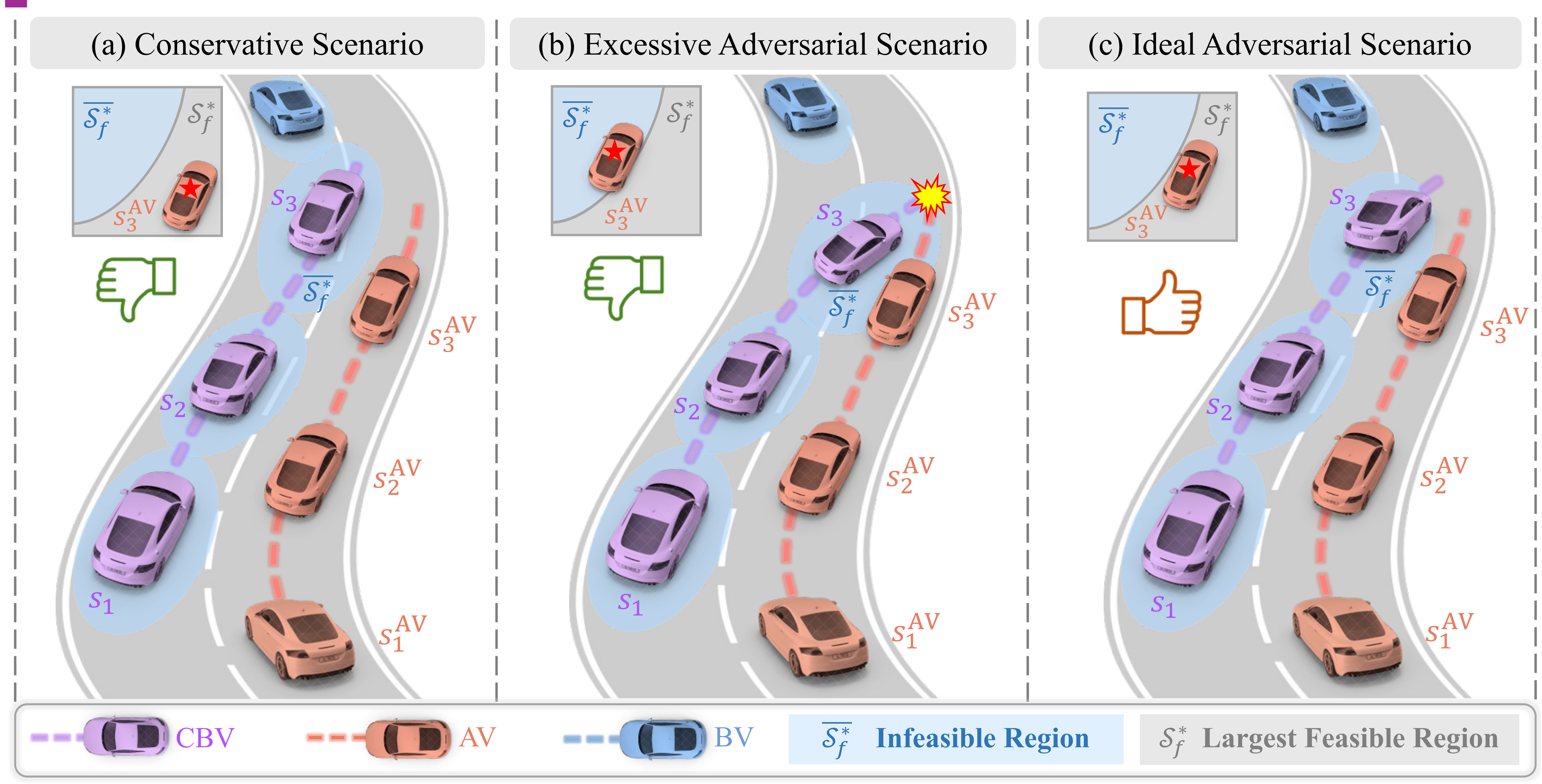
FREA: Feasibility-Guided Generation of Safety-Critical Scenarios with Reasonable Adversariality

Keyu Chen[1], Yuheng Lei[2], Hao Cheng[1], Haoran Wu[1], Wenchao Sun[1], Sifa Zheng[1]
[1]SVM, Tsinghua University & [2]The University of Hong Kong.

