Applying Deep Reinforcement Learning Methods in VISTA

Oliver Chang (elochang@ucsc.edu), Leilani Gilpin (lgilpin@ucsc.edu)

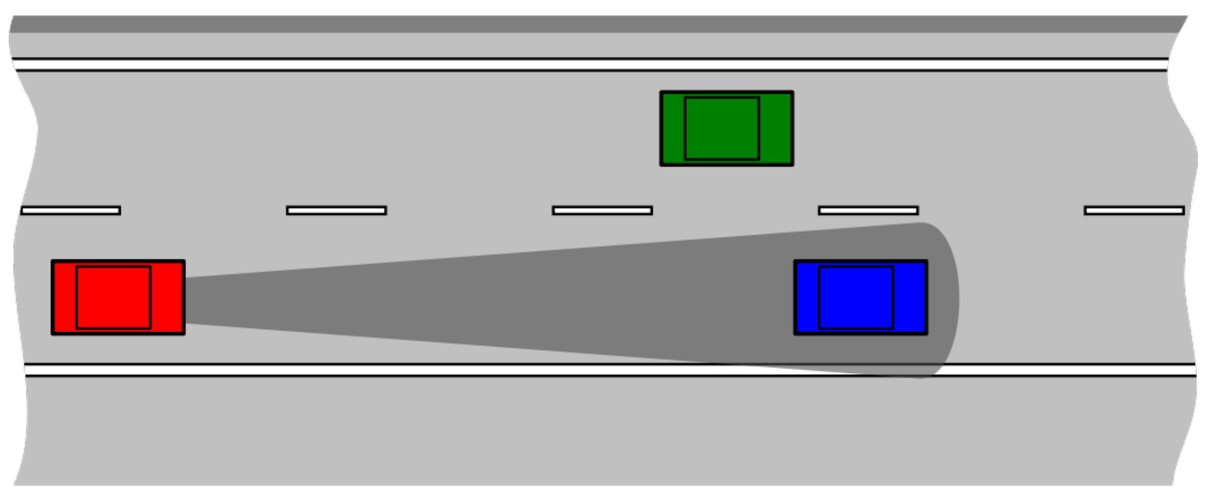
DRL in AVs via Simulator

Motivation

- Lack of well-establish DRL benchmark in AVs
- VISTA is an open-source and computationally cheap software
- Photorealistic data augmentation → robust Sim-to-Real transferability

