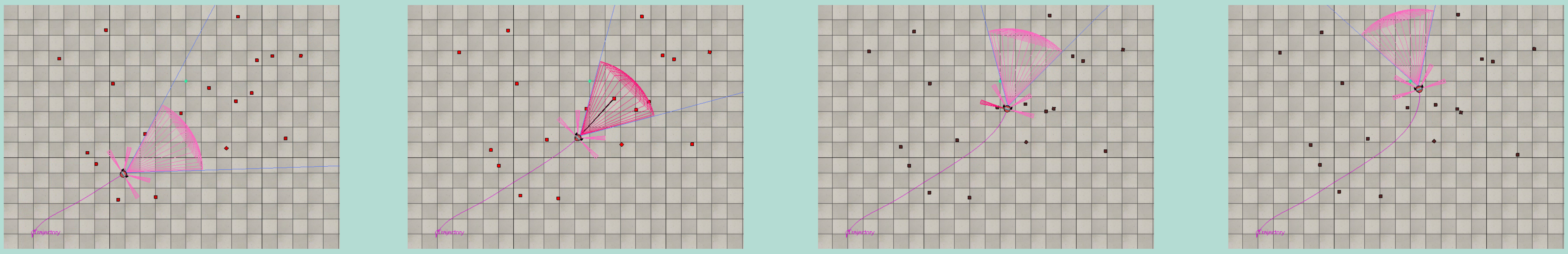
DEEP REINFORCEMENT LEARNING FOR ON-LINE MOTION PLANNING

Leonid Butyrev[1], Georgios Kontes[1], Thorsten Edelhäußer[1] and Christopher Mutschler[1,2]

[1] Fraunhofer IIS, Machine Learning and Information Fusion Group, Nuremberg, Germany
[2] Friedrich-Alexander University Erlangen-Nürnberg (FAU), Machine Learning and Data Analytics Lab
{butyreld | georgios.kontes | thorsten.edelhaeusser | christopher.mutschler} @iis.fraunhofer.de



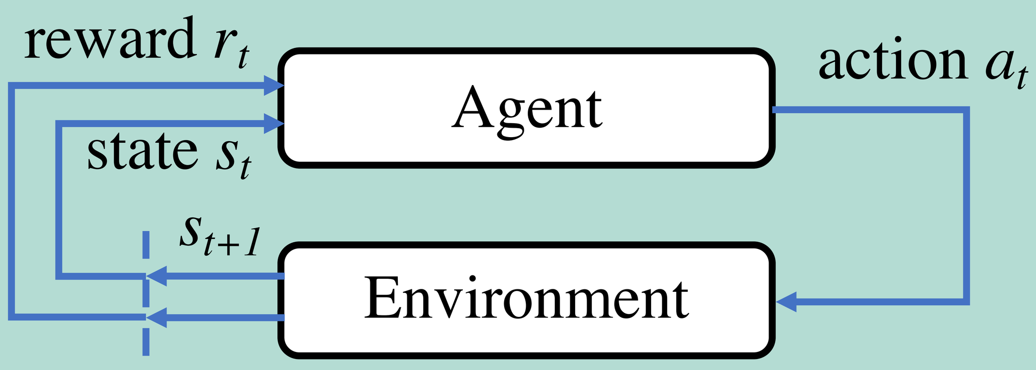
A two-wheeled balancing robot discovers a collision-free path in a dynamic environment with moving obstacles.

MOTIVATION

Trajectory (or Motion) Planning is a fundamental and well-studied problem in mobile robotics. Classical path/motion planning and following methods usually are not suitable for on-line settings, while they also require a model of the robot and the environment.

OUR APPROACH

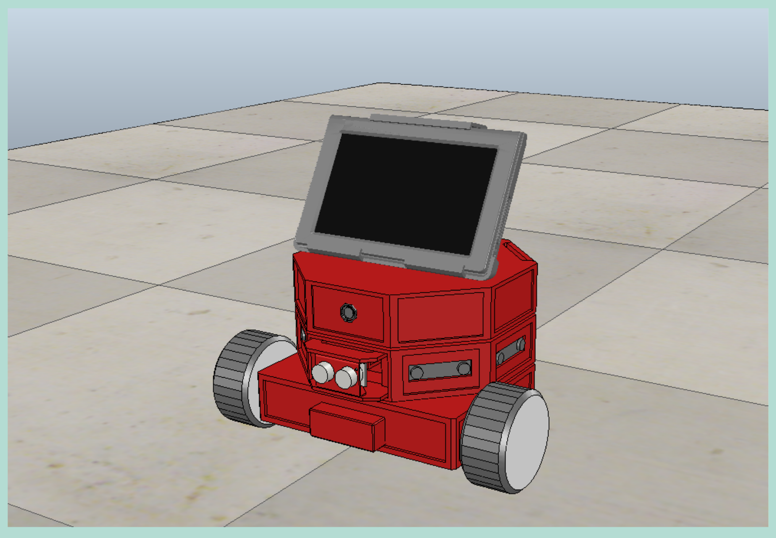
We leverage recent advances in (Deep) Reinforcement Learning (RL) to address the path planning and following problem in dynamic environments. In model-free RL a model of the system is not required but the agent learns to perform a predefined task by trial-and-error through interacting with the environment. In our case, we use the Deep Deterministic Policy Gradient (DDPG) algorithm for the robot/agent training.



The Reinforcement Learning paradigm.