Code is available

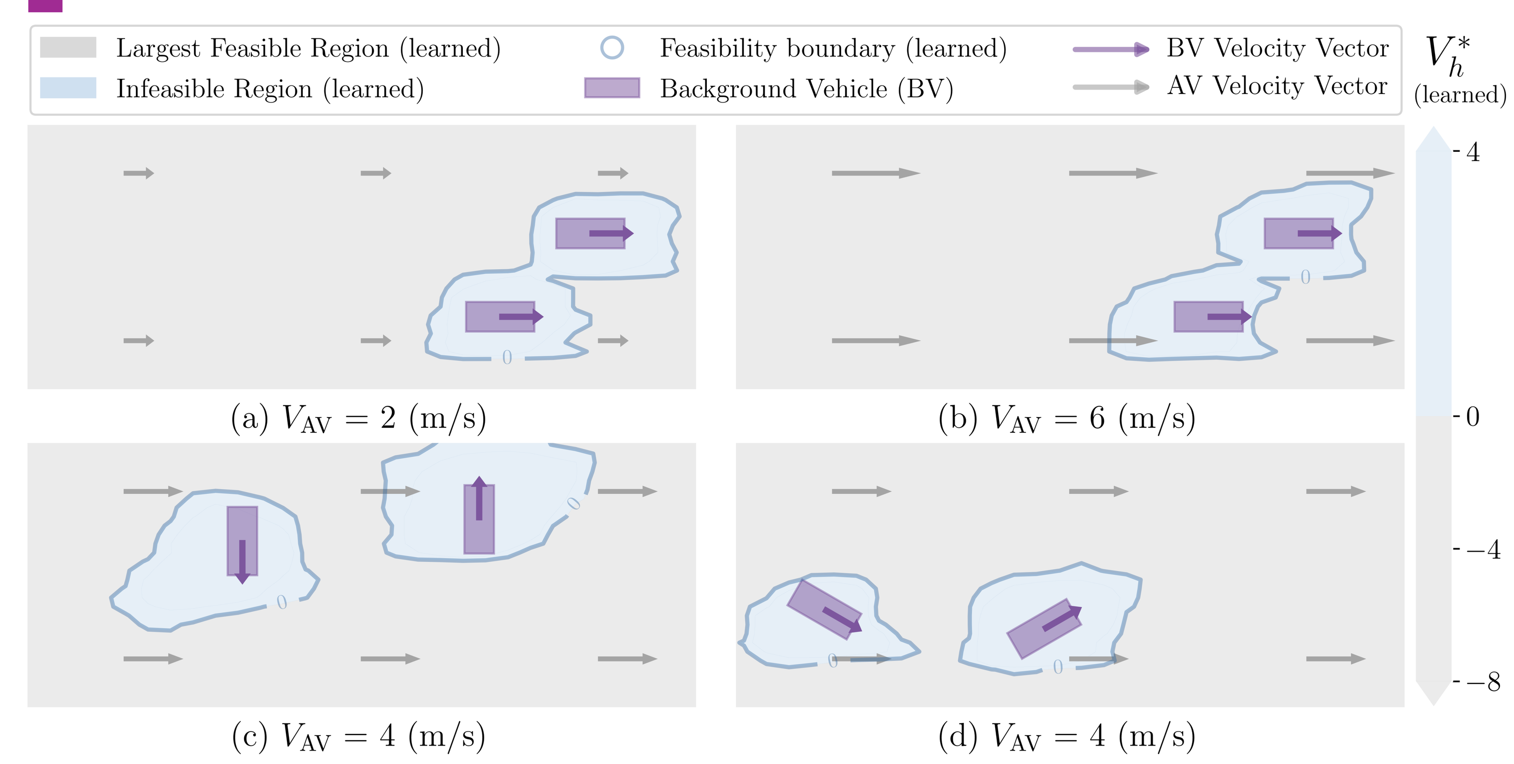
Introduction



Core Contribution:

FREA incorporates feasibility as guidance to generate adversarial yet AV-feasible, safety-critical scenarios for autonomous driving.

LFR Visualization



The well-trained LFR is reliable under various traffic scenarios.

Feasibility Metrics

Table 1: Feasibility evaluation using Expert [26] as AV under different CBV methods. Results are the average of 10 runs in "Scenario9" with varied seeds.

CBV	Feasibility	Town05 intersections			Town02 intersections		
		CR (↓)	IR (↓)	ID (↓)	CR (↓)	IR (↓)	ID (↓)
KING[5]	✗	76.67%	65.97%	7.54m	N/A	N/A	N/A
PPO	✗	37.5%	35.56%	10.57m	30.0%	51.18%	12.40m
FPPO-RS	✓	11.25%	34.92%	9.13m	24.29%	45.36%	9.16m
FREA	✓	**5.0%**	**31.10%**	**6.25m**	**5.71%**	**27.18%**	**4.94m**

FREA balances adversariality with AV feasibility for minimal collision severity.

Methods

Algorithm 1 Feasibility-guided reasonable adversarial policy (*FREA*)

1: **Offline Part** (Section 3.1)
2: Initialize feasibility value networks V_h, Q_h.
3: **for** each gradient step **do**
4: Update V_h using Eq. (3) # Optimal feasible state-value function learning
5: Update Q_h using Eq. (4) # Optimal feasible action-value function learning
6: **end for**
7: **Online Part** (Section 3.2)
8: Initialize policy parameters θ_0, reward value function parameters ψ_0
9: **for** $k = 0, 1, 2, \ldots$ **do**
10: Collect set of trajectories $\mathcal{B}_k = \{\tau_i\}$ with policy π_{θ_k}, where τ_i is a T-step episode.
11: Compute reward advantage $A_r^{\pi_{\theta_k}}(s,a)$, using generalized advantage estimator (GAE [23]).
12: Compute feasibility advantage using Eq. (9).
13: Derive overall advantage using Eq. (6) # Advantage calculating
14: Fit reward value function, by Smooth L1 Loss. # Value function learning
15: Update the policy parameters θ by maximizing Eq. (5). # Policy learning
16: **end for**

Largest Feasible Region (LFR):

Definition 1 (Optimal feasible value function). *Based on [13, 15, 22], the optimal feasible state-value function V_h^* and the optimal feasible action-value function Q_h^* are defined in Eqs. (1) and (2).*

$$V_h^*(s^{AV}) := \min_{\pi^{AV}} \max_{t \in \mathbb{N}} h\left(s_t^{AV}\right), s_0^{AV} = s^{AV}, a_t^{AV} \sim \pi^{AV}\left(\cdot \mid s_t^{AV}\right), \tag{1}$$

$$Q_h^*(s^{AV}, a^{AV}) := \min_{\pi^{AV}} \max_{t \in \mathbb{N}} h\left(s_t^{AV}\right), s_0^{AV} = s^{AV}, a_0^{AV} = a^{AV}, a_{t+1}^{AV} \sim \pi^{AV}\left(\cdot \mid s_{t+1}^{AV}\right). \tag{2}$$

