DeFIX: Detecting and Fixing Failure Scenarios with Reinforcement Learning in Imitation Learning Based Autonomous Driving

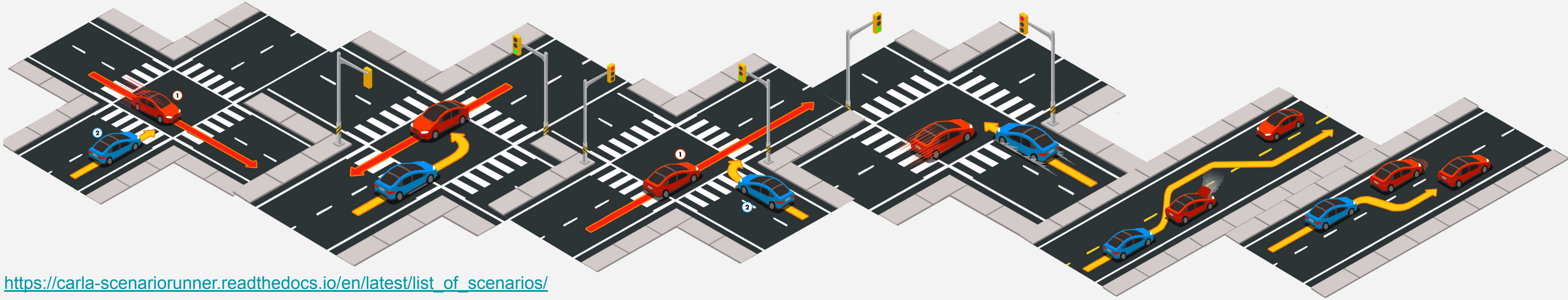
October 8-12, 2022, Macau, China

Resul Dagdanov, Feyza Eksen, Halil Durmus, Ferhat Yurdakul, and Nazim Kemal Ure

- Istanbul Technical University
- ITU Artificial Intelligence and Data Science Research Center
- Eatron Technologies

Problem Statement

- Decision making in adversarial urban driving scenarios is a **hard problem**.
- Designing a **perfect** rule-based agent would necessitate a **lot of hard-coding** to tackle these scenarios.
- Objective is to come up with a methodology that **solves shortcomings** of suboptimal policy.
- CARLA Challenge environment is chosen as it features **state-of-the-art** urban driving scenarios:
  - vehicles run in red lights
  - pedestrians cross the road irregularly
  - decisions in uncontrolled traffic intersections occur
  - vehicles get stuck on the road



