End-to-End Autonomous Driving using Proximal Policy Optimization

Sai Harshitha Peddi (018222087), Venkata Satya Charan Naga Deep (019127238), Niharika Janapati (017553744), Bhargava Phaneendra Kumar Suryadevara (018176379)

Artificial Intelligence has advanced rapidly, and autonomous driving stands as one of its most ambitious applications. Self-driving cars promise safer roads and improved mobility, but training such systems directly in the real world is risky, expensive, and ethically complex. Driving is a continuous decision-making task where an agent must interpret high-dimensional sensory input and output precise control signals. Mistakes can cascade quickly, making safe training environments essential. To address these challenges, simulation and reinforcement learning provide a framework for building and evaluating driving agents before deployment.

This project investigates how Proximal Policy Optimization (PPO), a stable on-policy reinforcement learning algorithm, can be applied to train an agent to navigate urban environments. We will implement PPO from scratch, incorporating clipped objectives, value functions, generalized advantage estimation, and entropy regularization. The agent will learn from raw camera images or latent representations, gradually improving through interaction. Reward shaping will encourage route completion and smooth lane-following while penalizing collisions, traffic violations, and unsafe maneuvers. To improve robustness, we will explore curriculum learning, beginning with simple driving tasks and progressing to complex scenarios, as well as domain randomization across weather, lighting, and traffic conditions.

While PPO relies mainly on online interaction with simulators, we may also use publicly available CARLA-related datasets and driving image collections to pretrain encoders or benchmark results. The final goal is to develop a policy that follows routes, stays in lane, and minimizes infractions, providing insights into reinforcement learning for autonomous driving. Development will begin with open-source software on local systems, but due to the computational demands of pixel-based reinforcement learning, we request access to the university's high-performance computing cluster for GPU-accelerated training.

Each member will contribute to complementary aspects of the project. Charan will manage simulation setup and integration, Bhargava will implement PPO, Niharika will design rewards, handle curriculum learning, and optimize PPO, and Harshitha will run evaluations, oversee baselines, and coordinate documentation. Our work will be guided by recent literature, including [1] and [2].

References

[1] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal Policy Optimization Algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[2] A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, "CARLA: An Open Urban Driving Simulator," *arXiv preprint arXiv:1711.03938*, 2017.