Yunpeng WANG, Kunxian ZHENG, Daxin TIAN, Xuting DUAN, Jianshan ZHOU, 2021. Pre-training with asynchronous supervised learning for reinforcement learning based autonomous driving. *Frontiers of Information Technology & Electronic Engineering*, 22(5):673-686. https://doi.org/10.1631/FITEE.1900637

Pre-training with asynchronous supervised learning for reinforcement learning based autonomous driving

Key words: Self-driving; Autonomous vehicles; Reinforcement learning; Supervised learning

Corresponding author: Kunxian ZHENG
E-mail: zhengkunxian@buaa.edu.cn
ORCID: https://orcid.org/0000-0002-2887-9294

Motivation

- Although the existing reinforcement learning (RL) theory has achieved some success in a variety of domains, its applicability has previously been focused on gaming or robotic control domains or on other domains in which poor initial performance can be tolerated. Few efforts have been made to explore the applicability of SL theory in promoting the initial performance of RL-based models for practical applications in real-world settings. We conduct some innovative work to demonstrate the effectiveness and the potential of the joint training framework of SL and RL in the design of the pre-training method, for enhancement of the initial performance of the RL-based model.

Motivation

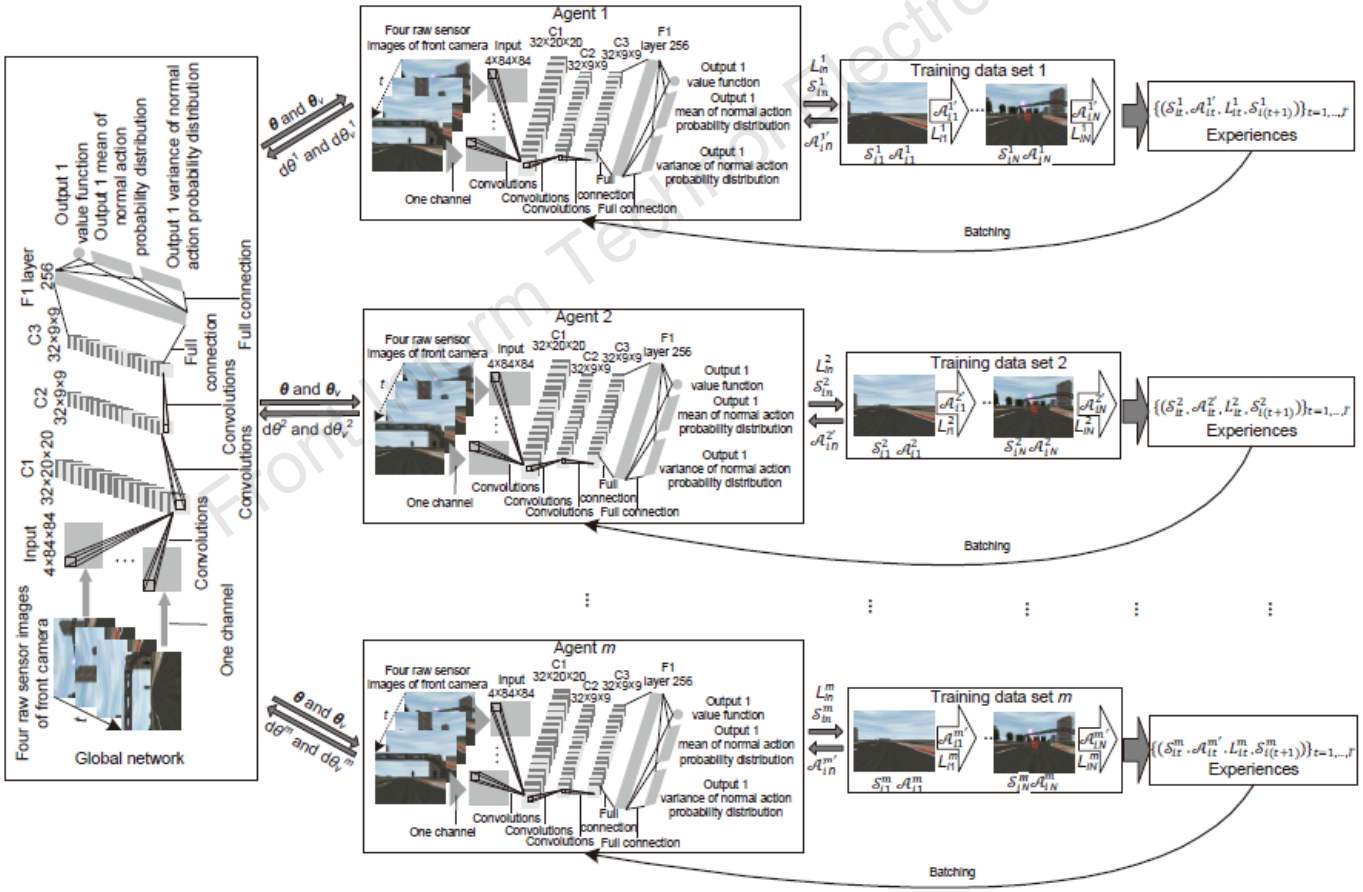
- Moreover, while our study is based on the existing RL theory, the methodological framework proposed is not a direct application of existing RL algorithms. To facilitate the practical application of the RL-based autonomous driving model in real-world settings, we propose a novel pre-training approach, in which two new components are designed and combined to adapt to the environments and improve the initial decision-making performance of RL agents in autonomous driving problems: (1) an asynchronous supervised learning (ASL) method based on the joint methodological framework of RL and SL for the pre-training stage, and (2) a manually designed heuristic driving policy (MDHDP) for automatic collection of pre-training demonstration data.

