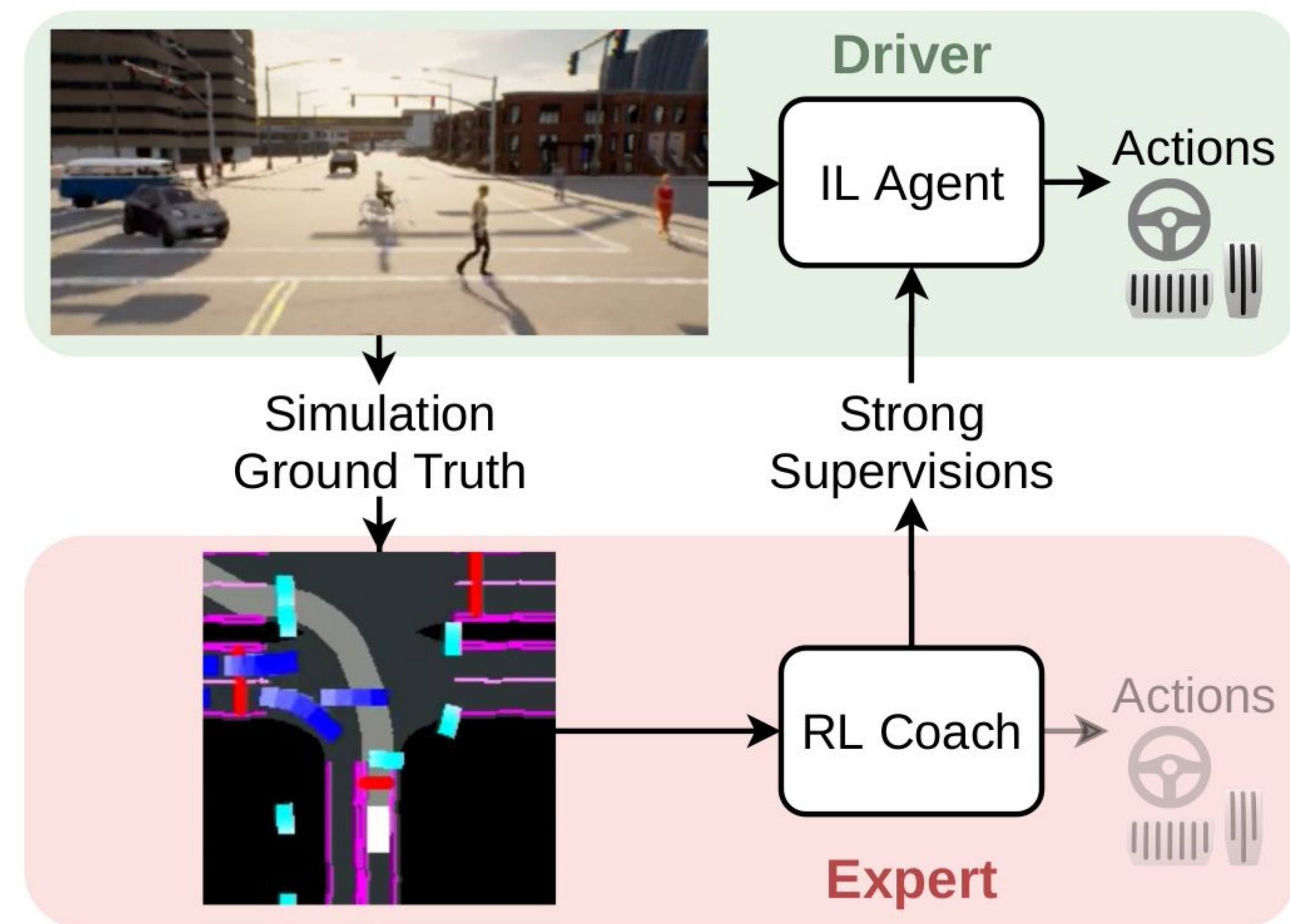


Motivation

Data aggregation (DAGger) alleviates covariate shift by training imitation learning (IL) agents using **on-policy** data.

- Human experts can provide **sparse** on-policy supervision.
- Automated experts can provide **dense** on-policy supervision.

Prior automated experts on CARLA rely on ground-truth information and hand-crafted rules and they perform suboptimally. This limits the performance of all IL agents.