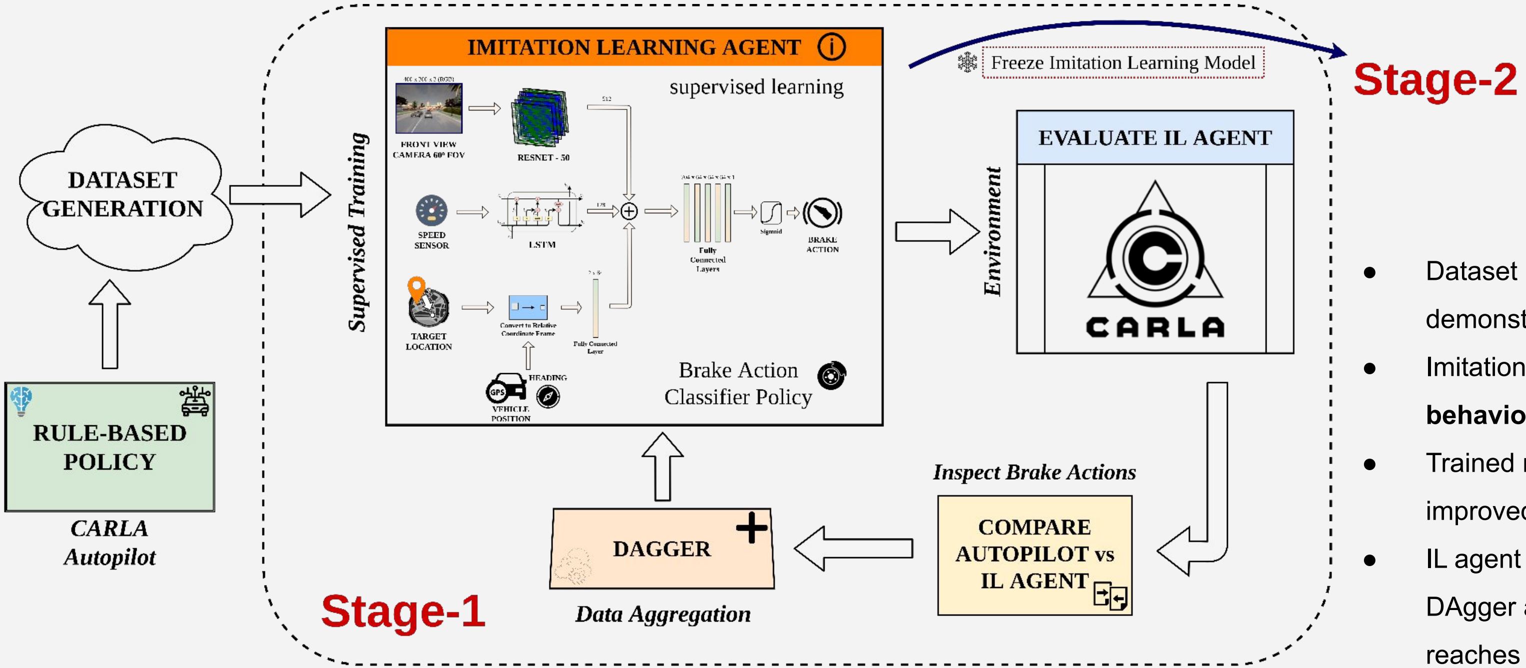
https://carla-scenariorunner.readthedocs.io/en/latest/list_of_scenarios/

Motivation

autopilot vehicle gets stuck

autopilot vehicle brakes late

- Hard-coded autopilot is **sensitive** to predefined rules.
