \sum_i x_i/3$. The overall goal of the micro-robot is to acquire effective locomotory gaits that generate a net displacement of its centroid in the positive x-direction in different environments. We immerse the micro-robot in a non-stationary environment that can change dynamically between a viscous fluid medium and a dry frictional medium. We consider these two drastically different media as a proof-of-principle of an "amphibious" micro-robot that adapts its locomotory gaits for both aquatic (viscous fluid medium) and terrestrial (dry friction medium) locomotion.

To simulate the dynamics of the micro-robot in a purely viscous fluid environment, we model the hydrodynamic interaction between the spheres by the Oseen tensor [5,40]. This approximation is valid in the asymptotic limit where the spheres are relatively far apart (*i.e.*, $R \ll L_1, L_2$). We also assume that the arms have negligible hydrodynamic influences. The velocity of the spheres are related to the forces by $\dot{x}_i = G_{ij} F_j$, where G_{ij} is the one-dimensional Oseen tensor [41–43] given by

$$G_{ij} = \begin{cases} \frac{1}{6\pi\mu R} & \text{if } i = j \\ \frac{1}{4\pi\mu|x_i - x_j|} & \text{if } i \neq j, \end{cases} \tag{1}$$

μ is the dynamic viscosity of the fluid, and F_j is the hydrodynamic force on the sphere j. The arm actuation rates $\dot{L}_i$ are related to the position of the spheres kinematically as $\dot{L}_1 = \dot{x}_2 - \dot{x}_1$ and $\dot{L}_2 = \dot{x}_3 - \dot{x}_2$. The kinematics of the micro-robot is fully determined upon applying the force-free condition, $\sum_i F_i = 0$, at low Re, where inertial effect is considered negligible.

To simulate the locomotion of the micro-robot on a dry frictional medium, we consider a standard Coulomb sliding friction law, which has been applied to study different locomotion problems on land [44,45]. The friction experienced by the spheres F_i depends on the net driving force f_i exerted on these spheres by the arms and the Coulomb sliding friction F_c as

$$F_i = \begin{cases} -F_c \dot{x}_i/|\dot{x}_i|, & \text{if } |f_i| > F_c \\ -f_i, & \text{if } |f_i| \le F_c. \end{cases} \tag{2}$$

Here, when the net driving force f_i on the sphere is greater than the Coulomb sliding friction F_c (*i.e.*, $|f_i| > F_c$), the sphere experiences a constant frictional force given by F_c, independent of the magnitude of its velocity. When $|f_i| \le F_c$, the sphere instead experiences a static friction that balances the net driving force, $F_i = -f_i$. Consistent with locomotion in the low Re regime, here we again consider the physical regime where inertial effect is negligible [45]. The motion of the spheres is therefore determined by enforcing the force free condition in the inertialess regime.

We consider a non-stationary environment that changes between the purely viscous fluid medium and frictional medium described above, where the change of the environment occurs instantaneously [46]. The micro-robot is not prescribed any locomotory gaits *a priori*. Instead, we apply a deep RL approach combined with a context detection method to enable the micro-robot to learn and adapt its locomotory gaits autonomously in such a dynamically changing environment. In this work, we scale lengths by the fully extended arm length L, velocities by the maximum arm actuation rate V_c, time by L/V_c, and forces by $\mu L V_c$. Hereafter we refer only to scaled quantities and use the same symbols for convenience. The micro-robot can vary the arm lengths L_1 and L_2 in the range of [0.6, 1]. In our simulations, during each action step we consider uniform arm actuation rates $\dot{L}_1$ and $\dot{L}_2$ in the range of [−1, 1]. We set the time duration for each action step as $\Delta t = 0.1$.

