

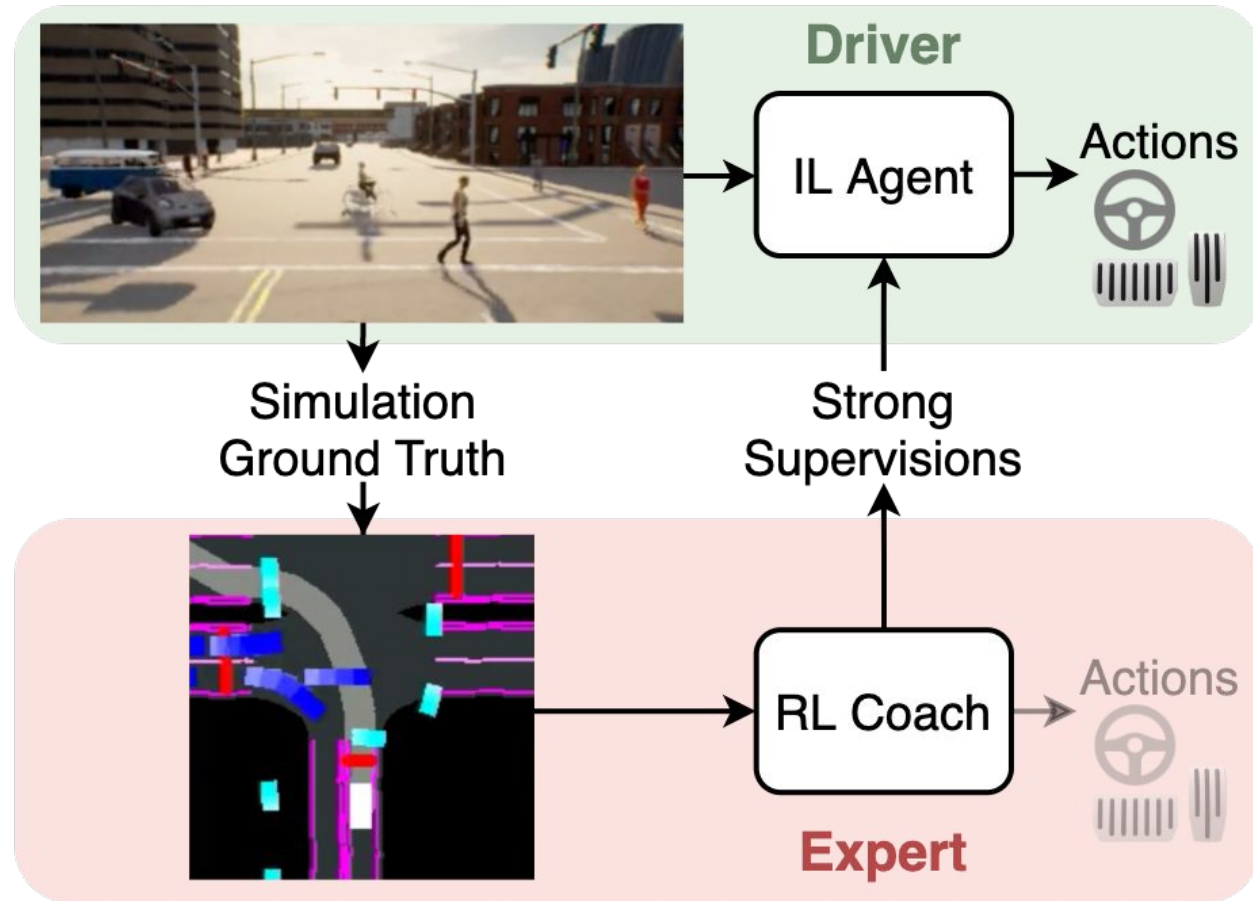
# End-to-End Urban Driving by Imitating a Reinforcement Learning Coach

Zhejun Zhang<sup>1</sup>, Alexander Liniger<sup>1</sup>, Dengxin Dai<sup>1,2</sup>,  
Fisher Yu<sup>1</sup> and Luc Van Gool<sup>1,3</sup>

<sup>1</sup>Computer Vision Lab, ETH Zürich, <sup>2</sup>MPI for Informatics,  
<sup>3</sup>PSI, KU Leuven



