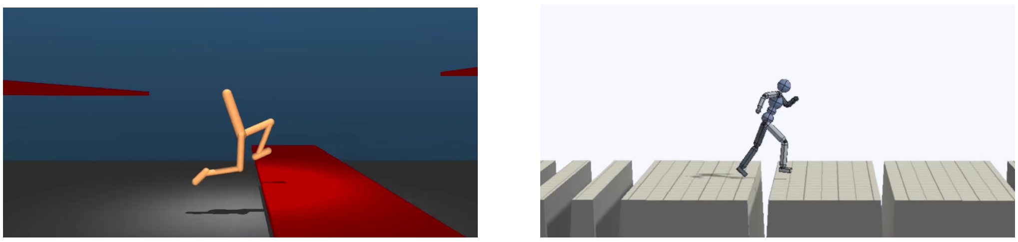
Chapter 1

Introduction

Recent advances in the field of Deep Reinforcement Learning (DRL) have achieved impressive results in various complex tasks, like being able to play Atari games at a super human level (Mnih et al., 2015), and beating the world's Go champion (Silver et al., 2016), just to name a few. These represent discrete action spaces problems, in which the action space can be defined by a finite-size set of values. More recently, this approach has also been applied to continuous actions spaces, i.e. in continuous control tasks. This has given good results in various locomotion benchmarks, like in Heess et al. and Peng et al., shown in Figure 1.1.



(a) (b)

Figure 1.1: Some results from applying Deep Reinforcement Learning to locomotion. a) Biped traversing simulated environment (Heess et al., 2017). b) Humanoid traversing simulated environment (Peng et al., 2018)

This approach is very promising because it allows an agent to learn an appropriate controller for a given task after training in the simulated environment, which reduces the need to implement highly sophisticated control pipelines. Currently available training environments have been presented in various articles, and we will talk more about these in Chapter 3. Some of these training environments are shown in Figure 1.2.

The ultimate goal of applying these techniques is to be able to develop complex behaviours, similar to those shown in nature by animals. However, to develop and test these complex behaviours there is a need to **create a wide set of rich and complex environments**, which is not provided by currently available locomotion benchmarks.