3. Reinforcement learning with context detection

We employ a deep RL approach [8,13,22,39] to train the micro-robot to generate net displacement of its centroid in the positive x-direction. In the RL algorithm, the state $s \in (x_1, x_2, x_3)$ contains all the x-coordinates of the spheres. The observation $o \in (L_1, L_2)$ is extracted from the state as geometric configurations. The RL agent determines the next action based on the current observation through an actor neural network. The micro-robot then performs the action $a \in (\dot{L}_1, \dot{L}_2)$ by actuating both of its arms for the duration of one action step. The RL agent evaluates the success of the action by measuring the net centroid displacement: $r_t = x_{c_{t+1}} - x_{c_t}$. The training process is divided into a total of N_e episodes, with each episode containing $N_t = 100$ action steps. We randomly initialize a state s_0 at the beginning of each episode to facilitate a full exploration of the observation space. The actor and critic networks are further updated for every 20 episodes by maximizing the expected long-term rewards $\mathbb{E}[R_{t=0}|\pi_\phi]$. Here, π_ϕ is the stochastic control policy, $R_t = \sum_{t'}^{\infty} \gamma^{t'-t} r_{t'}$ is the infinite-horizon discounted future returns, and γ is the discount factor measuring the greediness of the algorithm. We set $\gamma = 0.99$ to ensure the farsightedness of the RL algorithm.

The RL framework described above has been shown effective for learning locomotory gaits for a single, stationary environment [8,13]. Here we adapt this RL framework for a non-stationary environment by applying it in each individual environment with a context change detection algorithm. For context change detection, we employ a high confidence change point detection method developed recently by Alegre et al. [38], which enables the agent to quickly detect a change of environment and trains a set of partial models for different environments as illustrated by Fig. 2. Consider a non-stationary environment consisting of a list of environments $E_1 \ldots E_K$. Let C be a random environment change point switching from E_i to E_j. A proper detection method should consider the time it takes to detect the change point C as well as false detection before C occurs. The high confidence change point detection method minimizes both by computing quality signals based on the experience of the micro-robot (*i.e.*, the action performed, the state transition, and the reward). The quality signals, $W_{k,t} = \max(0, W_{k,t-1} + L_{k,t})$, are computed for each partial model k at every action step t utilizing a multivariate variant of cumulative sum (MCUSUM) method [38], where $L_{k,t}$ is the log-likelihood ratio indicating how likely a particular model k becomes a better fit than the current model. The algorithm will activate the partial model with the highest quality signal that surpasses a threshold h and thereby enables the swimmer to adapt its locomotory gaits in response to the change of environment. Furthermore, in the context detection algorithm, a new partial model will be generated when all other partial models become ineffective in describing the current environment, allowing the micro-robot to explore unlimited distinct environments (Fig. 2 top, see Supplementary Information for more details of the algorithm).

4. Results and discussion

4.1. Training and locomotory gaits

In Fig. 3(a), we show the average training results of RL agents with context detection in a non-stationary environment switching between the viscous and frictional media (blue line). In the training process, we periodically switch between the two media every 100 episodes, which was empirically found to provide satisfactory training outcomes, leading to the oscillatory episode reward shown in Fig. 3(a). As the training proceeds, the RL agent with context detection gradually builds up its experience with the two media, which is used to improve its ability to both detect a change in the environment and develop specialized locomotory strategies of the two partial models. We contrast the result against those by RL agents trained separately without context detection in the viscous