# End-to-End Urban Driving by Imitating a Reinforcement Learning Coach



## Imitation Learning Agents

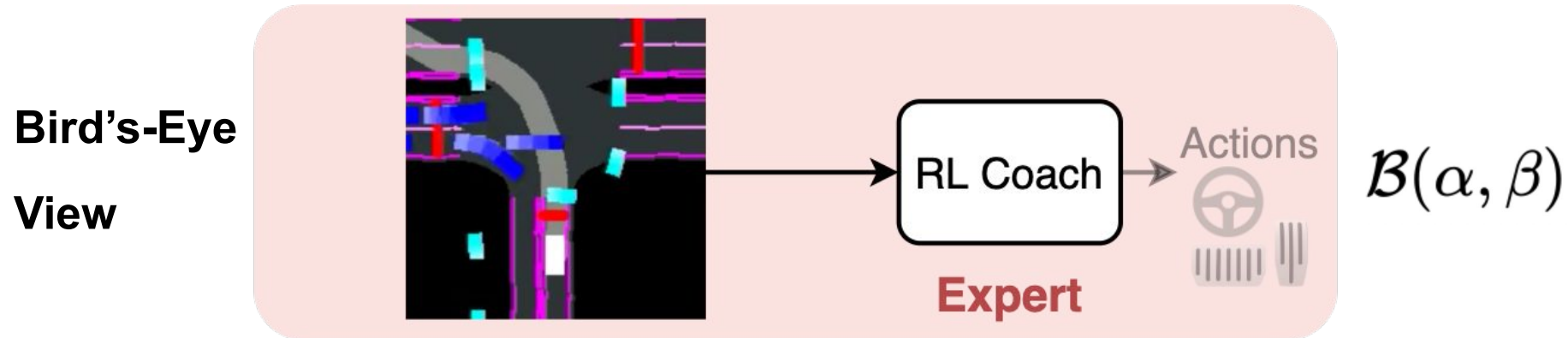
learn from

**dense and informative on-policy** supervisions

provided by the

**Roach,**  
**the Reinforcement Learning Coach.**

# STEP 1: Roach, the RL Coach



Proximal Policy Optimization (PPO) with

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

# STEP 1: Roach, Highlights

- Excellent Performance
  - Better than any other agents, including the CARLA Autopilot.
- High Sample Efficiency
  - Training from scratch takes 4 days on one 2080 Ti.
- Complex Traffic Scenarios



Stop Sign



Roundabout



Highway



Lane Changing

# STEP 2: Data Aggregation (DAgger)

DAgger uses **on-policy** supervisions.

1. Execute the IL policy, record behavior and mistakes made by the policy.
2. Query the expert for the correct actions.
  - a. Human experts: **sparse**
  - b. Automated experts: **dense**

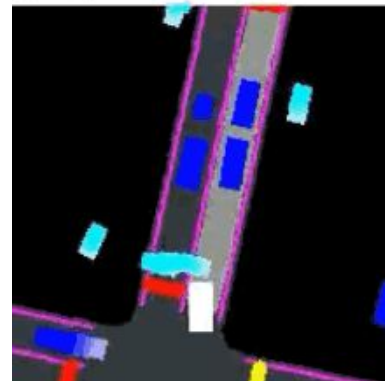


↑ Input to IL, driving  
2x Speed (10 FPS)

