Adversarial Imitation Learning via Random Search in Lane Change Decision-Making

MyungJae Shin (Chung-Ang University), Joongheon Kim (Chung-Ang University)

E-mails: mjshin.ml@gmail.com, joongheon@gmail.com

Introduction

Advanced Driver Assistance Systems for Autonomous Driving
→ The complex ADAS controllers are actively studied to improve road safety.
→ The advanced sensors (i.e., LiDAR sensor) are developed to assist the ADAS controllers.
→ However,
   (1) Coordination methods between the ADAS controllers are not considered.
   (2) Long term driving assistance strategies with the ADAS controllers is not considered.
   (3) Robust driving policy learning method for supervisor is required.
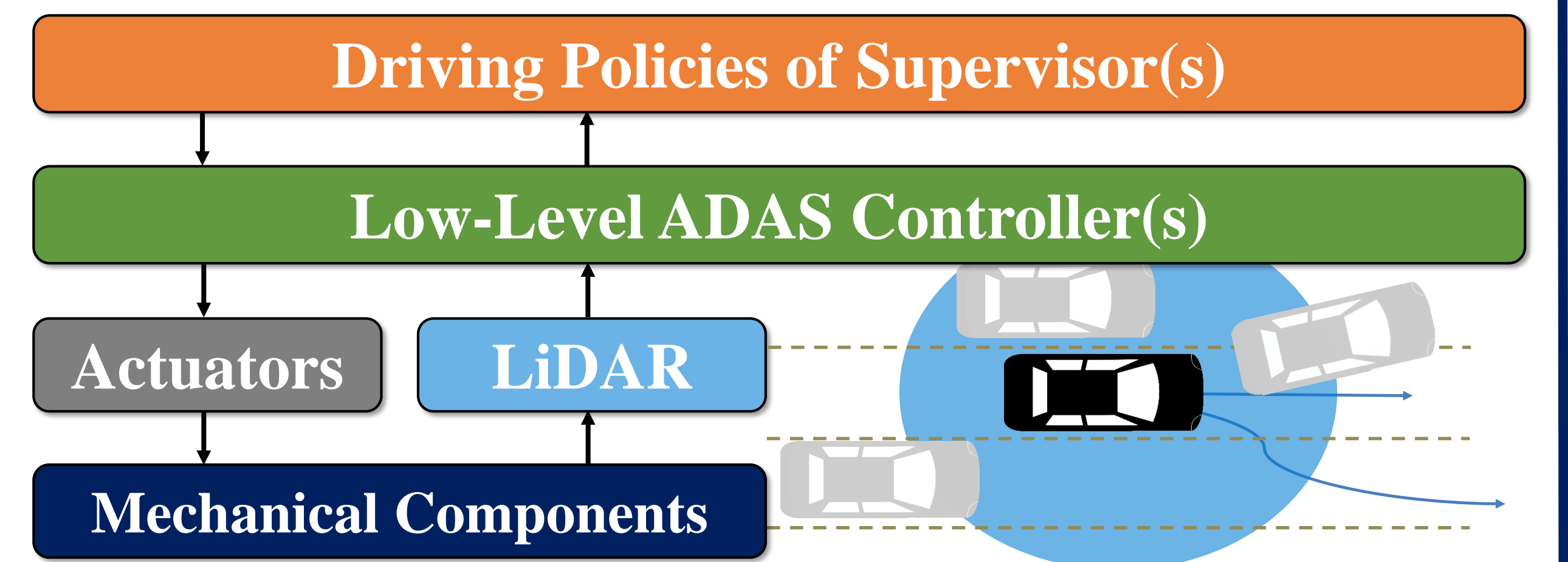
Random Search and Adversarial Imitation Learning for Decision Making

I. Random Search Strategy

The driving agent takes Gaussian noise matrices $\delta = \{\delta_1, \delta_2, \cdots, \delta_N\}$.
→ Adds/Subtracts it to the weights.
→ Perform a driving task with various different weights $\theta_t \pm \nu\delta_i$.
→ Parallel processing is used to collect the rewards $r(\theta_t \pm \nu\delta_i)$.
→ Based on the collected rewards, the update direction is determined.

II. Rewards from Discriminator

The rewards are collected from the discriminator that distinguishes whether the trajectories are that of the agent or that of the expert.