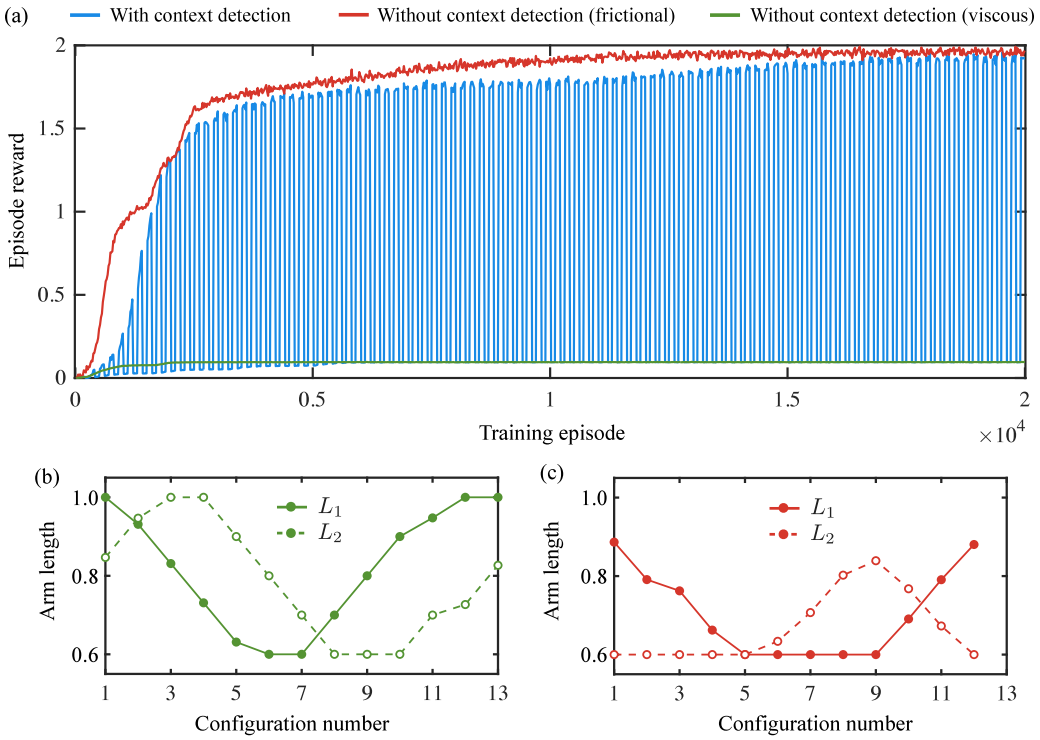


Fig. 3. (a) The average episode reward of the training process is plotted against the training episode. Generally, all RL agents increase their episode reward as the number of training episode increases. Each line represents the training results averaged over five agents. For the RL agents trained in the viscous (green line) and frictional (red line) environments without context detection, the values of episode reward quickly plateau. We periodically switch between two environment for the RL agents with context detection (blue), leading to the oscillatory behaviour in the episode reward. As the training proceeds, the RL agents with context detection gradually improve the partial models and eventually reach a similar episode reward for both environments. (b) A typical cycle of the swimming gaits used by the RL agents with context detection in the viscous fluid medium. The lengths of the two arms, L_1 and L_2, modulate in a manner akin to harmonic oscillations with a mismatch in phases. (c) A typical cycle of the crawling gaits used by the RL agents with context detection in the frictional medium. The crawling gaits are characterized by the use of only a single arm while the other arm remains static and fully contracted, significantly different from the swimming gaits.

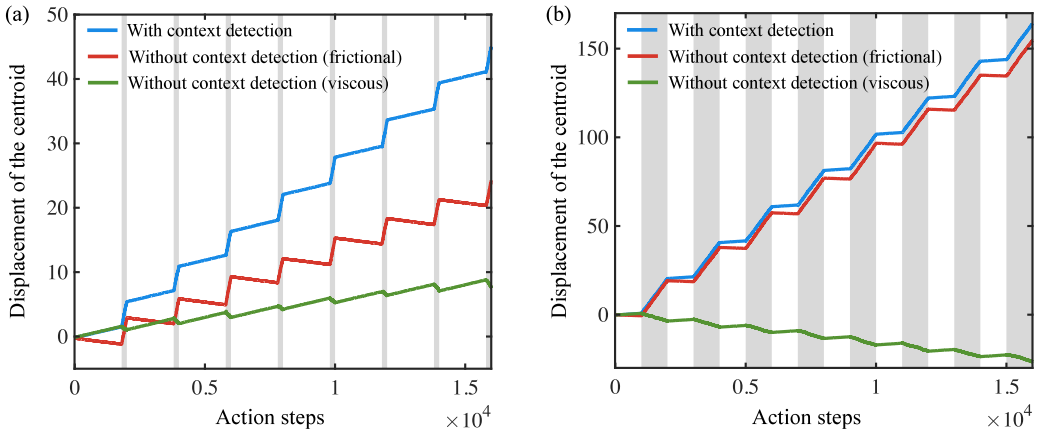


Fig. 4. Displacement of the micro-robot's centroid in non-stationary environments that periodically switch between viscous (white regions) and frictional (grey regions) media, with (a) 10% and (b) 50% of the total duration in the frictional medium. The blue lines track the performance of RL agents with context detection, whereas the green and red lines track, respectively, the performance of RL agents trained separately in the viscous and friction environments without context detection.

environment (green line) and the frictional environment (red line). We observe that the episode reward of the RL agents with context detection gradually approaches to those by RL agents trained only for specific environments. In particular, the RL agents with context detection quickly acquire similar levels of reward as the RL agents trained in the viscous environment at around 6×10^3 episodes.