- Impossible to **generalize** with rule-based system.
- Imitating a suboptimal policy will result in a **lack of performance** of trained model.
- General **methodology / framework** is required to efficiently solve safe autonomous driving tasks.



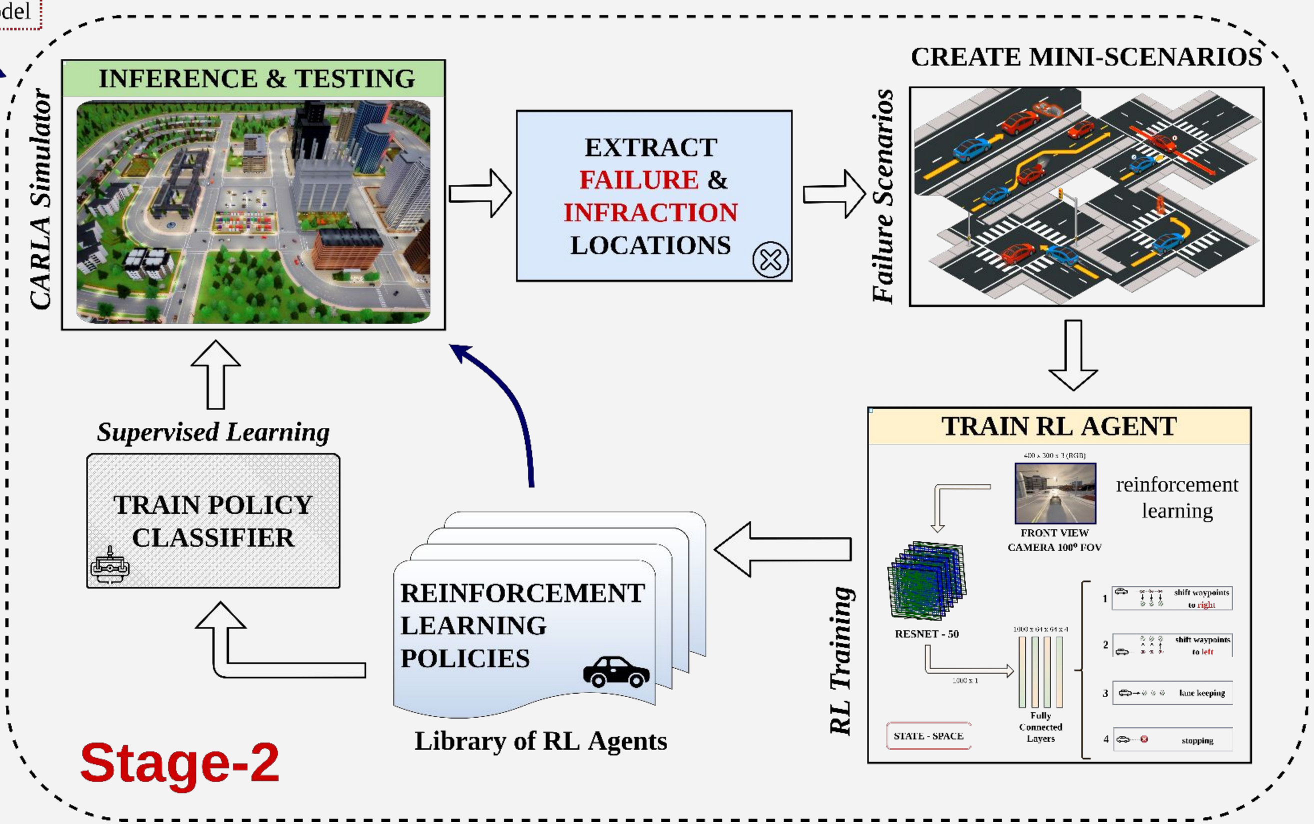
Stage-1:

