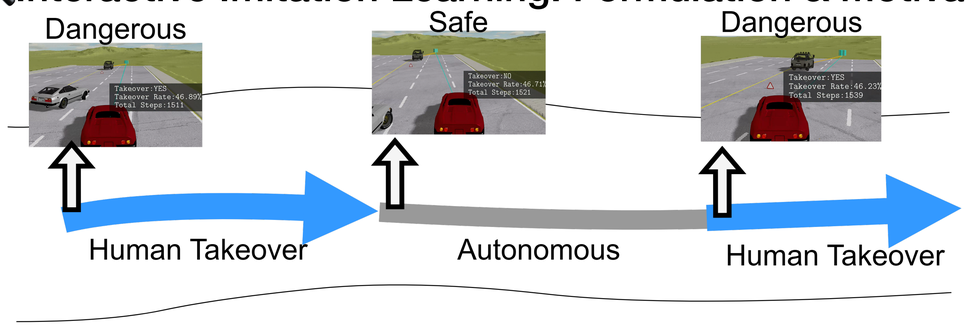
Robot-Gated Interactive Imitation Learning with Adaptive Intervention Mechanism

Haoyuan Cai, Zhenghao Peng, Bolei Zhou
UCLA Computer Science Department

Interactive Imitation Learning: Formulation & Motivation



Human-Gated Intervention:

Human: ***Bad action***, *I help you!* — Human: *Good action, go.* — Human: *I want a rest ...*

Robot-Gated Interaction:

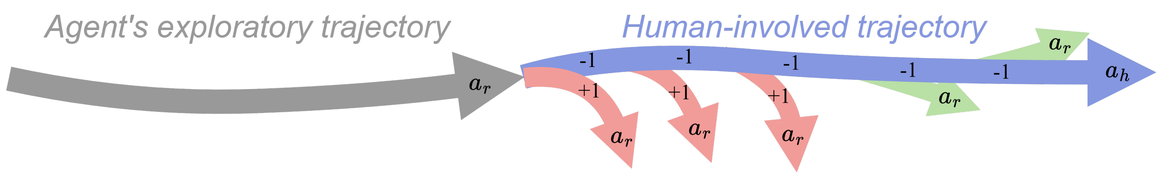
Agent: *I'm **uncertain**, help!* — Agent: *I can proceed safely.* — Human: *Still energetic!*

Human: *Sure, I help you!* — Human: *I can rest now.*

- **Human-Gated**: human watch every step → **heavy cognitive load**.
- **Robot-Gated**: agent **ask help when needed** → saves human effort.

Can we build a robot-gated intervention mechanism that reliably detects danger and requests help before a crash?

Adaptive Intervention Mechanism (AIM)



TD update rule

$$Q(s,a) \equiv \gamma Q(s',a')$$

$$J^{\mathrm{TD}}(\theta) = \mathop{\mathbb{E}}_{(s,a,s')} |Q_\theta(s,a) - \gamma \max_{a'} Q_{\hat{\theta}}(s',a')|^2.$$

*Only for **bad** agent actions*

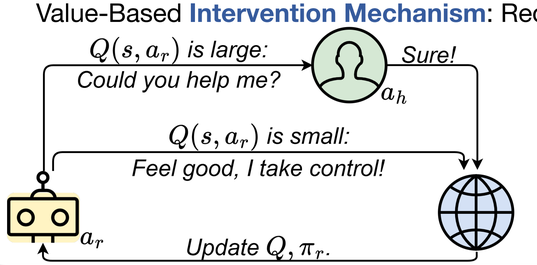
$$Q(s,a_r) \equiv +1$$

For all human actions

$$Q(s,a_h) \equiv -1$$

$$J^{\mathrm{AIM}}(\theta) = \mathop{\mathbb{E}}_{(s,a_r,a_h)} \begin{cases} |Q_\theta(s,a_r) - 1|^2 + |Q_\theta(s,a_h) + 1|^2, & \|a_h - a_r\| > \epsilon \\ |Q_\theta(s,a_h) + 1|^2, & \|a_h - a_r\| \le \epsilon \end{cases}$$

Value-Based **Intervention Mechanism**: Request human help when $Q_\theta(s,a_r) > (1-\delta)$-th quantile