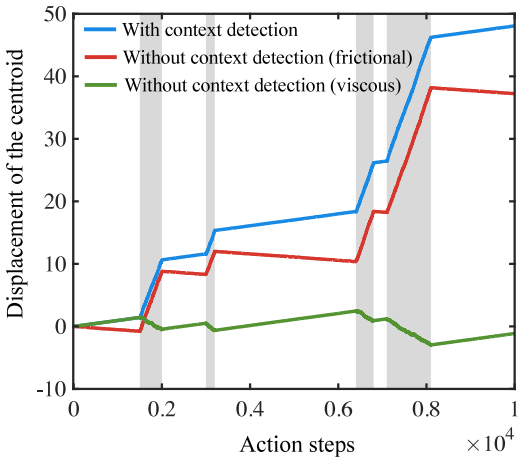


Fig. 5. Displacement of the micro-robot's centroid in a non-stationary environment with non-periodic changes between viscous (white regions) and frictional (grey regions) media. The blue line tracks the performance of the RL agent with context detection, whereas the green and red lines track, respectively, the performance of RL agents trained separately in the viscous and friction environments without context detection.

After around 2×10^4 episodes, the RL agent with context detection successfully learns a set of locomotory gaits that generate similar episode rewards in both environments, while also acquiring the ability to effectively detect environment change. The RL agents with context detection are capable of detecting a change in the environment with typically 3 to 7 action steps. As a remark, we note that, towards the end of the training, the episode rewards in the frictional environment (approximately 2) are much larger than that in the viscous environment (approximately 0.1) as shown in Fig. 3(a). This is due to the intrinsic difference in the dynamics of the two environments; the effective crawling gaits in the frictional environment displace the micro-robot substantially more than the effective swimming gaits do in the viscous medium.

In Fig. 3(b) and (c), we visualize the locomotory gaits employed by the RL agent with context detection in the viscous [panel (b)] and frictional [panel (c)] environments by plotting the variation of the arm lengths L_1 and L_2 of a representative stroke sequence. Since uniform actuation rates are employed for each action step, the arm length variations are piece-wise linear functions. We note the significant differences between the locomotory gaits in these two environments: in the viscous environment, the micro-robot actuates the two arms in a manner akin to harmonic oscillations with a mismatch in phases [Fig. 3(b); see also SI Movie 1], reminiscent of the swimming gait previously studied for the Najafai–Golestanian swimmer [47]; the mismatch in phases was shown essential for generating net swimming motion in a viscous fluid medium. In contrast, the crawling gaits in the frictional environment are characterized by some sequential movements of only a single arm while the other arm remains static in the fully contracted state [Fig. 3(c); see also SI Movie 2]. The RL agents are capable of learning these specialized locomotory gaits without any prior knowledge of the locomotion physics in these different environments. We also note that the epochs at which the changes in environment occur are unknown to the micro-robot. With context detection, the micro-robot can detect the changes in the environment and deploy the swimming and crawling gaits adaptively (see SI Movie 3).

4.2. Locomotion performance in different non-stationary environments

To illustrate the advantage of the context detection method, we next compare the locomotion performance of the RL agents with context detection against RL agents without context detection but trained separately for the viscous or frictional environment, by placing them in different non-stationary environments. We track their centroid displacements as they perform different actions in a non-stationary environment. In Fig. 4(a), we consider a non-stationary environment that periodically switches between the viscous (white regions) and frictional (grey regions) media, with 10% of the total duration in the frictional medium. We observe that the RL agent with context detection (blue line) shows substantially better performance in terms of the larger displacement of the centroid than the RL agents without context detection (red and green lines), demonstrating the significance of detecting the change in the environment and adaptively deploying specialized locomotory gaits for the corresponding environments. Specifically, the RL agent with context detection displays an approximately six-fold (two-fold) enhancement in the centroid displacement compared with the RL agent trained in the viscous (frictional) environment at the end of the simulation. In particular, we observe that the RL agent trained in the frictional environment (red line) is able to move in the desired direction only when it is in the frictional environment; it indeed moves in the opposite direction when the environment becomes viscous, as indicated by the negative slopes in the white regions shown in Fig. 4(a). Similarly, the locomotion policy of the RL agent trained in the viscous environment (green line) is only effective in the viscous environment, moving the agent in the opposite direction when the environment becomes frictional. In stark contrast, the RL agent with context detection (blue line) always moves in the target direction by adapting its locomotory gaits based on the detected changes in the environment, as indicated by the positive slopes in all regions.