- Dataset is created from **autopilot** demonstrations.
- Imitation Learning agent is trained with **behavioral cloning**.
- Trained model is **continuously** improved with DAgger technique.
- IL agent is freezed after multiple DAgger approaches until **performance** reaches an autopilot performance.

Methodology (Stage-2)

Freeze Imitation Learning Model

Stage-1

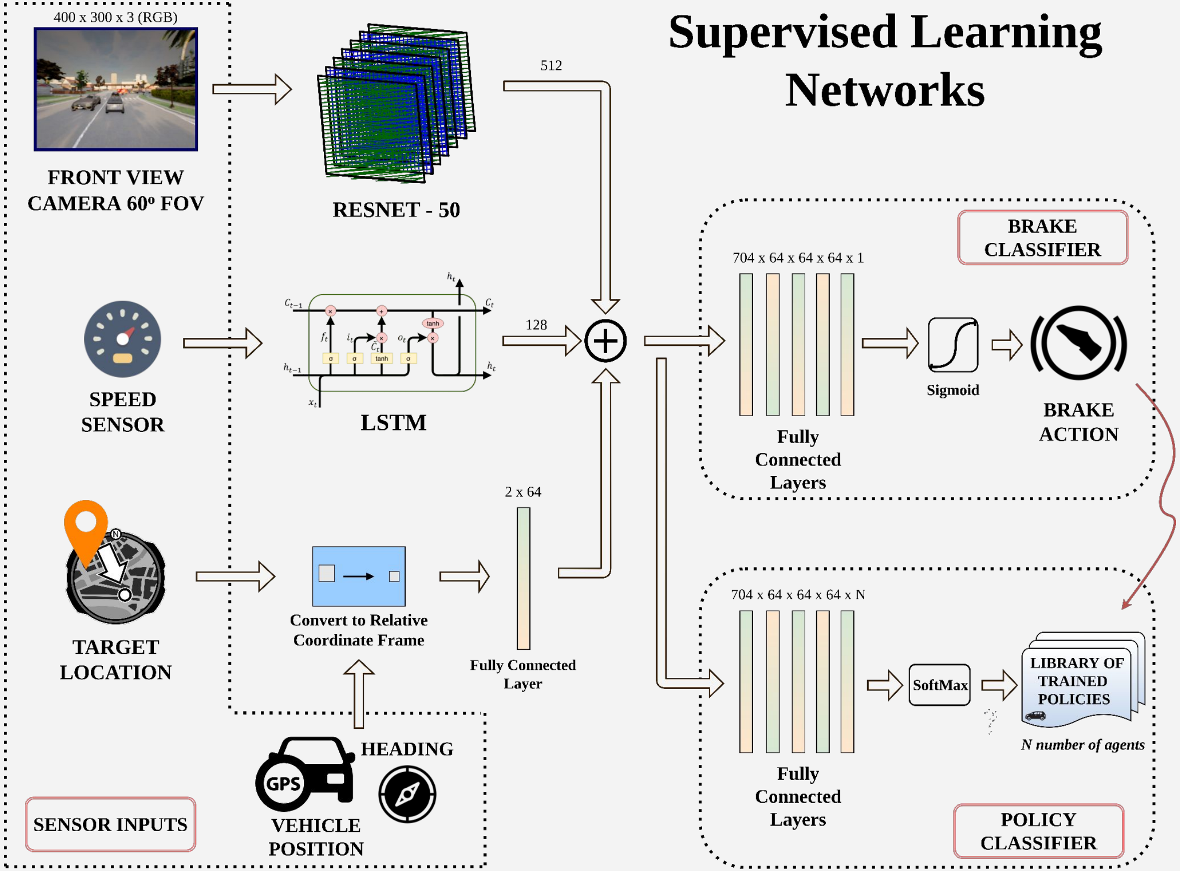


Stage-2:

- Trained IL model is **freezed** and tested in the simulation.
- Mini-scenarios are constructed from **infraction** locations.
- Online RL agent is **trained** on one of the mini-scenarios.
- Library of trained policies is **extended** with newly trained RL agent.
- **Policy classifier** network is trained in a supervised manner with ground truth determined from reward function.

Network Architectures

Supervised Learning Networks



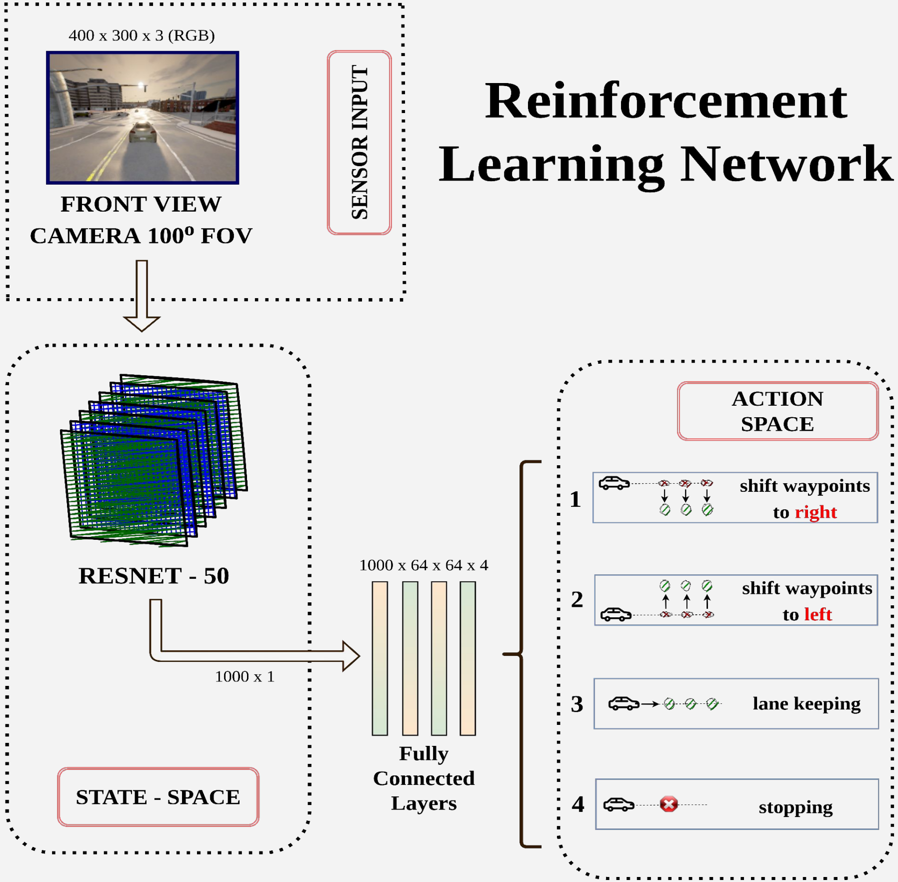
- Freezed Resnet-50 **image backbone**.
- Low-level **PID controllers** for throttle and steering actions.
- Ground truth labels for Policy Classifier are determined with **reward function**.

