Manav Gagvani
September 15, 2023

Senior Research Proposal: Using a Variational Autoencoder to Split Reinforcement Learning into Online and Offline Stages

1 Introduction and Objective

Self-driving cars are currently a hot topic in robotics, AI, and computer vision research. With some of the world's largest companies devoting billions of dollars towards building the world's first self-driving car, there have also been several projects and organizations devoted to creating research platforms for autonomy which are smaller, cheaper, and safer than real cars.

I aim to research reinforcement learning for self-driving cars, with the focus being improving the sample efficiency of the algorithm so that the learning process can occur in real life instead of in simulation. My goal is to do this through a two-stage learning process in which a basic policy is learned through imitation learning of a human operator before it is fine-tuned through reinforcement learning in the real world.

2 Prior Work

2.1 My Prior Work

My prior work has focused mostly on simpler approaches to autonomy such as end-to-end control. In 2016, Nvidia published a breakthrough paper in which they used a convolutional neural network to make a real car drive around simple highways [Bojarski et al., 2016]. The innovation behind their idea was the fact that the problem of autonomy, which is usually split up into perception, planning, and control, was solved all at once, hence the "end-to-end" moniker.

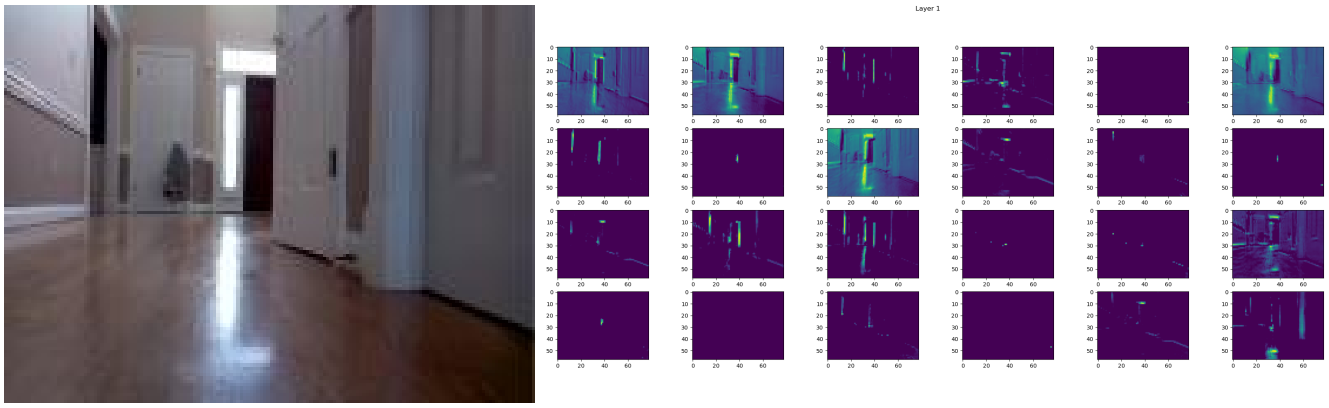

Figure 1: **Left:** A picture taken from the onboard camera of my 1/18 scale car. **Right:** The output from the first layer of a CNN I trained driving in loops around the hallway to the left. By visualizing a single layer we see how the model learns to steer in the absence of clear visual indicators such as a line on the ground.

I applied this idea to my own 1/18 scale car (the ones in the Sys-Lab are 1/10 scale, and I will talk more about those later). The CNN takes in an image from the front-facing camera as input and passes it through 4 convolutional layers before flattening the output and passing it through a fully-connected network until the two outputs, steering angle and throttle, are obtained. Having designed and assembled the hardware myself, this project helped me become familiar with the systems onboard these autonomous R/C cars as well as the software stack used, for instance, the way a remote computer interacts with the SBC on the car itself, and how the SBC sends the motor driver commands over the I2C bus.