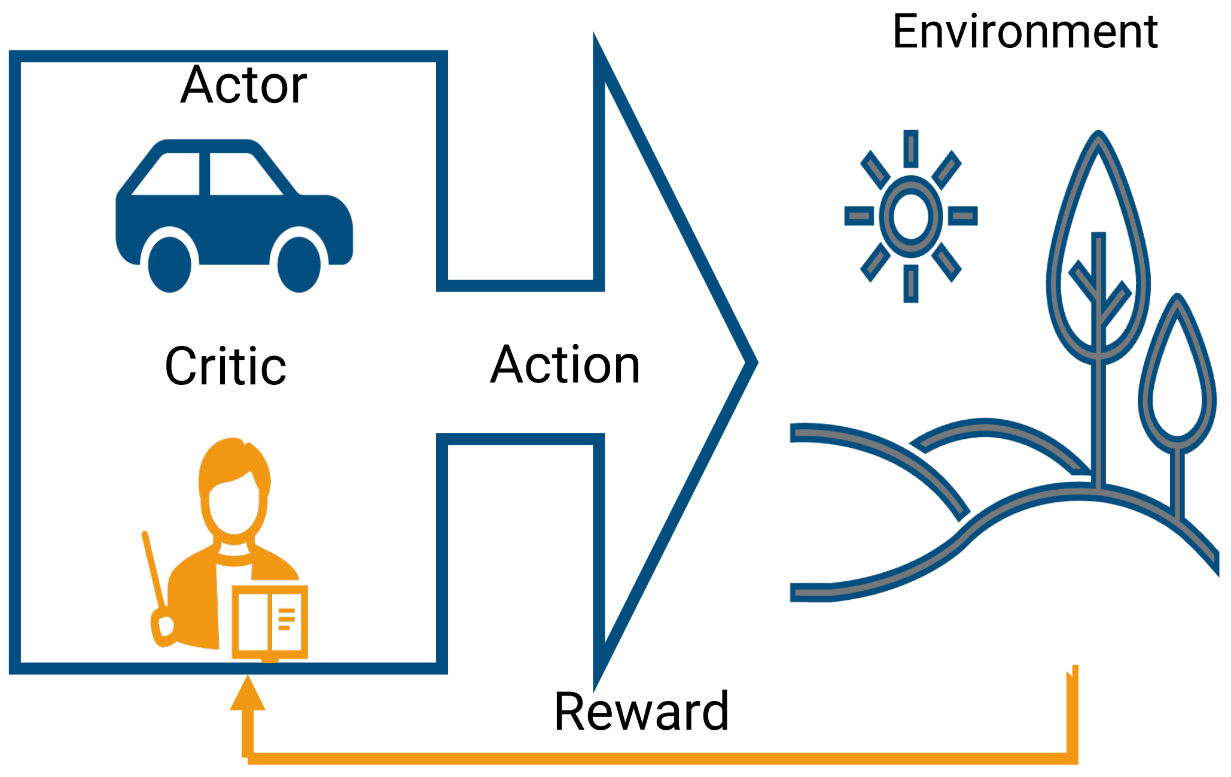
Goal

- Lane following
- Collision Avoidance
- Measure various state-of-the-art DRL approaches



DRL Algorithms

A2C



PPO

- Uses clipped surrogate objective optimization

$$L^{CLIP}(\theta) = E[\min(r_t(\theta)A_t, clip(r_t(\theta), 1-\epsilon, 1+\epsilon)A_t)]$$

where $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta old}(a_t|s_t)}$

SAC

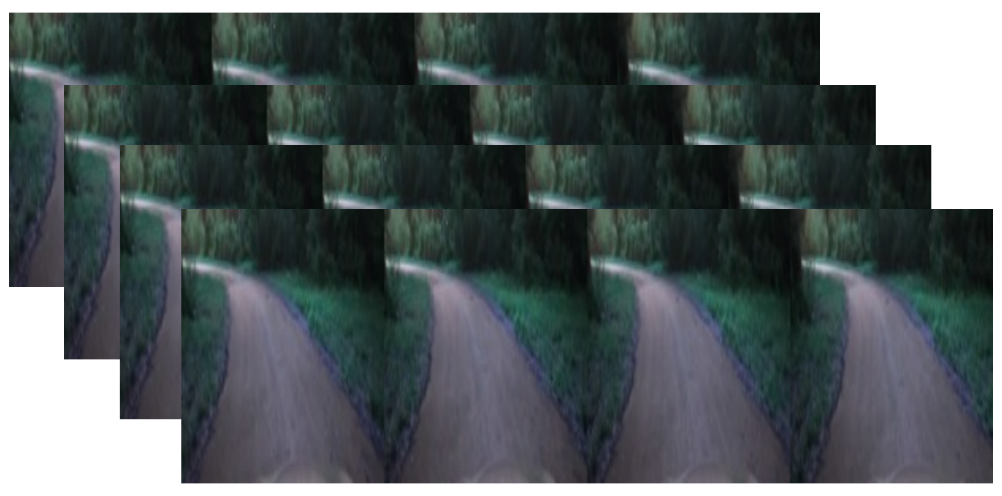
- Off-policy DRL approach
- Utilizes a replay buffer for efficient sampling
- Features exploration through maximum entropy
- Works with continuous action spaces