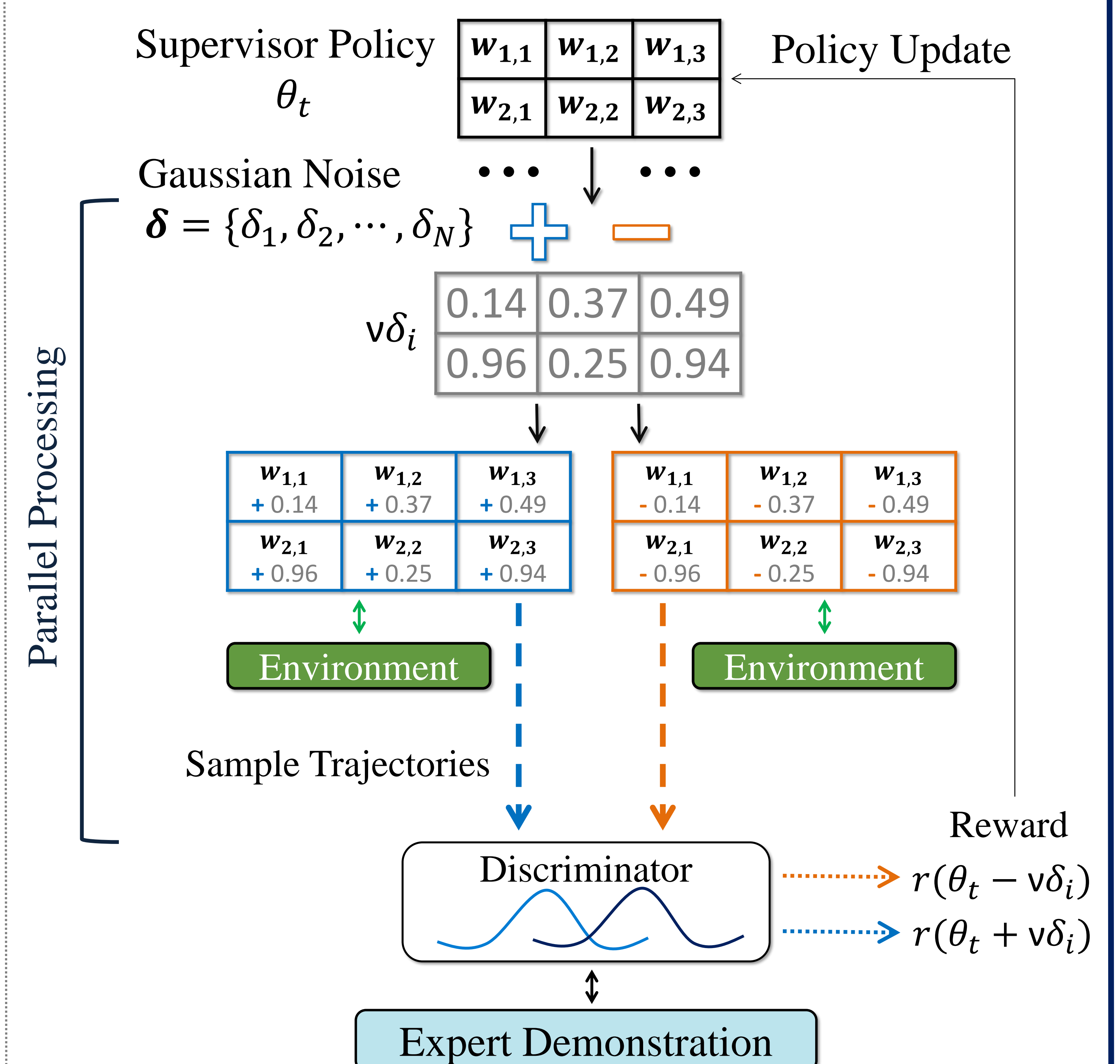$$r(\theta_t \pm \nu\delta_i) = \mathbb{E}[-\log(1 - \mathcal{D}_\phi(Trajectories))]$$

The least square GAN (LS-GAN) is used to improve the performance of the discriminator.