Imitation Learning Agent Model

Policy Classifier Model

Network Architectures

- **DQN** model is trained.
- Image backbone is **completely frozen**.
- **State-space** of RL agents is an image backbone features.
- Low-level **PID controllers** for throttle and steering movements.
- Generalized **action states** are:
    - Left lane change
    - Right lane change
    - Lane Keep
    - Stop



Experiments (Reward Function)

$$
\begin{aligned}
\mathcal{R}(s_i) = & \delta(s_i) + [1 - \xi_{c_s}] \cdot [(1 - \phi(s_i)) \cdot V(s_i) \\
& + 50 \cdot \phi(s_i) \cdot \beta(s_i) - \phi(s_i) \cdot V(s_i)] \\
& + \xi_{c_s} \cdot V(s_i) - 100 \cdot \tau(s_i) - 1500 \cdot \zeta(s_i)
\end{aligned}
$$

- s_i: observation state at a given step i
- $V(s_i)$: speed of the agent vehicle in m/s
- ξ_{c_s}: 1 - if $c_s(s_i, t_i) = 2$; 0 - otherwise
- $\beta(s_i)$: brake command of an agent vehicle [0 or 1]
- $\phi(s_i)$: 1 - if an affecting traffic light is red or yellow, existence of vehicles and pedestrians in the field of influence of a rule-based agent; 0 - otherwise
- $\delta(s_i)$: Euclidean distance between agent's location and latest passed waypoint location in $meters$
- $\tau(s_i)$: 1 - if agent velocity is 0.0 m/s for 60 consecutive seconds; 0 - otherwise
- $\zeta(s_i)$: boolean function revealing any collisions

GENERAL LIST OF DEPICTED CARLA SCENARIOS

ID $c_s(s_i, t_i)$	**Scenario Name**
0	Dynamic Vehicle Collision
1	Emerging Pedestrian Collision
2	Stuck Vehicle & Static Objects
3	Vehicle Running Red Light
4	Crossing Signalized Traffic Intersections
5	Crossing Un-signalized Intersections

$$
label(s_i) = \begin{cases} IL\ Agent, & \text{if} \quad \mathcal{R}(s_i) > 0.0 \\ c_s(s_i, t_i), & \text{if} \quad \mathcal{R}(s_i) \leq 0.0 \end{cases}
$$

Experiments (RL Training)

- An autopilot and an imitation learning agent **gets stuck** in these scenarios and could not make lane changing maneuvers.
- One mini-scenario of stuck vehicle case is **constructed** and DQN agent is trained on this scenario.
- To generalize, **weather conditions** are dynamically changed at every 20 steps while training RL agent.
- At each training episode, RL agent is **initialized randomly** at a distance 3 meters to 10 meters to the infraction location.