We note that the level of enhancement in the locomotion performance of the RL agent with context detection depends on the composition of the non-stationary environment. To illustrate, we consider in Fig. 4(b) a non-stationary environment that switches periodically between viscous and frictional environments with equal duration in both environments. In such a non-stationary environment, the disadvantage of the RL agent trained in the viscous environment (green line) becomes more apparent because of the longer duration spent in the frictional environment, where the agent moves in the wrong direction; overall, the RL agent moves in the negative x-direction at the end of the simulation in this non-stationary environment, opposite to the target direction. On the other hand, the RL agent trained in the frictional environment (red) has significantly improved performance because of the increased duration spent in the frictional medium. In this non-stationary environment, the RL agent with context detection still displays enhanced locomotion performance relative to the RL agent trained in frictional environment, but in a diminished manner compared with that in the non-stationary environment in Fig. 4(a).

Last, we remark that although periodic changes in the environment are considered in Fig. 4, the epochs at which the changes in environment occur are unknown to the agent. The context detection approach continues to work whether the change of environment is periodic or not. To illustrate this point, we consider in Fig. 5 a non-stationary environment with non-periodic changes between the viscous (white regions) and frictional (grey regions) media. We observe that the RL agent with context detection (blue line) remains capable of detecting the change in the environment and adapting its locomotory gaits to always move in the positive x-direction in such a non-stationary environment with non-periodic changes. This capability again results in superior locomotion performance compared with RL agents without context detection (green and red lines), illustrating the impact of the context detection method.

5. Conclusions

In their potential applications, micro-robots may encounter heterogeneous surrounding environments including both fluid and solid terrains. Optimal locomotory gaits in one environment, however, may become largely ineffective in a different environment. To achieve robust locomotion performance, these micro-robots need to adapt their locomotory gaits like living organisms in response to changes in the environment. In this work, we explore a first use of a context detection method combined with deep RL as a potential tool to facilitate micro-robotic locomotion in a dynamically changing environment. As a model micro-robot, we consider a simple reconfigurable system consisting of three linked spheres immersed in a non-stationary environment that switches between a viscous fluid medium and a dry frictional medium. These two specific media are considered here to realize an ‘‘amphibious’’ micro-robot performing aquatic (viscous fluid medium) and terrestrial (dry frictional medium) locomotion. We demonstrate that the use of high-confidence change-point detection statistics empowers the RL agent to detect changes in the environment and deploy effective swimming (crawling) gaits in the viscous fluid (dry frictional) environment, leading to superior locomotion performance compared with RL agents without context detection. These results serve as a proof-of-principle of integrating deep RL with context detection to enable smart micro-robotic locomotion in a dynamically changing environment.

We remark on some limitations of the current study and suggest potential directions for subsequent investigations. First, the context change detection method assumes individual environments to be distinct, meaning that either the effective locomotory strategies (observation transition function) or the reward function should be significantly different in order for the method to be effective. A non-stationary environment consisting of environments with very similar observation transition and reward functions may cause failure in training the detection mechanism; however, even without context detection in these scenarios, the RL agent is expected to achieve sufficiently effective performance in similar environments. Second, as a first step in modelling a dynamically changing environment, the change in environment is assumed to be instantaneous in this work. Subsequent efforts should account for a transition phase where the micro-robot moves across the boundary between two different environments. In this more complex scenario, the heterogeneity in the dynamics of individual parts of the micro-robot will need to be captured by the governing equations. It will be interesting to examine the locomotion strategy and the use of the context detection approach during such a transition phase in future work. Third, the influences of external disturbances such as flows, obstacles, and Brownian noise [48–52] is another important practical aspect to address, given the small sizes of the micro-robots and the presence of other uncontrolled environmental factors within complex biological settings. Understanding the impact caused by these hydrodynamic and thermal fluctuations is an ongoing focus, both during the training phase of the micro-robot and its resulting navigation performance. Finally, we consider in this work a non-stationary environment consisting of the purely viscous fluid and dry frictional media only as a proof-of-principle demonstration. In their potential medical applications, micro-robots are expected to traverse diverse and complex biological environments, including different types of non-Newtonian fluids (e.g., blood and mucus) and lubricated solid surfaces (e.g., gastrointestinal walls). Without adaptability like living organisms, it remains formidable for micro-robots to achieve robust locomotion performance across these different environments. The proof-of-principle demonstration here suggests the possibility of addressing these outstanding challenges by integrating deep RL with context detection to empower adaptive locomotion of micro-robots across different environments.