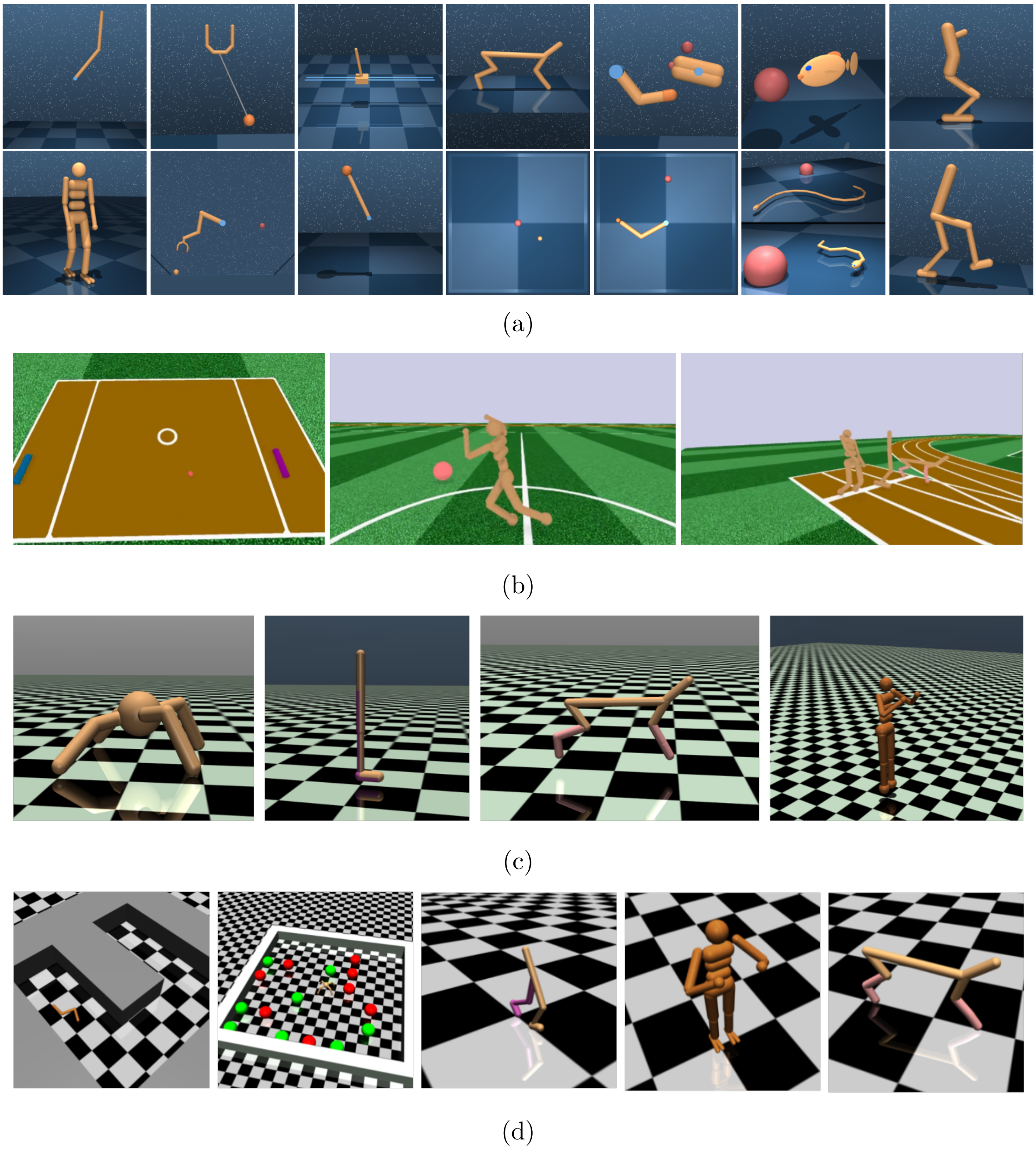


Figure 1.2: Available benchmarks for locomotion tasks: a) controlsuite (Tassa et al., 2018), b) roboschool (OpenAI, 2017), c) gym (Brockman et al., 2016) and d) rllab (Duan et al., 2016)

1.1 Problem Statement

As explained previously, currently there are no robot locomotion benchmarks that allow to **create diverse learning environments**, which limits the range of behaviours that can be learned from scratch using Deep Reinforcement Learning techniques. Besides, there are some other issues that arise in the context of learning locomotion behaviours:

- **Exploitation** of the dynamics of the simulated environments: Because of the nature of the objective the agents try to optimize for, they usually end up developing behaviours that cheat by exploiting some specific aspects of the environment and getting stuck sub-optimal behaviours, e.g. walking in very unnatural way in walking tasks. This is specially true when designing a reward function. The reward is the signal that the agent gets to improve its behaviours, so poorly engineered reward functions could lead to sub-optimal and even bad performance.

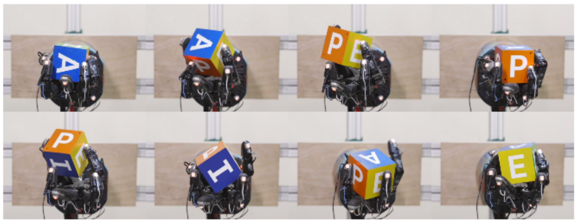
- **Inability to transfer** learned policies into the real world: In general, policies learned in simulation do not transfer directly to a real world robotics platform. This is because of discrepancies between the dynamics of the simulation and the dynamics of the real world, which is usually called the **reality gap**. This issue is also a direct consequence of the previous one. As the agent exploits the dynamics of the simulation, it will overfit to the given environment, and will not generalize when transferred to the real world.

Some approaches used to deal with these issues are related to controlling the learning environments itself, like in Tan et al. and Andrychowicz et al. (Figure 1.3). These try to account for unmodeled dynamics and generalization by using randomization over the simulated environment and other techniques to force the learning agent not to overfit to a specific behaviour that only exploits some of the dynamics of the environment.

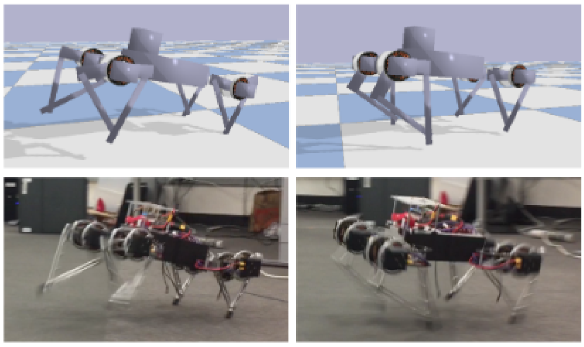
Then, one approach that can be used to solve these issues is to have a framework that provides researchers with the tools required to create these diverse environments and have full control over them. Furthermore, such a framework could be used to develop environments in a progressive way via a **curriculum** (Bengio et al., 2009), Based on all this, we formulate the following hypothesis:

Hypothesis 1 *By having control of the learning environment we can control the learning procedure of an agent in such a way that it can develop complex and robust behaviours to solve the tasks presented, and also generalize to other environments.*

We will focus in testing half of this hypothesis by implementing the required framework to create these diverse environments, and evaluating current state of the art Deep RL algorithms to these new environments. As future work we will try to deal with the full hypothesis, by trying to develop appropriate curricula.



(a)



(b)

Figure 1.3: Some examples of sim2real results: a) Robotic hand generalizing in randomized environment (Andrychowicz et al., 2018). b) Quadruped trained in simulation and deployed in real world (Tan et al., 2018)

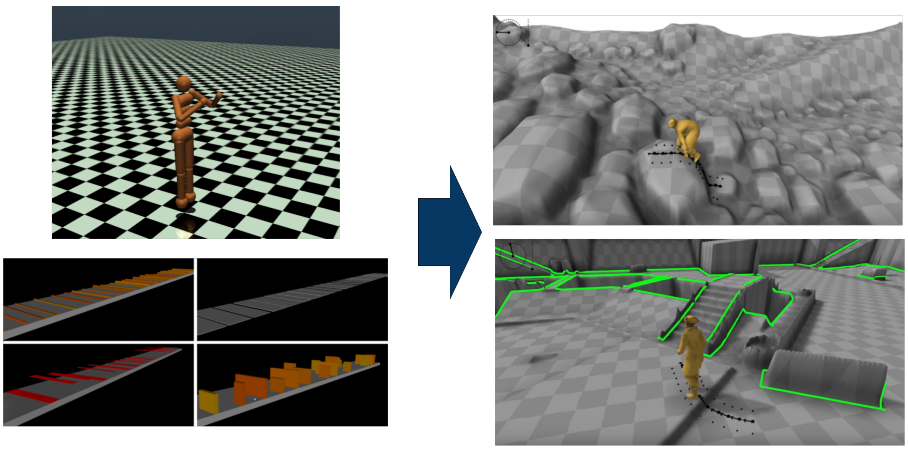


Figure 1.4: Comparison of the proposed environments to the environments provided by current benchmarks. Left: relatively simple environments (image adapted from Brockman et al. (2016)). Right: relatively more complex proposed environments (images adapted from Holden et al. (2017)).

1.2 Objectives

General Objective

Develop a framework that allows the creation of diverse learning environments for locomotion tasks (as shown in Figure 1.4), and evaluate the performance and generalization capabilities of current state of the art Deep Reinforcement Learning algorithms.

redacta mej el objetivo, make y buil tienen signi cados simi- lares, tal ve sea mejor d velop a fran work that allow us to build...

Specific Objectives

- Implement an abstract core framework for robot locomotion, decoupled from any concrete physics engine currently available.
- Implement wrapper functionality for the integration of various currently available physics engines following a common API.
- Implement a set of APIs that allows to create diverse and complex environments, and create diverse tasks using these environments.
- Implement user APIs via Python bindings to allow users to easily interact with the framework, set up experiments and train agents using current Deep Learning Frameworks.
- Make the necessary documentation for the whole framework to allow further development of this framework and adoption as a standardized benchmark for robot locomotion.
- Evaluate the performance of state of the art Deep Reinforcement Learning algorithms in these environments.

solo hay do objetivos es pecíficos?

1.3 Organization

The structure of this proposal is as follows :

- **Chapter 2 - Related background**: we present the necessary background and terminology used throught this document.
- **Chapter 3 - Related works**: we discuss the currently available benchmarks by analyzing the provided functionality and use cases. We also discuss some state of the art algorithms used with these benchmarks, which will be implemented and evaluated with the proposed framework.

- **Chapter 4 - Proposal in depth**: we discuss more details of the proposed framework, including: architecture of the framework, core features to be implemented, etc.
- **Chapter 5 - Current progress**: we discuss the current state of the implementation of the proposed framework.
- **Chapter 6 - Discussion**: finally, in this chapter we present a discussion on the whole proposal and the road-map that we are following towards the objectives proposed earlier.

a bueno
capítulo de
clusiones