Main idea

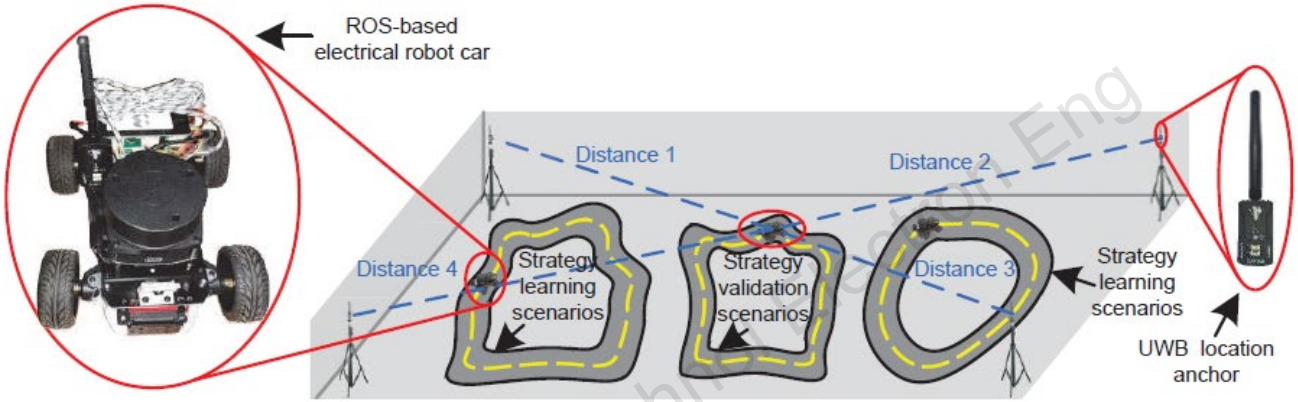
- We propose an ASL method for pre-training the end-to-end autonomous driving model, similar to the Gorila framework (Nair et al., 2015) and the asynchronous RL framework (Mnih et al., 2016); however, instead of executing multiple RL processes, we asynchronously execute multiple SL processes in parallel, on multiple real-world training data sets. The aim in designing this asynchronous method is to propose an SL algorithm that can learn a sequential decision-making policy reliably and without large resource requirements. By running different SL processes in different threads, the overall changes being made to the model parameters by multiple agents applying online updates in parallel are likely to be less correlated in time than a single agent applying online updates, which stabilizes the SL process. After pre-training by the ASL method, the autonomous driving model is trained in real-world settings by RL.