2.2 The F1Tenth Project

The F1Tenth project has formed the base of the research I plan to do. This project, funded by an NSF grant since 2017, has created open source hardware platforms and corresponding simulators for the hardware for use in reinforcement learning and other research [O'Kelly et al., 2020].

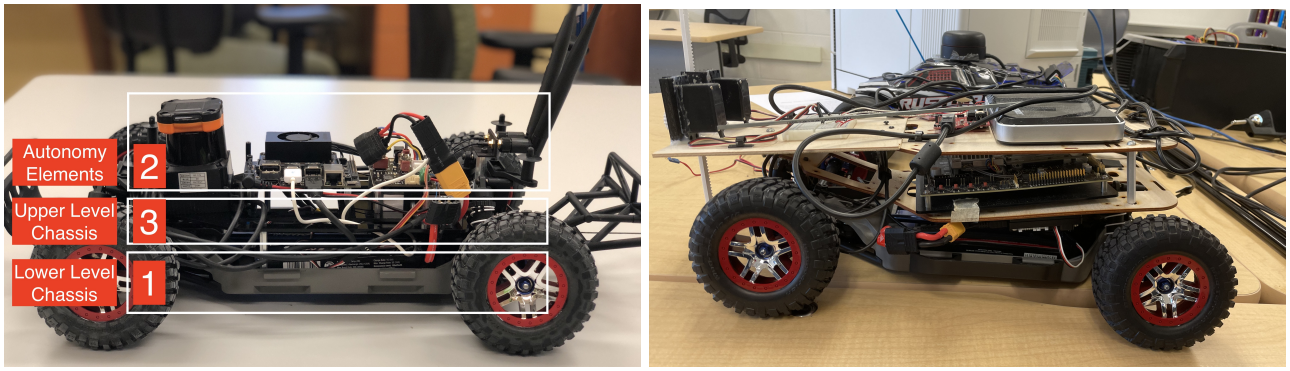

Figure 2: **Left:** A fully completed F1Tenth car with Lidar and stereo camera sensors. **Right:** The current state of the car in the lab. It is missing a stereo camera and Lidar sensor but I will have to disassemble it to determine what else is needed to get it up and running.

Several research projects involving using a learned strategy to navigate have previously been completed, such as [Zheng et al., 2022] which structures multi-agent driving as a game tree, or [Bosello et al., 2022] which uses a Deep Q-Network to train a car to drive on a simple track in real life without a corresponding physics simulators. However, this research uses a Lidar scan, which is a 1-D array of distances, as the input to the algorithm, constraining the usefulness of this approach.

2.3 Reinforcement Learning with Vision

I aim to use reinforcement learning with the input being a front-facing picture from the view of the car. This has the potential to be far more generalizable than using Lidar as previous approaches have done. Although there are several variations of this approach, I plan to use some form of encoder-decoder network to speed up the real-life training process. In a physics simulation, it often takes thousands of runs to train a policy which can drive without crashing. This would be far too slow for real life, so the initial exploration stages of reinforcement learning need to be sped up. The paper "Learning to Drive in a Day" [Kendall et al., 2019] uses a deep deterministic policy gradient as the reinforcement learning algorithm. It uses a variational autoencoder (VAE) which is trained *online* during the reinforcement learning process. However, the key benefit of using an autoencoder is the fact that it does not have to be trained online, and can instead be trained in two parts. The encoder converts an input into a set of latent features, also known as the latent space. Latent features are feature which the model has learned are important for determining the proper output. To train the encoder, a decoder is also trained with the inputs being the latent space and the outputs being a reconstruction of the encoder's input. Dreamer [Hafner et al., 2023] is a recently developed approach to reinforcement learning employing a latent space for increased explainability and sample-efficiency of RL algorithms which has been applied to autonomous cars in [Brunnbauer et al., 2022]. Manually observing the latent space of these algorithms can provide insight into neural networks which are normally "black boxes". In simulation, [Raffin and Sokolkov, 2019] decoupled the VAE training process and the reinforcement learning process. Given a pre-trained VAE network, a reinforcement learning algorithm can be trained in simulation in minutes rather than days. I plan on training the algorithm in real life instead, this could potential take a few hours or days but is still much more tractable.