Definition 2 (Largest Feasible Region (LFR)). *The largest feasible region is the sub-zero level set of the optimal feasible state-value function.*

$$\mathcal{S}_f^* := \left\{ s^{AV} \mid V_h^*(s^{AV}) \leq 0 \right\}$$

Approximate LFR through Offline Learning:

$$\mathcal{L}_{V_h}(\omega) = \mathbb{E}_{\mathcal{D}}\left[L_{\text{rev}}^{\tau}\left(Q_h(s^{AV}, a^{AV}; \phi) - V_h(s^{AV}; \omega)\right)\right], \tag{3}$$

$$\mathcal{L}_{Q_h}(\phi) = \mathbb{E}_{\mathcal{D}}\left[\left(\left((1-\gamma)h(s^{AV}) + \gamma \max\left\{h(s^{AV}), V_h(s^{AV\prime}; \omega)\right\}\right) - Q_h(s^{AV}, a^{AV}; \phi)\right)^2\right], \tag{4}$$

Optimal Feasible Advantage Function:

Lemma 1. *As the BVs follow deterministic policy, the optimal feasible action-value function of AV can be achieved by AV's current state and next state (see Appendix A.2 for proof).*

$$Q_h^*\left(s^{AV}, a^{AV}\right) = \begin{cases} V_h^*(s^{AV}) & h(s^{AV\prime}) \geq h(s^{AV}) \\ \max\{h(s^{AV}), V_h^*(s^{AV\prime})\} & h(s^{AV\prime}) < h(s^{AV}) \end{cases} \tag{7}$$

$$A_h^*(s^{AV}, a^{AV}) = Q_h^*\left(s^{AV}, a^{AV}\right) - V_h^*\left(s^{AV}\right) \tag{8}$$

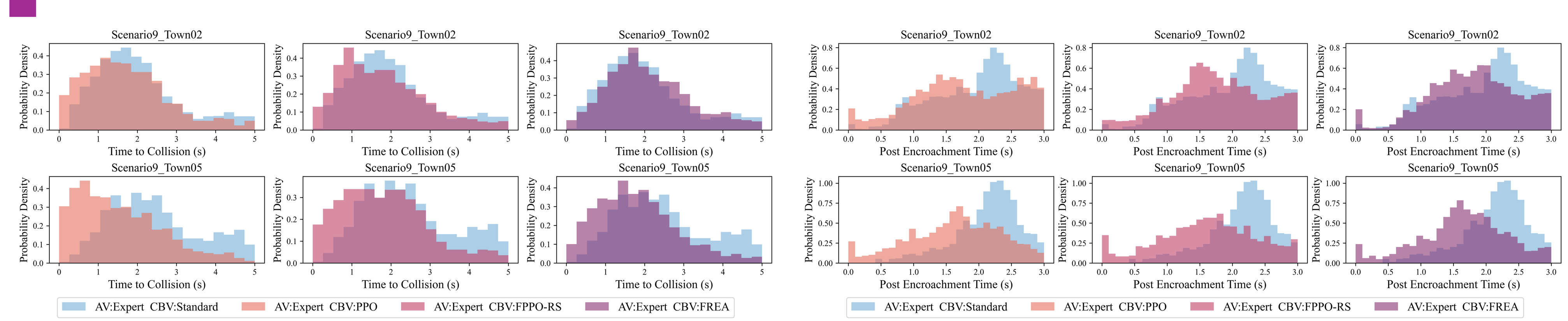
$$= \begin{cases} V_h^*\left(g\left(s'\right)\right) - V_h^*\left(g\left(s\right)\right) & h\left(g\left(s'\right)\right) \geq h\left(g\left(s\right)\right) \\ \max\{h\left(g\left(s\right)\right), V_h^*\left(g\left(s'\right)\right)\} - V_h^*\left(g\left(s\right)\right) & h\left(g\left(s'\right)\right) < h\left(g\left(s\right)\right) \end{cases} \tag{9}$$