Roach (RL Coach) is an automated expert **end-to-end trained** using reinforcement learning (RL).

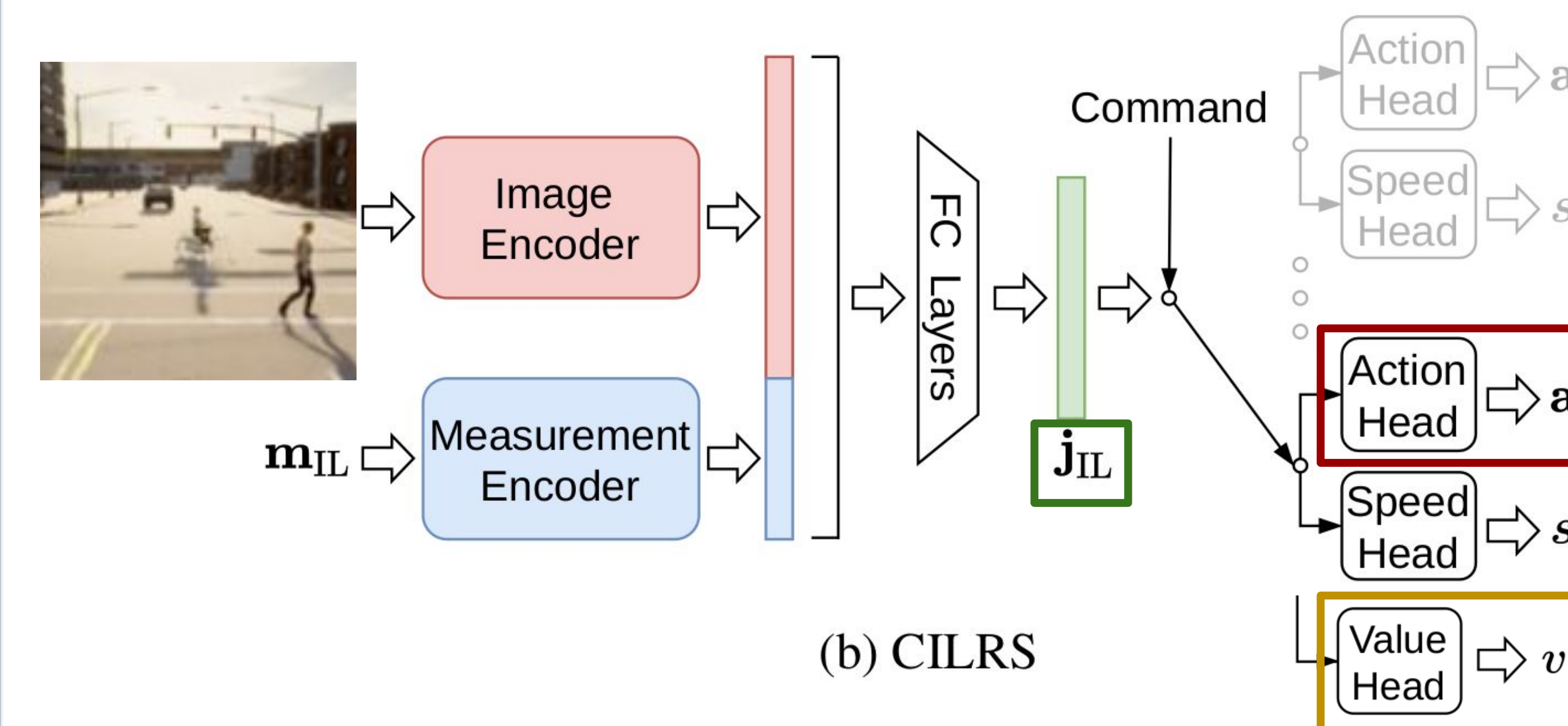