Method

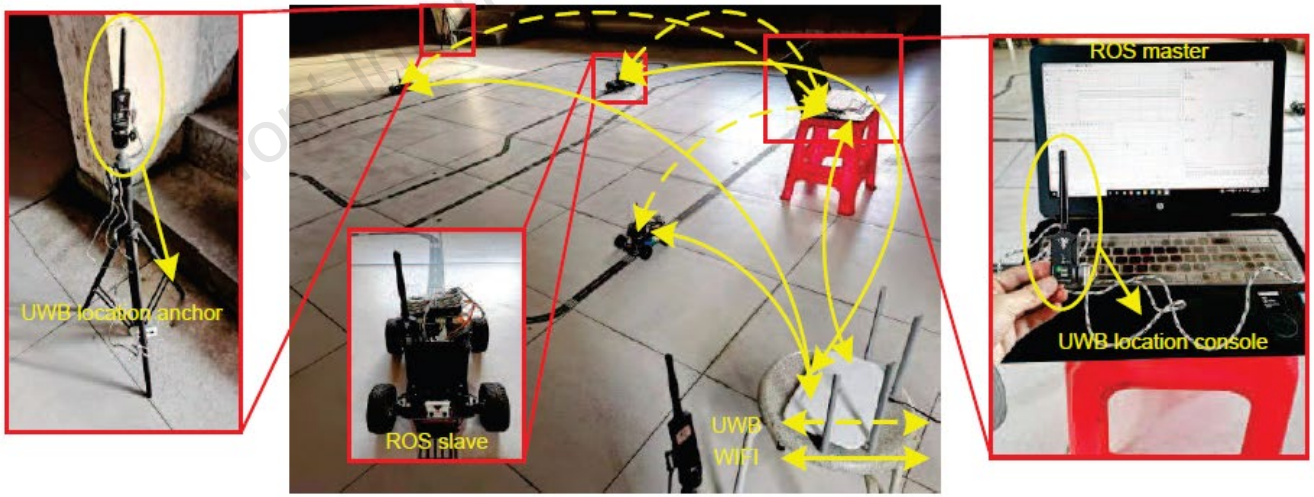
1. Methodological framework of ASL: the presented pre-training method asynchronously executes multiple SL processes in parallel on multiple driving demonstration data sets



2. The real-vehicle verification system: (a) the schematic diagram of real-vehicle scenarios; (b) the physical picture of real-vehicle scenarios



(a)



(b)

Major results

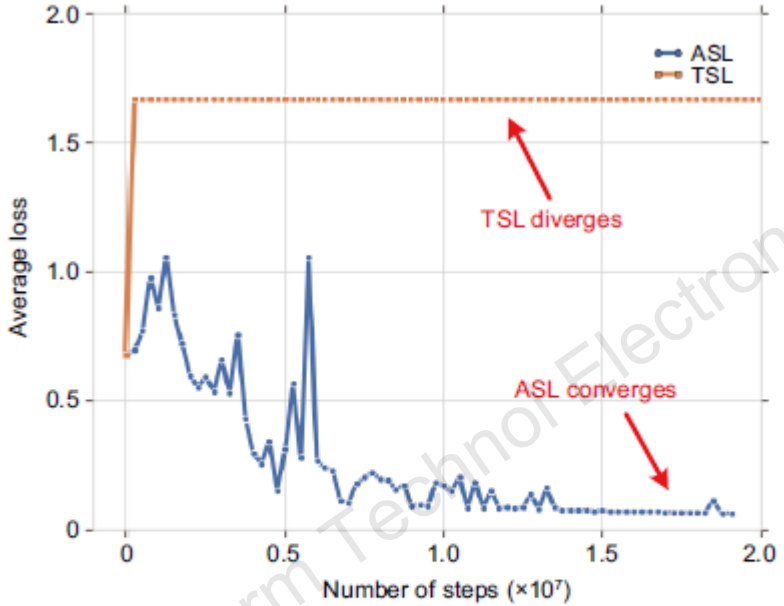


Fig. 5 Convergence performance in the pre-training stage, showing the superiority of our proposed ASL method compared with another SL method in convergence