THE SIMULATION ENVIRONMENT

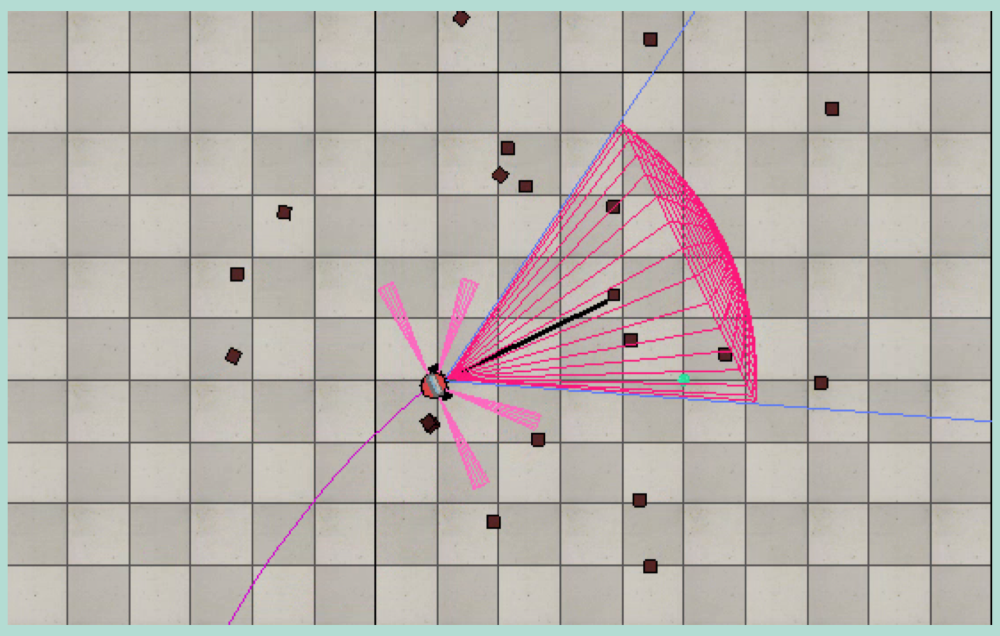
We simulate a Segway-like two-wheeled *balancing* robot and a dynamic environment with moving obstacles in V-REP. The robot is equipped with 5 proximity sensors with different ranges. One of the main constraints of the sensor infrastructure is that each sensor can detect only one obstacle (the closest one), which makes the problem significantly more challenging for our training algorithm.



The simulated two-wheeled robot.

EVALUATION AND RESULTS

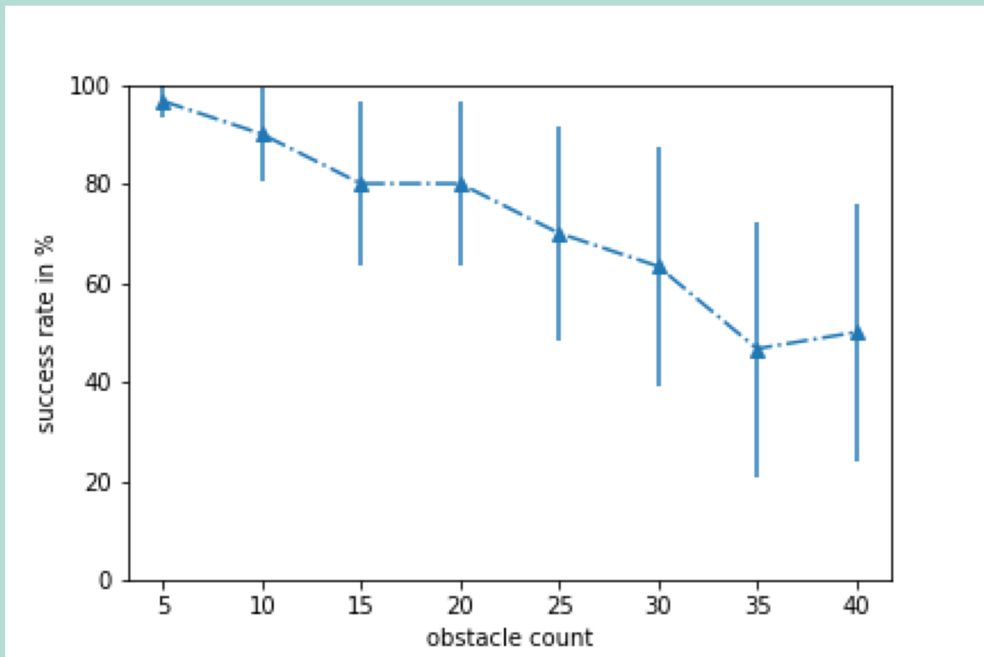
We train the agent in an environment with 30 random moving obstacles. The controller/policy utilizes the sensor inputs to directly generate speed commands for the two wheels, while the only global information is the goal position.



The simulated dynamic environment.

We evaluate the learned policy in different dynamic environments with varying numbers of obstacles.

The robot achieves high success rates in settings with low number of obstacles as expected. For environments with 20 – 30 moving obstacles has a slight drop of performance due to the limited sensor system. For settings with 35 and 40 obstacles the performance drops even further both due to the high number of obstacles and the fact that the agent was trained with 30 obstacles and has not encountered higher obstacle densities during training.



Evaluation of the task with varying number of obstacles.

Even though the limited sensor capabilities hindered the robot's task, the agent was able to achieve high success rates.

Fraunhofer IIS