CRediT authorship contribution statement

Zonghao Zou: Methodology, Software, Validation, Formal analysis, Investigation, Visualization, Writing – original draft. **Yuexin Liu:** Methodology, Software, Validation, Formal analysis, Investigation, Visualization, Writing – original draft. **Alan C.H. Tsang:** Conceptualization, Methodology, Investigation, Writing – review & editing, Supervision, Funding acquisition. **Y.-N. Young:** Conceptualization, Methodology, Investigation, Writing – review & editing, Supervision, Funding acquisition. **On Shun Pak:** Conceptualization, Methodology, Investigation, Writing – review & editing, Supervision, Funding acquisition, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Acknowledgements

Funding support by the National Science Foundation, United States (Grant Nos. 1830958, 1931292, and 2323046 to O.S.P. and Grants Nos. 1614863 and 1951600 to Y.N.Y.) is gratefully acknowledged. Y.N.Y. acknowledges support from Flatiron Institute, part of Simons Foundation, United States. A.C.H.T. acknowledges funding support from the Croucher Foundation, Hong Kong. Z.Z. and O.S.P. also acknowledge the computational resources from the WAVE computing facility at Santa Clara University, United States.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.cnsns.2023.107666.

References

[1] Purcell EM. Life at low Reynolds number. Amer J Phys 1977;45:3–11.
[2] Yeomans JM, Pushkin DO, Shum H. An introduction to the hydrodynamics of swimming microorganisms. Eur Phys J Spec Top 2014;223(9):1771–85.
[3] Fauci LJ, Dillon R. Biofluidmechanics of reproduction. Annu Rev Fluid Mech 2006;38(1):371–94. http://dx.doi.org/10.1146/annurev.fluid.37.061903.175725.
[4] Lauga E, Powers TR. The hydrodynamics of swimming microorganisms. Rep Progr Phys 2009;72:096601.
[5] Najafi A, Golestanian R. Simple swimmer at low Reynolds number: Three linked spheres. Phys Rev E 2004;69:062901. http://dx.doi.org/10.1103/PhysRevE.69.062901.
[6] Nelson BJ, Kaliakatsos IK, Abbott JJ. Microrobots for minimally invasive medicine. Annu Rev Biomed Eng 2010;12(1):55–85.
[7] Gao W, Wang J. The environmental impact of micro/nanomachines: A review. ACS Nano 2014;8:3170–80.
[8] Jiao Y, Ling F, Heydari S, Heess N, Merel J, Kanso E. Learning to swim in potential flow. Phys Rev Fluids 2021;6:050505. http://dx.doi.org/10.1103/PhysRevFluids.6.050505, URL https://link.aps.org/doi/10.1103/PhysRevFluids.6.050505.
[9] Gazzola M, Hejazialhosseini B, Koumoutsakos P. Reinforcement learning and wavelet adapted vortex methods for simulations of self-propelled swimmers. SIAM J Sci Comput 2014;36(3):B622–39. http://dx.doi.org/10.1137/130943078.
[10] Gazzola M, Tchieu AA, Alexeev D, de Brauer A, Koumoutsakos P. Learning to school in the presence of hydrodynamic interactions. J Fluid Mech 2016;789:726–49. http://dx.doi.org/10.1017/jfm.2015.686.
[11] Tsang ACH, Tong PW, Nallan S, Pak OS. Self-learning how to swim at low Reynolds number. Phys Rev Fluids 2020;5:074101. http://dx.doi.org/10.1103/PhysRevFluids.5.074101, URL https://link.aps.org/doi/10.1103/PhysRevFluids.5.074101.
[12] Liu Y, Zou Z, Tsang ACH, Pak OS, Young Y-N. Mechanical rotation at low Reynolds number via reinforcement learning. Phys Fluids 2021;33(6):062007. http://dx.doi.org/10.1063/5.0053563.