Total objective for Q network

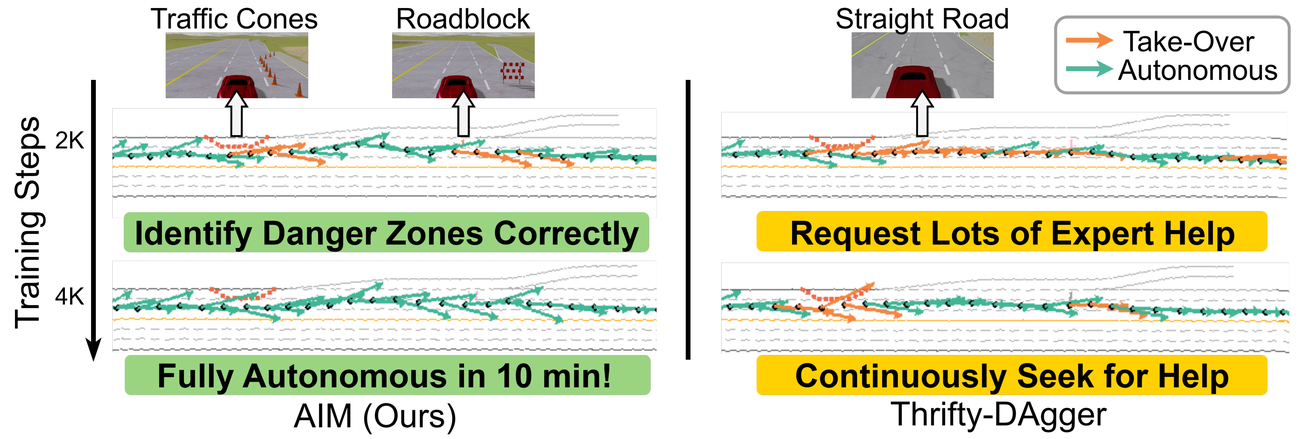
$$J(\theta) = J^{\mathrm{AIM}}(\theta) + J^{\mathrm{TD}}(\theta).$$

BC loss for policy update

$$J(\pi_r) = -\mathop{\mathbb{E}}_{(s,a_h)} \|\pi_r(s) - a_h\|^2.$$

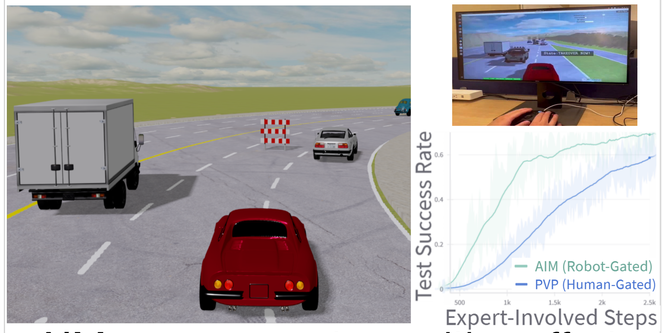
Q_θ - proxy Q network
a_r - agent action
a_h - human action
s - state
π_r - agent policy
δ - switching threshold

Request Less Expert Help Than Robot-Gated Baselines



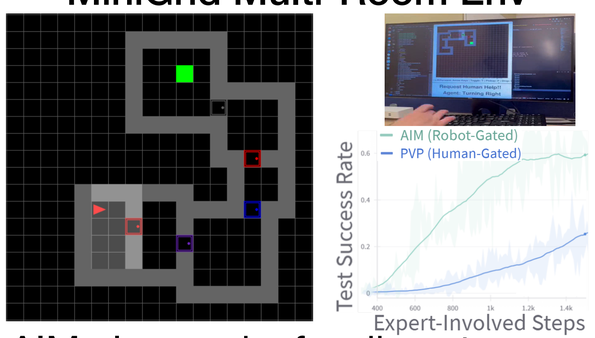
- AIM precisely pinpoints where expert intervention is most necessary.
- AIM adaptively reduces help requests as the agent becomes proficient.

MetaDrive Safety Benchmark



AIM saves expert cognitive effort.

MiniGrid Multi-Room Env



AIM also works for discrete env.

Efficiency & Performance

MetaDrive Env	Training		Testing	
	Expert-Involved Steps	Take-Over Rate	Episodic Return	Success Rate
Neural Expert	-	-	**336.5**	**0.84**
BC	2K	-	243.0	0.33
HG-DAgger	2K	0.45	310.8	0.61
PVP	2K	0.19	270.4	0.62
Ensemble-DAgger	2K	0.55	267.4	0.60
Thrifty-DAgger	2K	0.21	250.0	0.58
AIM (Ours)	**1.9K**	0.24	**328.4**	**0.82**

MiniGrid Env	Training			Testing
	Expert-Involved Steps	Take-Over Rate	Total Steps	Success Rate
Neural Expert	-	-	-	**0.78**
BC	2K	-	-	0.01
HG-DAgger	2K	0.12	-	0.20
PVP	2K	0.12	-	0.34
Ensemble-DAgger	2K	0.36	5.6K	0.38
Thrifty-DAgger	2K	0.27	7.4K	0.42
AIM (Ours)	**0.4K**	**0.09**	**4.4K**	**0.63**

- AIM learns **most performant** agent with **least expert involvement**.
- AIM outperforms both human-gated and robot-gated imitation baselines.

Code is available at metadriverse.github.io/aim

This work was supported by NSF grants CCF-2344955 and IIS-2339769. ZP is supported by the Amazon Fellowship via UCLA Science Hub.