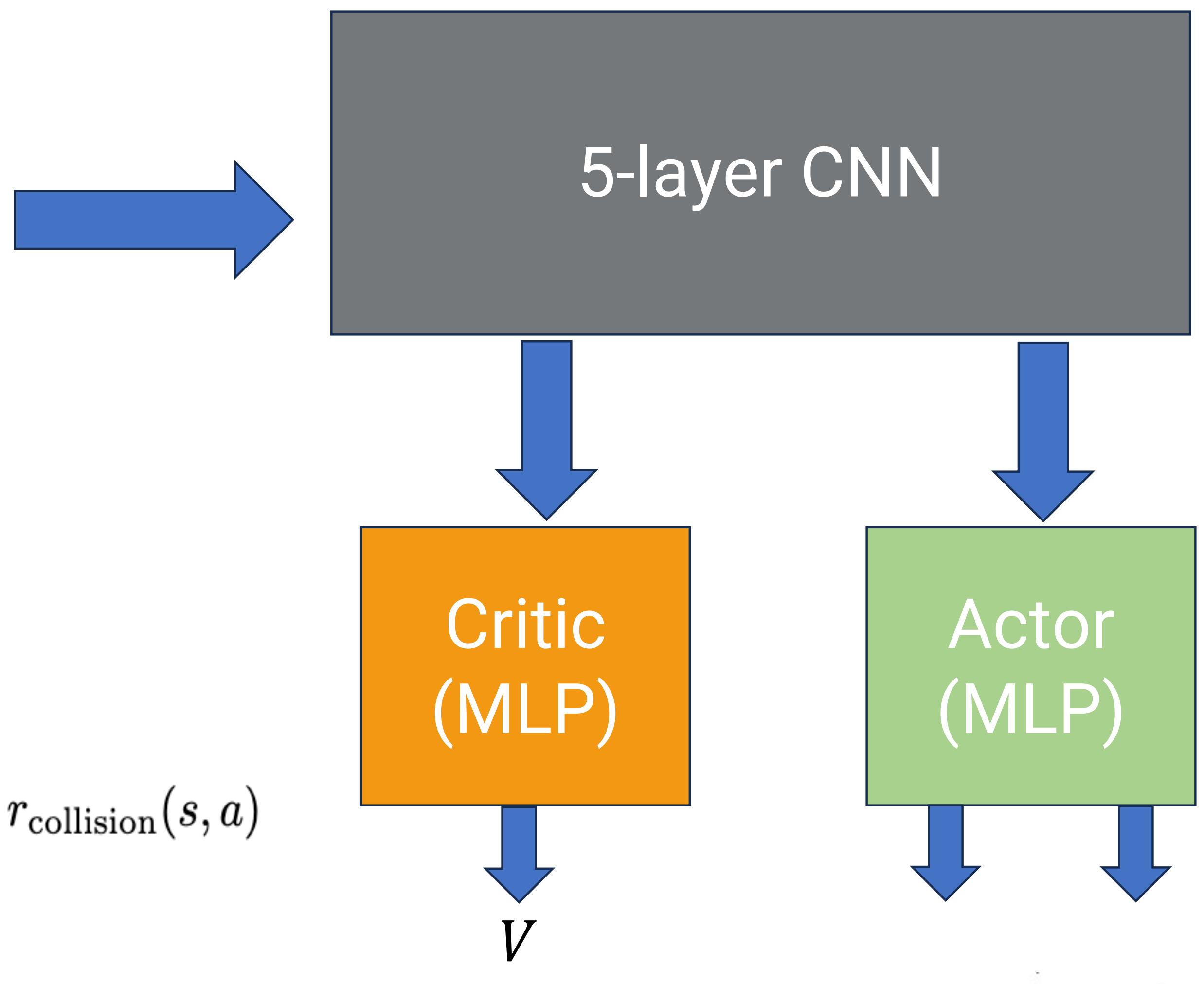
DDGP

- Off-policy DRL approach
- Intentionally adds noise to increase exploration
- Works with continuous action spaces