For the ASL method, it can be seen that the loss can converge to a steady state (about 0.1) after about 1×10^7 steps, which proves that our proposed ASL method can make the RL-based model effectively learn the prior knowledge from the training data set. For the TSL method, it remains in a high loss state. TSL cannot adapt to the high-dimensional input or continuous output of the autonomous driving scenario, and therefore loses the ability to converge. This verifies the superiority of our proposed ASL method in imitation learning of autonomous driving model's pre-training tasks.

Major results (Cont'd)

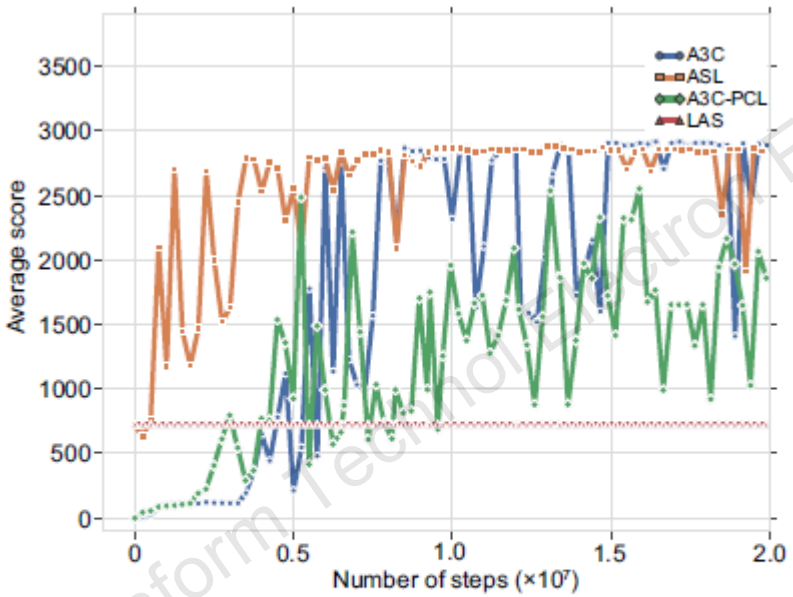


Fig. 6 Simulation results for different methods in the RL training stage, showing the reward against the total RL training step (References to color refer to the online version of this figure)

It can be seen that the autonomous driving model pre-trained by our ASL method with 1.0×10^7 pre-training steps converges faster in the RL training process than the A3C and A3C-PCL methods.

Conclusions

- We have proposed an ASL method to improve the convergence speed of the autonomous driving model in an RL training process, and our trained model can achieve better performance than the typical autonomous driving methods.
- We have found a feasible and effective visualization method to analyze the improvement due to pre-training from the detailed point of view, which uses heatmaps to visualize the specific influences of the input units in the neural network. The visualization results are meaningful and explicable, which helps us determine if the pre-trained model has learned prior knowledge.
- Simulation and experimental results have demonstrated the feasibility, effectiveness, and superiority of our method from both the microscopic and macroscopic points of view.

王云鹏，博士，教授，主要从事车辆运行状态联网感知、行车安全预警和车路协同控制研究

段续庭，博士，讲师，主要从事车联网系统、V2X通信、协作定位研究

郑坤贤，硕士，主要从事车载自组网信道分配、自动驾驶系统研究

周建山，博士，主要从事车联网系统、群体智能、计算卸载研究

田大新，博士，教授，主要从事车联网系统、群体智能、智能交通系统研究