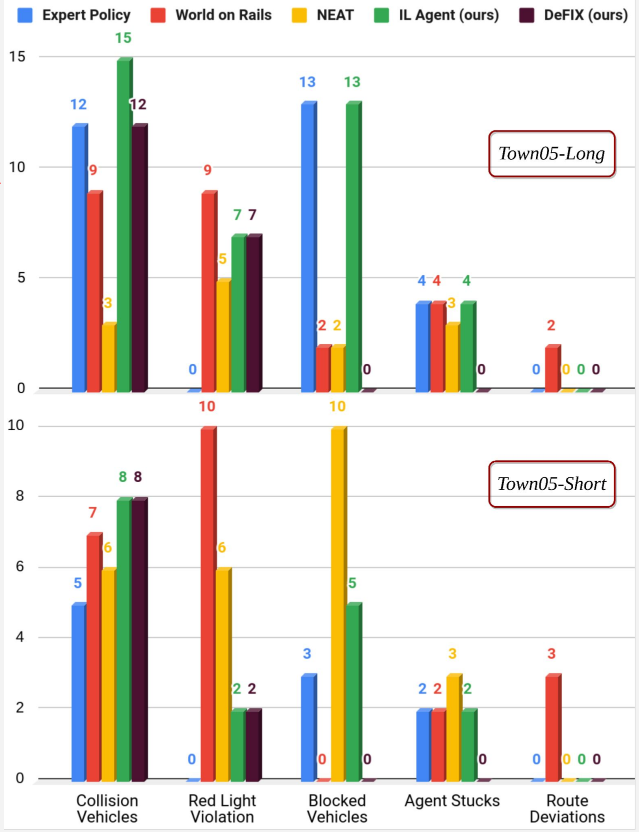
Stuck Vehicle Scenarios in Town05 Map

Blue: Training Scenario
Orange: Evaluation Scenarios

Results

RC: Route Completion (higher is better)
IS: Infraction Score (higher is better)
DS: Driving Score (higher is better)

Total Number of Infractions



IMITATION LEARNING AGENT PERFORMANCE ON TOWN05

		Performance Metrics		
Town05-Short	Trained Dataset	**RC** ⇑	**IS** ⇑	**DS** ⇑
Imitation Learning Agent	$D_0 \rightarrow 22K$	89.8±2.27	0.74±0.01	64.95±1.3
	$D_1 \rightarrow 24K$	90.0±2.57	0.68±0.01	62.91±3.4
	$D_2 \rightarrow 26K$	**90.1**±1.12	**0.76**±0.02	**68.47**±2.9
Autopilot Agent	N/A	90.94	0.91	82.75
Town05-Long	Trained Dataset	**RC** ⇑	**IS** ⇑	**DS** ⇑
Imitation Learning Agent	$D_0 \rightarrow 22K$	75.1±0.33	0.39±0.09	25.59±4.8
	$D_1 \rightarrow 24K$	75.4±0.03	0.41±0.03	30.91±2.2
	$D_2 \rightarrow 26K$	**75.4**±0.06	**0.44**±0.03	**33.17**±2.2
Autopilot Agent	N/A	75.41	0.693	48.60

Results

RC: Route Completion (higher is better)
DS: Driving Score (higher is better)

- DeFIX outperforms **state-of-the-art** benchmark.
- **Better RC score** than autopilot with only relying on sensor information without privileged knowledge.
- Achieve **%15** better RC score and **%3** closer DS.
- Town05 is the most challenging map of CARLA challenge environment.

BENCHMARK COMPARISON OF DRIVING PERFORMANCE SCORES

Method	Town05-Short		Town05-Long	
	RC ⇑	DS ⇑	RC ⇑	DS ⇑
LBC [7]	55.01	30.97	32.09	7.05
Late Fusion [22]	83.66	51.56	68.05	31.30
CILRS [23]	13.40	7.47	7.19	3.68
AIM [22]	81.07	49.00	60.66	26.50
TransFuser [22]	78.41	54.52	56.36	33.15
NEAT [3]	69.34	58.21	88.78	**57.49**
Geometric Fusion [22]	86.91	54.32	69.17	25.30
World on Rails [9]	52.60	38.14	60.57	32.18
DeFIX (ours)	**96.34**	**72.41**	**89.61**	39.42
IL Agent (ours)	90.10	68.47	75.40	33.17
RL Agent (ours)	30.14	24.65	5.17	4.15
Autopilot Agent	90.94	82.75	75.41	48.60

Conclusion

- Imitation learning policy is **limited** with the performance of the demonstrator policy.
- Proposed and tested a new methodology called **DeFIX**.
- Achieved better **DS** and **RC** scores than the methods in the literature on the most challenging map.
- The **future** of this research:
  - Extensive tests on different maps.
  - Expand the library of RL agents and breaden the effectiveness.
  - Carry the evaluation of DeFIX method into official CARLA challenge.