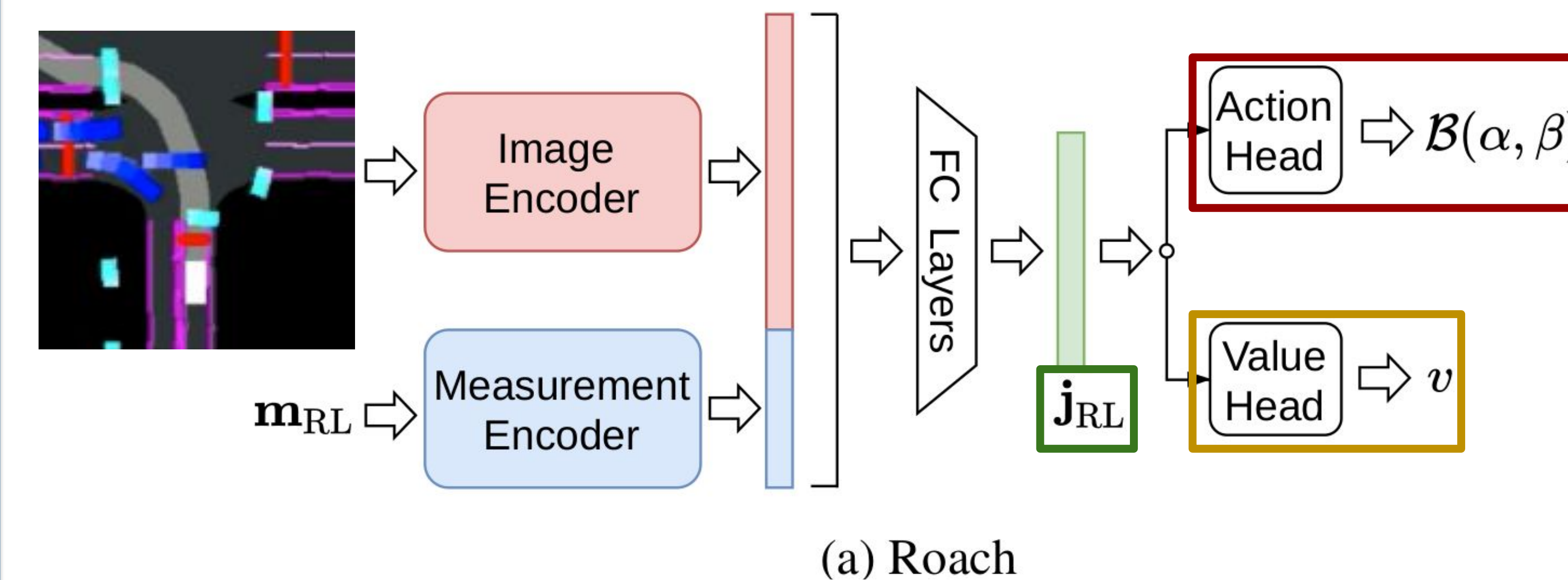
Roach can boost the performance of IL methods on CARLA because