[13] Zou Z, Liu Y, Young Y-N, Pak OS, Tsang ACH. Gait switching and targeted navigation of microswimmers via deep reinforcement learning. Commun Phys 2022;5(158). http://dx.doi.org/10.1038/s42005-022-00935-x.
[14] Paz S, Ausas RF, Carbajal JP, Buscaglia GC. Chemoreception and chemotaxis of a three-sphere swimmer. Commun Nonlinear Sci Numer Simul 2023;117:106909.
[15] Liu Y, Zou Z, Pak OS, Tsang ACH. Learning to cooperate for low-Reynolds-number swimming: a model problem for gait coordination. Sci Rep 2023;13:9397.
[16] Qin K, Zou Z, Zhu L, Pak OS. Reinforcement learning of a multi-link swimmer at low Reynolds numbers. Phys Fluids 2023;35(3):032003.
[17] Colabrese S, Gustavsson K, Celani A, Biferale L. Flow navigation by smart microswimmers via reinforcement learning. Phys Rev Lett 2017;118:158004. http://dx.doi.org/10.1103/PhysRevLett.118.158004, URL https://link.aps.org/doi/10.1103/PhysRevLett.118.158004.
[18] Gustavsson K, Biferale L, Celani A, Colabrese S. Finding efficient swimming strategies in a three-dimensional chaotic flow by reinforcement learning. Eur Phys J E 2017;40(12):110. http://dx.doi.org/10.1140/epje/i2017-11602-9.
[19] Colabrese S, Gustavsson K, Celani A, Biferale L. Smart inertial particles. Phys Rev Fluids 2018;3:084301. http://dx.doi.org/10.1103/PhysRevFluids.3.084301, URL https://link.aps.org/doi/10.1103/PhysRevFluids.3.084301.
[20] Schneider E, Stark H. Optimal steering of a smart active particle. Europhys Lett 2019;127(6):64003. http://dx.doi.org/10.1209/0295-5075/127/64003.
[21] Alageshan JK, Verma AK, Bec J, Pandit R. Machine learning strategies for path-planning microswimmers in turbulent flows. Phys Rev E 2020;101:043110. http://dx.doi.org/10.1103/PhysRevE.101.043110, URL https://link.aps.org/doi/10.1103/PhysRevE.101.043110.
[22] Hartl B, Hübl M, Kahl G, Zöttl A. Microswimmers learning chemotaxis with genetic algorithms. Proc Natl Acad Sci USA 2021;118(19).
[23] Tsang ACH, Demir E, Ding Y, Pak OS. Roads to smart artificial microswimmers. Adv Intell Syst 2020;2(8):1900137. http://dx.doi.org/10.1002/aisy.201900137, URL https://onlinelibrary.wiley.com/doi/abs/10.1002/aisy.201900137.
[24] Nasiri M, Löwen H, Liebchen B. Optimal active particle navigation meets machine learning. Europhys Lett 2023;142(1):17001.
[25] Mishra S, van Rees WM, Mahadevan L. Coordinated crawling via reinforcement learning. J R Soc Interface 2020;17(169):20200198.
[26] Elder B, Zou Z, Ghosh S, Silverberg O, Greenwood TE, Demir E, et al. A 3D-printed self-learning three-linked-sphere robot for autonomous confined-space navigation. Adv Intell Syst 2021;3(9):2100039. http://dx.doi.org/10.1002/aisy.202100039.
[27] Nassif X, Bourdoulous S, Eugène E, Couraud P-O. How do extracellular pathogens cross the blood–brain barrier? Trends Microbiol 2002;10(5):227–32.
[28] Celli JP, Turner BS, Afdhal NH, Keates S, Ghiran I, Kelly CP, et al. Helicobacter pylori moves through mucus by reducing mucin viscoelasticity. Proc Natl Acad Sci USA 2009;106(34):14321–6.
[29] Mirbagheri SA, Fu HC. Helicobacter pylori couples motility and diffusion to actively create a heterogeneous complex medium in gastric mucus. Phys Rev Lett 2016;116:198101.
[30] McCloskey M, Cohen NJ. Catastrophic interference in connectionist networks: The sequential learning problem. In: Bower GH, editor. In: Psychology of learning and motivation, Vol. 24, Academic Press; 1989, p. 109–65. http://dx.doi.org/10.1016/S0079-7421(08)60536-8.