VISTA Simulator Setup

128x128 Horizontally Stacked Image Observations





$$\pi_\theta = \mathcal{N}(\mu, \sigma^2)$$
$$a \sim \pi_\theta(\cdot\,|\,s_t)$$

Reward Function

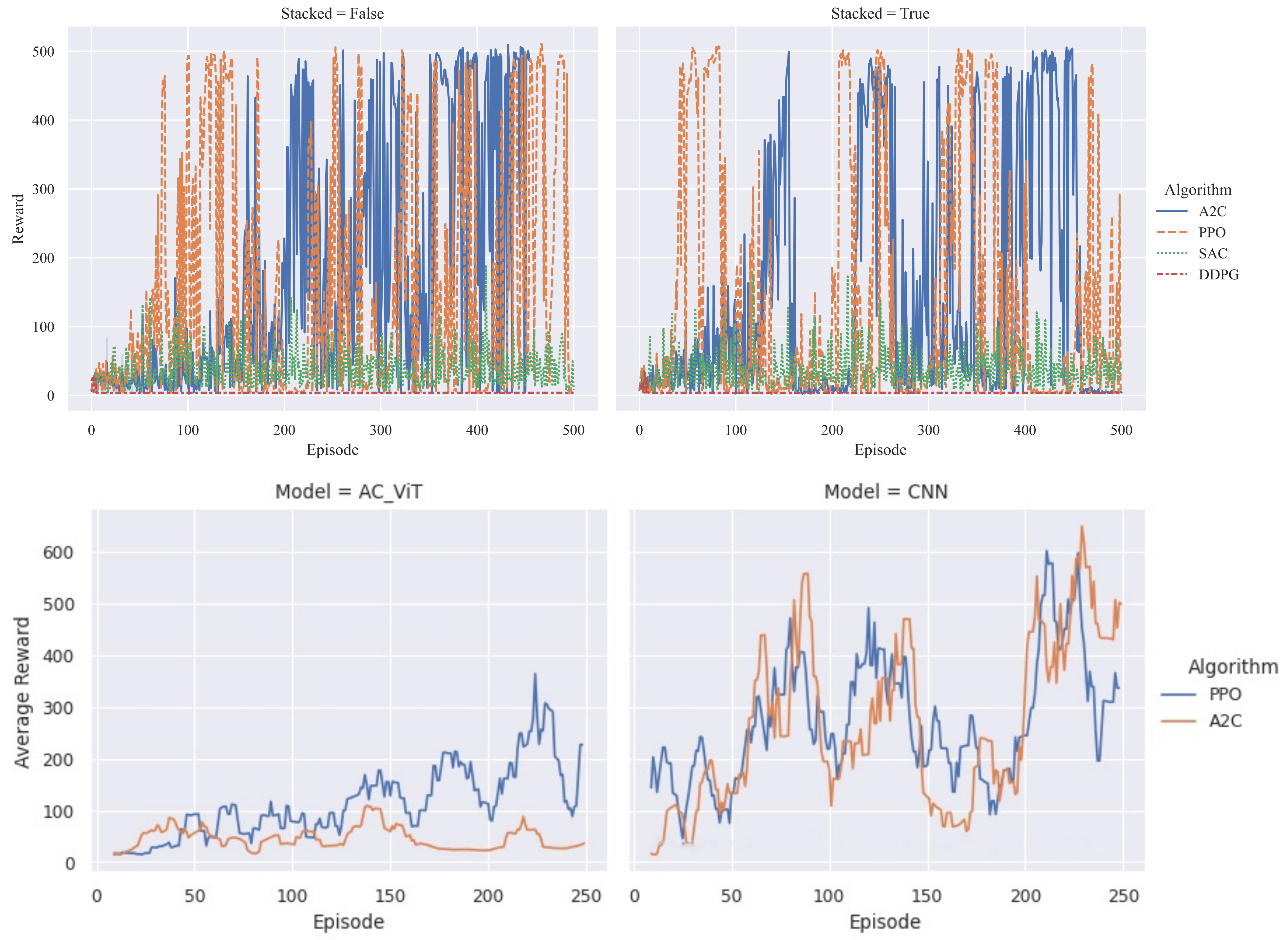
$$r(s,a) = r_{\text{lane}}(s,a) + r_{\text{rotation}}(s,a) + r_{\text{collision}}(s,a)$$

$$r_{\text{lane}} = 1 - (\frac{q_{\text{lat}}}{Z_{\text{lat}}})^2$$

$$r_{\text{collision}} = -\frac{|\text{Dilate}(P_{ego}) \cap P_{other}|}{|P_{ego}|}$$

$$r_{\text{rotation}} = -abs(\delta_{ego} - \delta prev)$$

Learning



- PPO tends to perform best in the lane follow and collision avoidance task
- A2C is the runner up but experiences frequent instability
- We also applied a vision transformer to extract temporal information

Future Work

- Cross domain examination in other simulators
- CNN + LSTM neural network backbone
- Memory optimization for image-based replay buffers