- it raises the performance upper-bound of IL by achieving **better driving performance** than other experts,
- it allows more effective IL by providing **dense and informative** on-policy supervision.

Our Method

STEP 1: Train Roach using Proximal Policy Optimization (PPO) with

- **bird's-eye view** as input,
- **Beta distribution** as output,
- **exploration suggest**, a generalized entropy loss that encodes basic traffic rules.



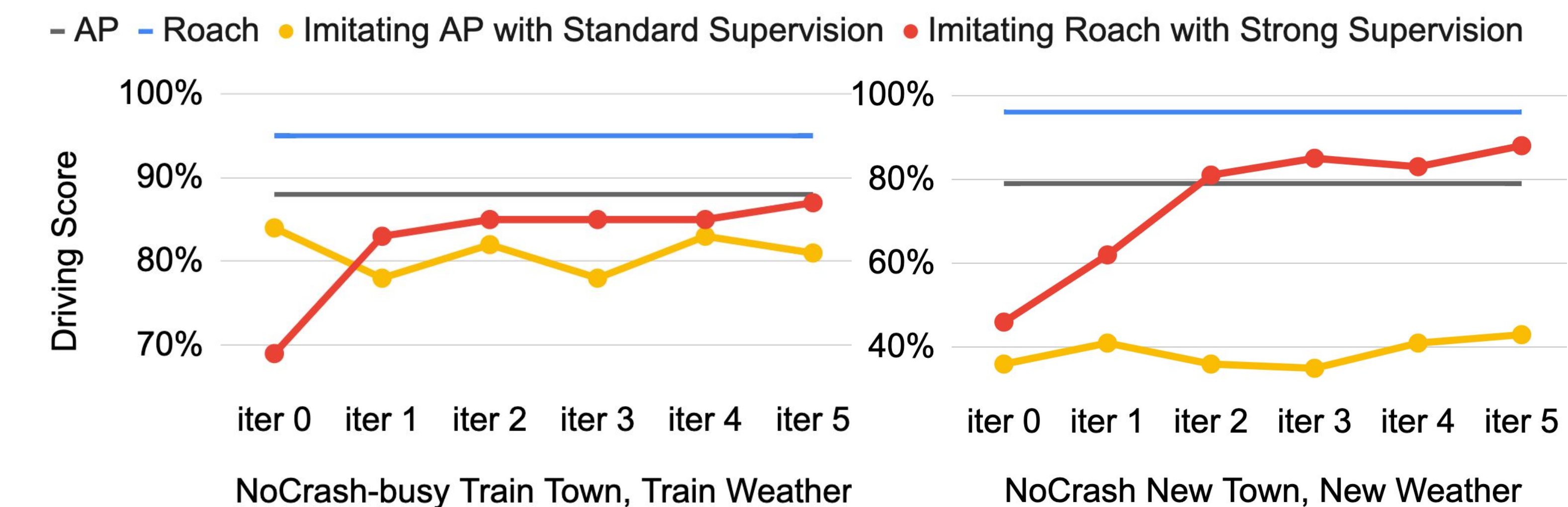
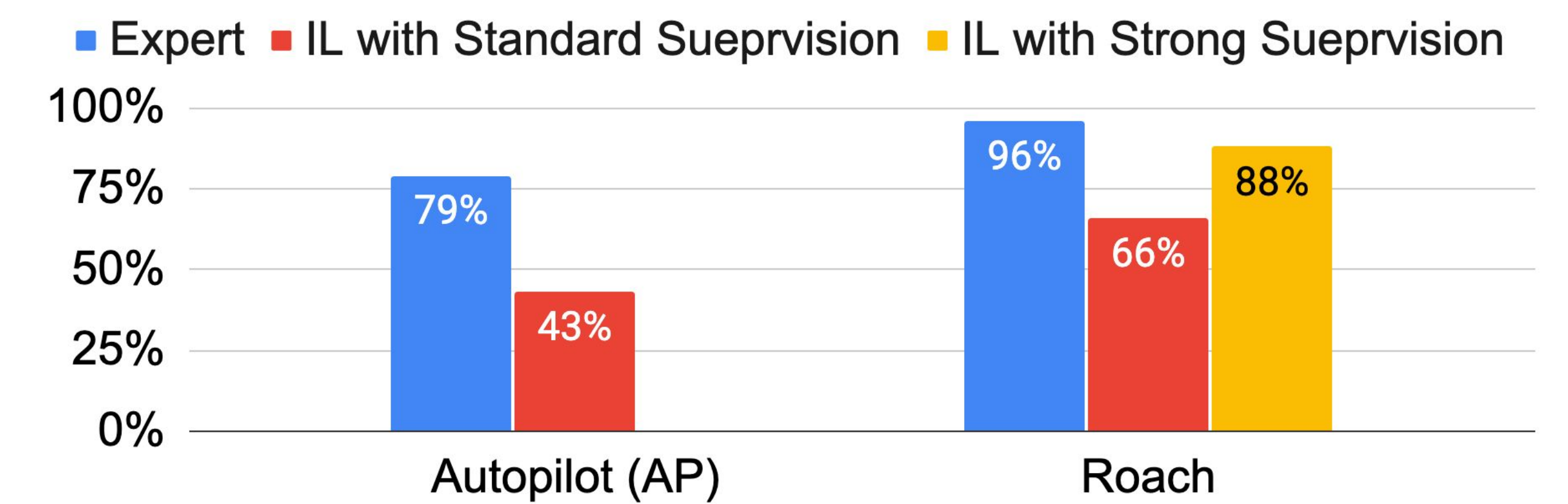
STEP 2: Train IL agents using DAGGER and strong supervisions from Roach.

The improved training scheme

- uses **action distribution** as a more informative target,
- regress **value estimation** as an auxiliary task,
- matches **latent feature** for knowledge distillation.

Performance

Driving Score (\uparrow) on NoCrash-Busy, New Town, New Weather



Complex Driving Scenarios



More videos are found at: <https://www.trace.ethz.ch/publications/2021/roach>