[31] French RM. Catastrophic forgetting in connectionist networks. Trends Cognit Sci 1999;3(4):128–35. http://dx.doi.org/10.1016/S1364-6613(99)01294-2, URL https://www.sciencedirect.com/science/article/pii/S1364661399012942.
[32] Webre DJ, Wolanin PM, Stock JB. Bacterial chemotaxis. Curr Biol 2003;13(2):R47–9.
[33] Wolf DM, Fontaine-Bodin L, Bischofs I, Price G, Keasling J, Arkin AP. Memory in microbes: quantifying history-dependent behavior in a bacterium. PLOS one 2008;3(2):e1700.
[34] Skoge M, Yue H, Erickstad M, Bae A, Levine H, Groisman A, et al. Cellular memory in eukaryotic chemotaxis. Proc Natl Acad Sci 2014;111(40):14448–53.
[35] Yang C-Y, Bialecka-Fornal M, Weatherwax C, Larkin JW, Prindle A, Liu J, et al. Encoding membrane-potential-based memory within a microbial community. Cell Syst 2020;10(5):417–23.
[36] da Silva BC, Basso EW, Bazzan ALC, Engel PM. Dealing with non-stationary environments using context detection. In: Proceedings of the 23rd international conference on machine learning. ICML '06, New York, NY, USA: Association for Computing Machinery; 2006, p. 217–24. http://dx.doi.org/10.1145/1143844.1143872.
[37] Padakandla S, J. PK, Bhatnagar S. Reinforcement learning algorithm for non-stationary environments. Appl Intell 2020;50(11):3590–606. http://dx.doi.org/10.1007/s10489-020-01758-5.
[38] Alegre LN, Bazzan ALC, da Silva BC. Minimum-delay adaptation in non-stationary reinforcement learning via online high-confidence change-point detection. In: Proceedings of the 20th international conference on autonomous agents and multiagent systems. AAMAS '21, Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems; 2021, p. 97–105.
[39] Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal policy optimization algorithms. 2017, arXiv:1707.06347.
[40] Dreyfus R, Baudry J, Stone HA. Purcell's "rotator": mechanical rotation at low Reynolds number. Eur Phys J B 2005;47(1):161–4.
[41] Happel J, Brenner H. Low reynolds number hydrodynamics: with special applications to particulate media. Noordhoff International Publishing; 1973.
[42] Kim S, Karrila SJ. Microhydrodynamics: principles and selected applications. Dover, New York; 2005.
[43] Dhont J. An introduction to dynamics of colloids. Elsevier; 1996.
[44] Guo ZV, Mahadevan L. Limbless undulatory propulsion on land. Proc Natl Acad Sci USA 2008;105(9):3179–84.
[45] Hu DL, Nirody J, Scott T, Shelley MJ. The mechanics of slithering locomotion. Proc Natl Acad Sci USA 2009;106(25):10081–5.
[46] Mattingly HH, Kamino K, Machta BB, Emonet T. E. coli chemotaxis is information-limited. 2021, http://dx.doi.org/10.1038/s41567-021-01380-3, bioRxiv.
[47] Golestanian R, Ajdari A. Analytic results for the three-sphere swimmer at low Reynolds number. Phys Rev E 2008;77:036308. http://dx.doi.org/10.1103/PhysRevE.77.036308, URL https://link.aps.org/doi/10.1103/PhysRevE.77.036308.
[48] Howse JR, Jones RAL, Ryan AJ, Gough T, Vafabakhsh R, Golestanian R. Self-motile colloidal particles: From directed propulsion to random walk. Phys Rev Lett 2007;99:048102. http://dx.doi.org/10.1103/PhysRevLett.99.048102, URL https://link.aps.org/doi/10.1103/PhysRevLett.99.048102.
[49] Lobaskin V, Lobaskin D, Kulić IM. Brownian dynamics of a microswimmer. Eur Phys J Spec Top 2008;157:149–56. http://dx.doi.org/10.1140/epjst/e2008-00637-7.
[50] Dunkel J, Zaid IM. Noisy swimming at low Reynolds numbers. Phys Rev E 2009;80:021903. http://dx.doi.org/10.1103/PhysRevE.80.021903, URL https://link.aps.org/doi/10.1103/PhysRevE.80.021903.
[51] Jabbarzadeh M, Hyon Y, Fu HC. Swimming fluctuations of micro-organisms due to heterogeneous microstructure. Phys Rev E 2014;90:043021.
[52] Stark H. Swimming in external fields. Eur Phys J Spec Top 2016;225:2369–87.