Speed  
7.5 Km/h

Command  
Turn Right

↓ Input to RL, coaching



Acc Output

0.35

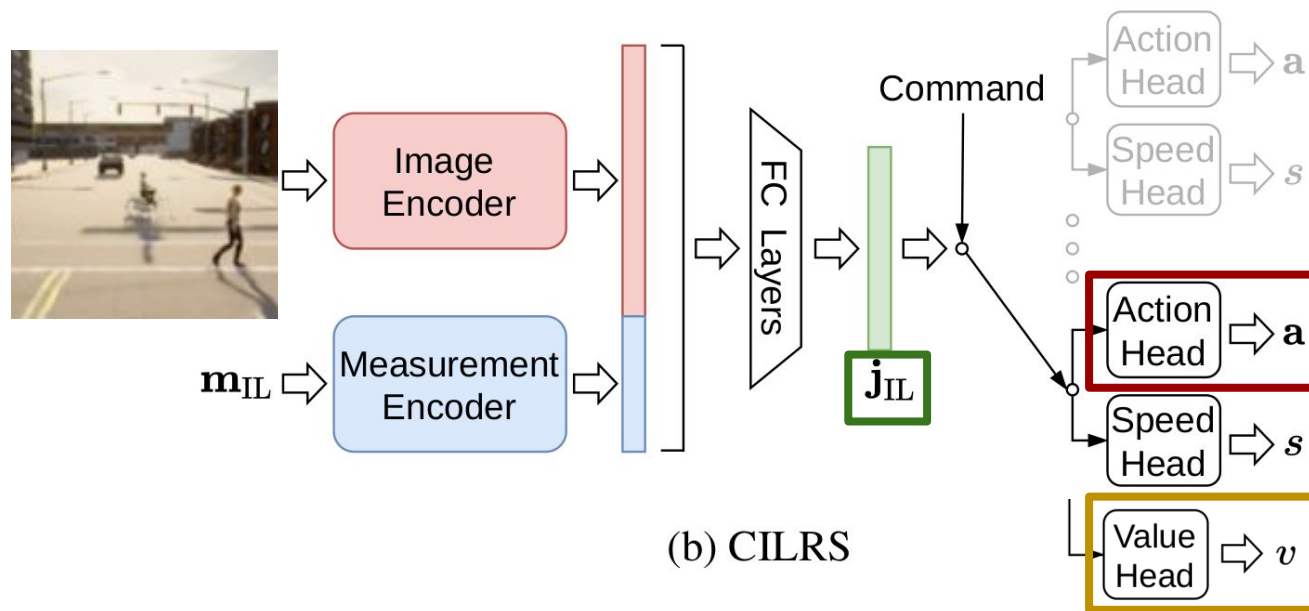
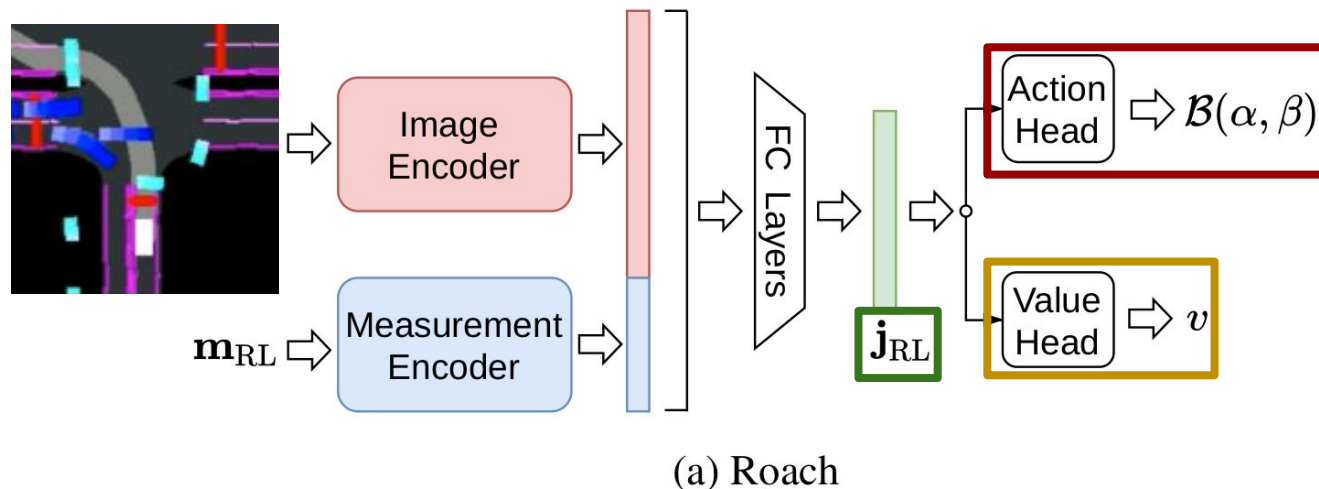
Steer Output

0.40

Brake 0 Gas Left 0 Right

Red: RL Expert. Blue: IL Agent.