Feasibility-dependent Objective Function :

$$L(\theta) = \mathbb{E}_{\pi_{\theta_k}}\left[\min\left(r_t(\theta) A^{\pi_{\theta_k}}(s,a), \operatorname{clip}\left(r_t(\theta), 1-\epsilon, 1+\epsilon\right) A^{\pi_{\theta_k}}(s,a)\right)\right], \tag{5}$$

$$A^{\pi_{\theta_k}}(s,a) = A_r^{\pi_{\theta_k}}(s,a) \cdot I(s,s') + A_h^*(s^{AV}, a^{AV}) \cdot (1 - I(s,s')), \tag{6}$$

Near-Miss Metrics



FREA effectively generate safety-critical scenarios, yielding considerable near-miss events.

Generalization of AV Testing

Table 2: Comparative performance of AVs across different maps, using CBV methods pre-trained with various surrogate AVs. Results are the average of 10 runs in "Scenario9" with varied seeds.

CBV	Surr. AV	AV	Town05 intersections						Town02 intersections					
			CR (↓)	OR (↓)	RF (↓)	UC (↓)	TS (↓)	OS (↑)	CR (↓)	OR (↓)	RF (↓)	UC (↓)	TS (↓)	OS (↑)
Standard	✗	Expert	0.0%	0.0m	7.0m	1%	55s	94.0	0.0%	0.0m	6.0m	2%	63s	93.0
		PlanT	1.0%	0.0m	7.0m	6%	70s	90.0	1.0%	0.0m	6.0m	6%	76s	90.0
PPO	Expert	Expert	36.0%	0.0m	6.0m	8%	66s	76.0	40.0%	0.0m	6.0m	15%	66s	72.0
		PlanT	61.0%	1.0m	7.0m	11%	70s	65.0	70.0%	0.0m	6.0m	27%	64s	57.0
PPO	PlanT	Expert	26.0%	0.0m	6.0m	8%	64s	80.0	21.0%	0.0m	6.0m	12%	74s	80.0
		PlanT	45.0%	0.0m	7.0m	7%	69s	72.0	51.0%	0.0m	6.0m	18%	70s	67.0
FREA	Expert	Expert	4.0%	0.0m	7.0m	7%	67s	**89.0**	9.0%	0.0m	6.0m	16%	75s	**83.0**
		PlanT	10.0%	0.0m	7.0m	5%	73s	**86.0**	10.0%	0.0m	7.0m	24%	86s	**79.0**
FREA	PlanT	Expert	5.0%	0.0m	7.0m	5%	62s	**90.0**	14.0%	0.0m	6.0m	15%	75s	**82.0**
		PlanT	9.0%	0.0m	7.0m	6%	73s	**87.0**	17.0%	0.0m	7.0m	18%	83s	**79.0**

FREA exhibits strong generalization in AV testing under various AV methods and traffic environment.

AV Training Results

Table 5: Comparative performance of AVs pretrained with various CBV methods across different maps. Results are the average of 10 runs in "Scenario9" with varied seeds.

Surr. CBV	CBV	Town05 intersections						Town02 intersections					
		CR (↓)	OR (↓)	RF (↓)	UC (↓)	TS (↓)	OS (↑)	CR (↓)	OR (↓)	RF (↓)	UC (↓)	TS (↓)	OS (↑)
Standard	Standard	11%	4m	19m	4%	68s	85	**39%**	8m	20m	18%	57s	70
PPO		17%	11m	17m	6%	66s	82	40%	11m	19m	15%	63s	70
FREA		**3%**	6m	18m	1%	75s	**89**	40%	3m	18m	19%	56s	**71**
Standard	**FREA**	39%	6m	17m	7%	66s	73	79%	3m	17m	35%	45s	51
PPO		36%	15m	21m	6%	66s	74	79%	4m	14m	37%	44s	51
FREA		**31%**	12m	17m	5%	71s	**76**	**73%**	3m	20m	28%	51s	**55**

FREA provides effective data (safety-critical data) for AV training, thus improving policy robustness.

Representative Scenarios