3 Research Plan

1. 1^{st} Quarter: My goals for the first quarter are to get all the infrastructure and hardware needed for the project to work. Once I finish my proposal, I plan on replicating other researcher's previous work on VAEs in simulation to ensure they are working and valid. I also plan on fully inventorying the current cars to determine which parts I will need to make the system work. As of right now, I think I will need a high-quality

stereo depth camera (something like this camera) since this project is so reliant on **1)** computer vision, and **2)**, spatial information, which stereo depth with structured light can provide. By the end of the 1st quarter I aim to have fully operational car(s) to test my algorithm on as well as an assurance that the idea of using a variational autoencoder is sound.

2. 2^{nd} Quarter: During the second quarter I plan on starting to implement basic control algorithms as baselines to compare against the reinforcement learning approach. These are mostly based on kinematic models and are very well-documented. Once that is done, I will start implementing the two-stage training process, manually collecting data to train the encoder.

3. 3^{rd} Quarter: I will continue working on implementing reinforcement learning with the autoencoder. During the third quarter, I foresee spending a significant chunk of time doing live reinforcement learning training with the cars. This will take a lot of time since I will be performing a process normally done in simulation in real life; while the pre-trained encoder will speed up the learning process, it will still be at least an order of magnitude slower than simulating.

4. 4^{th} Quarter: The fourth quarter is dedicated to creating a poster presentation, slide deck, and research paper. I also plan on creating a polished, repeatable live demo of a car driving using my reinforcement learning algorithm.

4 Works Cited

References

[Bojarski et al., 2016] Bojarski, M., Testa, D. D., Dworakowski, D., Firner, B., Flepp, B., Goyal, P., Jackel, L. D., Monfort, M., Muller, U., Zhang, J., Zhang, X., Zhao, J., and Zieba, K. (2016). End to end learning for self-driving cars. *CoRR*, abs/1604.07316.

[Bosello et al., 2022] Bosello, M., Tse, R., and Pau, G. (2022). Train in austria, race in montecarlo: Generalized rl for cross-track f1¡sup¿tenth¡/sup¿ lidar-based races. In *2022 IEEE 19th Annual Consumer Communications Networking Conference (CCNC)*, pages 290–298.

[Brunnbauer et al., 2022] Brunnbauer, A., Berducci, L., Brandstätter, A., Lechner, M., Hasani, R., Rus, D., and Grosu, R. (2022). Latent imagination facilitates zero-shot transfer in autonomous racing.

[Hafner et al., 2023] Hafner, D., Pasukonis, J., Ba, J., and Lillicrap, T. (2023). Mastering diverse domains through world models. *arXiv preprint arXiv:2301.04104*.

[Kendall et al., 2019] Kendall, A., Hawke, J., Janz, D., Mazur, P., Reda, D., Allen, J.-M., Lam, V.-D., Bewley, A., and Shah, A. (2019). Learning to drive in a day. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8248–8254. IEEE.

[O'Kelly et al., 2020] O'Kelly, M., Zheng, H., Karthik, D., and Mangharam, R. (2020). F1tenth: An open-source evaluation environment for continuous control and reinforcement learning. *Proceedings of Machine Learning Research*, 123.

[Raffin and Sokolkov, 2019] Raffin, A. and Sokolkov, R. (2019). Learning to drive smoothly in minutes. https://github.com/araffin/learning-to-drive-in-5-minutes/.

[Zheng et al., 2022] Zheng, H., Zhuang, Z., Betz, J., and Mangharam, R. (2022). Game-theoretic objective space planning. *arXiv preprint arXiv:2209.07758*.