# Roach: Dense and Informative On-Policy Supervision

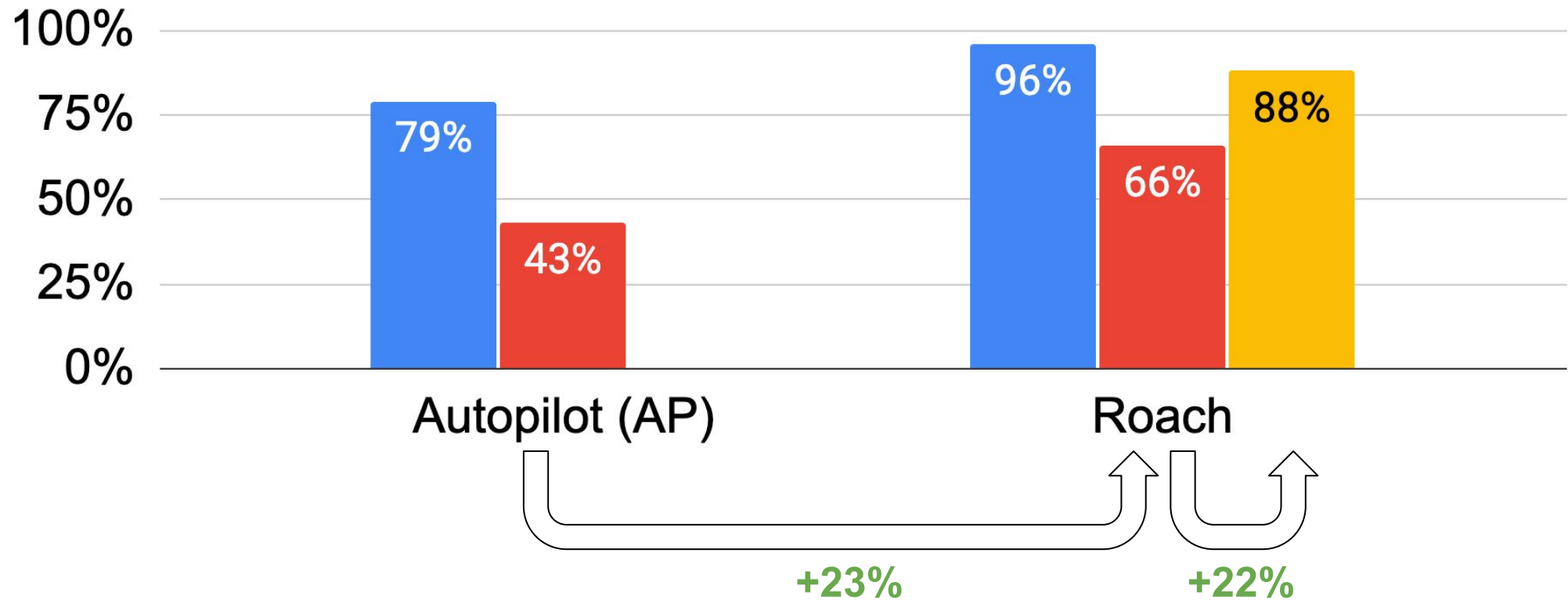


- Standard Supervisions
  - $\mathcal{L}_A$  Deterministic Action (L1 loss)
- Roach: Informative Supervisions
  - $\mathcal{L}_K$  Action Distribution (KL loss)
  - $\mathcal{L}_V$  Value Estimation (L2 loss)
  - $\mathcal{L}_F$  Latent Feature (L2 loss)

# Expert-Level Driving Performance

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

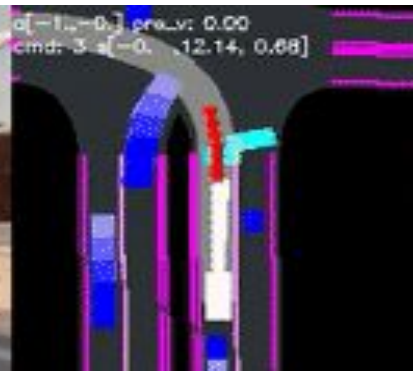
■ Expert ■ IL with Standard Supervision ■ IL with Strong Supervision



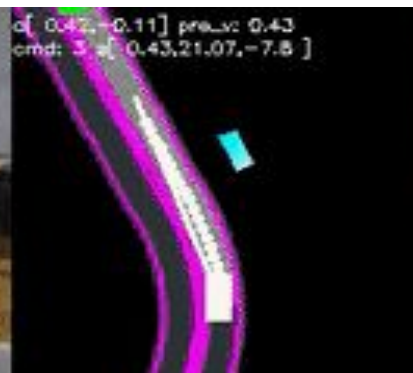
# Challenging Driving Scenarios



Roundabout



Unprotected  
Left Turn



US Style  
Traffic Lights





↑ Input to IL, driving  
2x Speed (10 FPS)

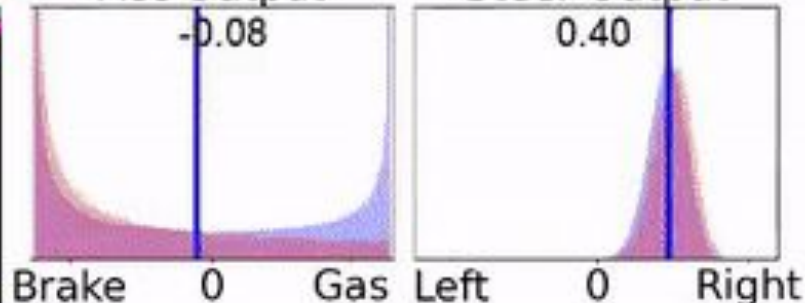
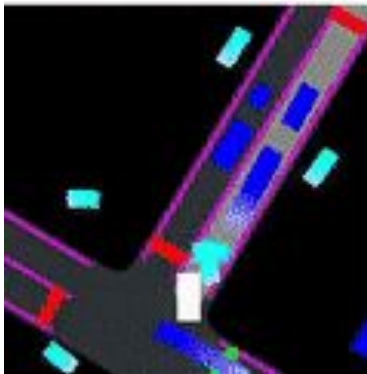
Speed  
0.0 Km/h  
Command  
Turn Right

**Stop for pedestrians while turning.**

↓ Input to RL, coaching

Acc Output

Steer Output



Brake 0 Gas Left 0 Right  
Red: RL Expert. Blue: IL Agent.

**Roach-Supervised IL Agent  
Driving in Town02  
NoCrash-Dense  
New Town & New Weather**

More videos and information: <https://www.trace.ethz.ch/publications/2021/roach>

Code and checkpoints: <https://github.com/zhejz/carla-roach>