III. Policy Update
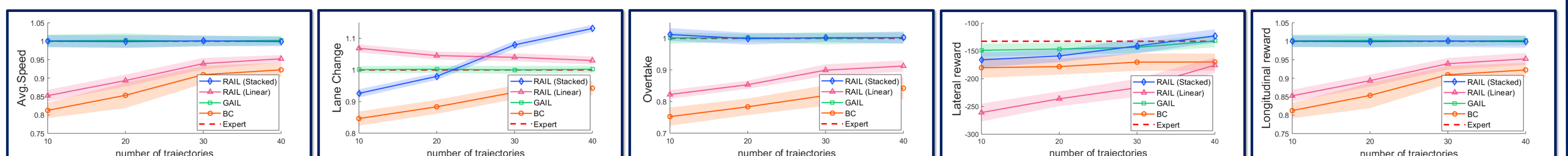
The policy is gradually updated as the follows:

$$\theta_{t+1} = \theta_t + \frac{\alpha}{\sigma_R} \sum_{i=1}^{N} (r(\theta_t + \nu\delta_i) - r(\theta_t - \nu\delta_i))\delta_i$$

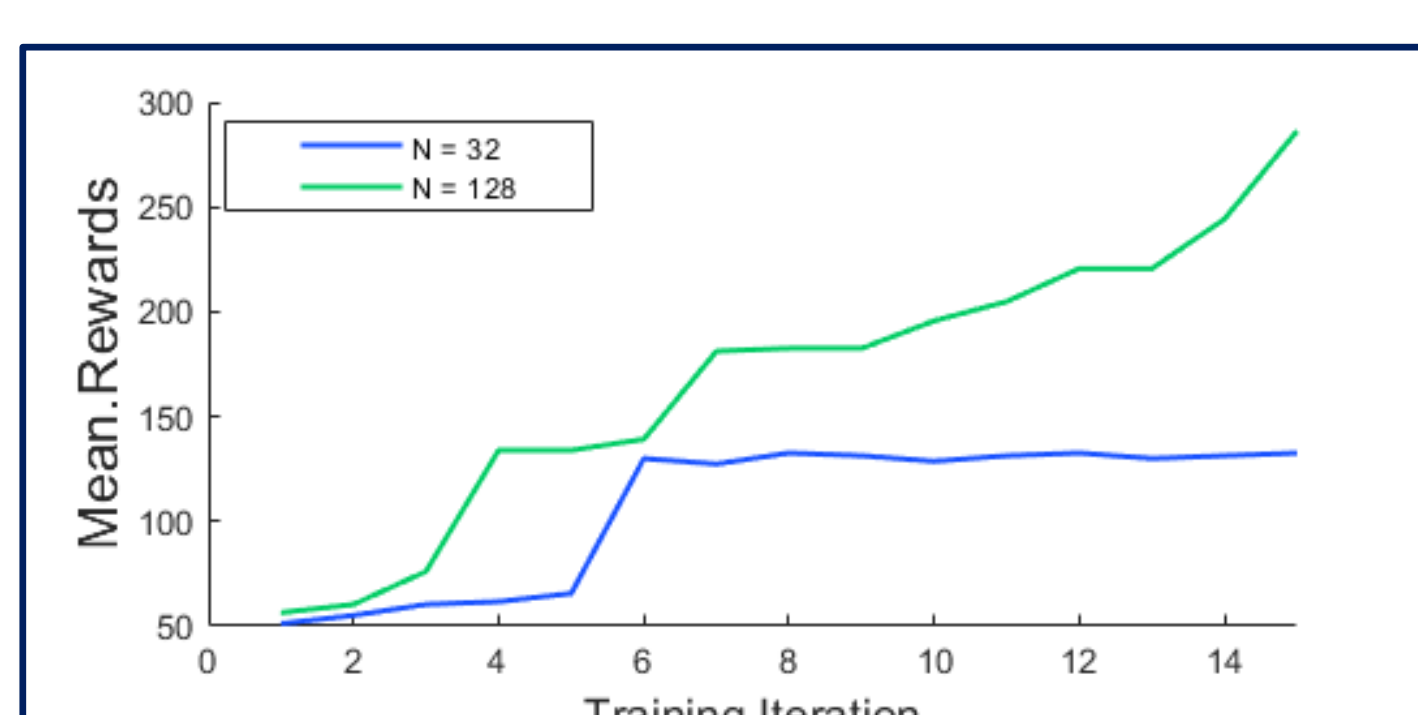To stabilize the update oscillation, the deviation of the collected rewards σ_R is used.

Experimental Results

Driving Performance

The proposed method was evaluated on the multi-lane highway simulator to test the impact of the number of expert trajectories.

Parallel Processing Performance

The impact of the number of parallel process was evaluated.
The experiment shows a change in the average reward value of one trajectory.

Simulator

Observation Space:
(1) LiDAR Sensor
(2) Front Camera
Action Space:
(1) Acceleration
(2) Deceleration
(3) Lane Change