Preparation of Papers for IEEE Sponsored Conferences & Symposia*

Florian Kuhnt[1] and J. Marius Zöllner[1]

Abstract— This electronic document is a ÒliveÓ template. The various components of your paper [title, text, heads, etc.] are already defined on the style sheet, as illustrated by the portions given in this document.

I. Introduction

The task of teaching vehicles how to drive autonomously in urban scenarios is a challenging and complex one to solve. Not only is there the problem of finding the adequate response to a given situation but also the challenge of taking into account the surrounding factors that have an influence on the state that a vehicle is in and its possible actions. To date, most approaches focus on the manual design of behavioral policies, such as defining a driving policy through the use of annotated maps. While these solutions might work in situations which are documented by the provided mapping infrastructure, they are often difficult to generalize or scale, as they do not necessarily enable the comprehension of any given local scene. In order to make autonomous driving truly feasible in a real-world scenario it would be better to develop systems which are able to find their way without having to rely on an explicit set of rules. One possible solution to this task is provided by reinforcement learning methods. Here, the agent, i.e. the vehicle, actively searches for the optimal driving policy whilst trying to maximize a numerical reward signal. As opposed to imitation learning techniques, which have been popular in finding driving policies (1), reinforment learning algorithms enable a car to exceed human abilities, if applied correctly. In recent years, deep reinforcement learning methods have proven to be succesful in solving complex tasks such as playing GO (2) or Atari (3) and there have been efforts in tackling various problems in the field of autonomous driving, including continous control tasks (4).

However, two major drawbacks of reinforcement learning methods are their heavy depency on adequate input state representations (5) and, as with other machine learning techniques, their need of a sufficient amount of accurate sample data to train on. In order to be able to train safely on an adequate amount of data, one approach is the use of data from other domains. For this purpose, the urban driving simulator CARLA has been developed, which is used as a simulation environment for this project.

[1]The authors are with FZI Research Center for Information Technology, Haid-und-Neu-Str. 10-14, 76131 Karlsruhe, Germany {kuhnt, zoellner}@fzi.de

In this paper, several state-of-the-art reinforcement learning algorithms are implemented and compared, with regard to their performance considering different driving tasks. Additionally, reward functions for the respective problems are tested and several input represantations are designed and evaluated.

?? What is it exactly that we contributed that is new to already existing research??

II. Related Work

III. Background

In this project, three differnt algorithms are used. Following the work of (Lillicrap et al.) and (Mnih et al.), the (DDPG)- and (A3C) algorithms are implemented and compared according to their performance in the Gym environment CarRacing-v0*. Later on, the PPO algorithm is applied to solve a continuous control task in CARLA.

A. A3C

B. DDPG

One very notable advance in reinforcement learning has been made by the development of the so-called "Deep Q Network" (Mnih et al., 2015). The DQN is able to solve tasks with high-dimensional observation spaces. However, it is only efficiently capable of working with discrete and low-dimensional action spaces. In order to adapt a (DQN) for the successful use with continuous control problems, as given in CarRacing-v0, a discretization of the action space has to be carried out, which can lead to two main difficulties: an explosion in the number of possible actions and the loss of important information (4).

To evade these obstacles (Lillicrap et al.) propose a new approach, called the Deep Deterministic Policy Gradient (DDPG), which is a model-free, off-policy actor-critic algorithm. They adopt the advantages of (DQN) and combine them with the actor-critic framework, resulting in the stabilization of Q-learning by using a replay buffer and soft updates on the target networks of both actor and critic, through

$$\tau << 1 : \theta' \leftarrow \tau\theta + (1-\tau)\theta' \tag{1}$$

a) Model architecture:

b) Training details:

C. CARLA

IV. Concept

V. Evaluation

A. Results

B. Comparison algorithms/reward functions

VI. Conclusions

References

[1] A. Barth and U. Franke, "Where will the oncoming vehicle be the next second?" 2008 IEEE Intelligent Vehicles Symposium, pp. 1068–1073, June 2008. [Online]. Available: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4621210