

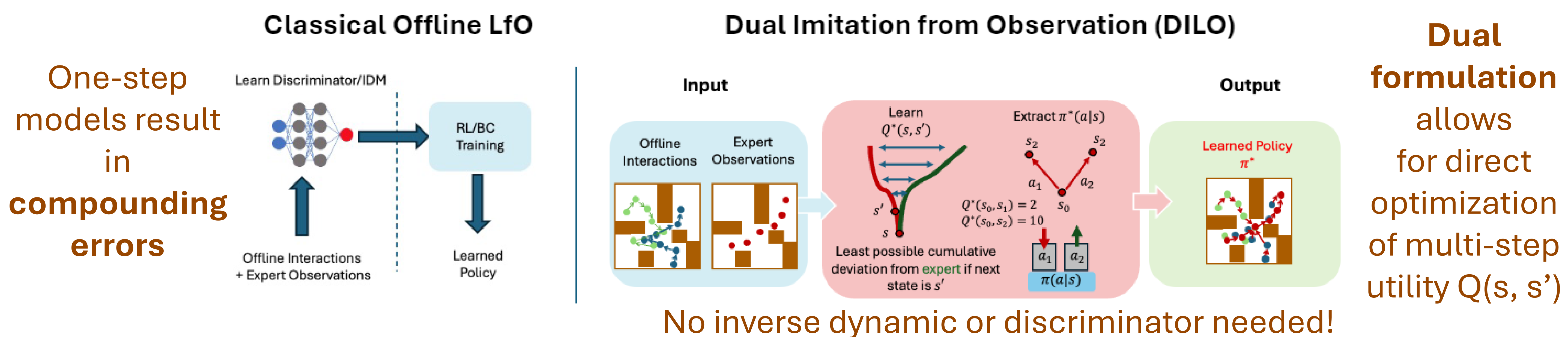
A Dual Approach to Imitation Learning from Observations with Offline Datasets

Harshit Sikchi*, **Caleb Chuck***, Amy Zhang, Scott Niekum

DILO learns to imitate from observations:

- Truly Off-Policy:** Leverage arbitrary prior interaction data
- Lower Errors:** Avoid Compounding Errors
- Stronger Guarantees:** Does not minimize loose upper bounds
- Efficient:** As efficient as vanilla Offline RL

Overview



Approach

Primal Formulation (Mixture Distribution Matching):

$$\max_d - D_f(\text{Mix}_\beta(d, \rho) || \text{Mix}_\beta(d^E, \rho))$$

$$s.t. \sum d(s', s'', a'') = (1 - \gamma)d_0(s', s'') + \gamma \sum d(s, s', a') p(s'' | s', a')$$

Dual Formulation (DILO):

$$\min_V \beta(1 - \gamma) \mathbb{E}_{d_0}[V(s, s')] + \mathbb{E}_{s, s' \sim \text{Mix}(d^E, \rho)} [f_p^*(\gamma V(s', s'') - V(s, s'))]$$

$$- (1 - \beta) \mathbb{E}_{s, s' \sim \rho} [\gamma V(s', s'') - V(s, s')]$$

Simulated Results

Only method to consistently succeed on variety of offline datasets

Suboptimal Dataset	Env	No expert actions			Expert
		ORIL	SMODICE	DILO	
random+expert	hopper	75.21±21.90	100.46±0.64	97.87±8.11	111.33
	halfcheetah	60.49±3.53	85.16±3.62	91.18±0.24	88.83
	walker2d	27.02±23.49	108.41±0.47	108.42±0.64	106.92
	ant	54.19±27.60	122.56±4.47	122.15±5.15	130.75
random+few-expert	hopper	29.86±22.60	78.80±3.09	93.73±7.59	111.33
	halfcheetah	25.76±9.52	4.10±1.50	52.32±10.72	88.83
	walker2d	3.22±3.29	107.18±1.87	108.42±0.25	106.92
	ant	36.52±9.37	-8.89±39.12	117.50±4.75	130.75

Real Robot Results

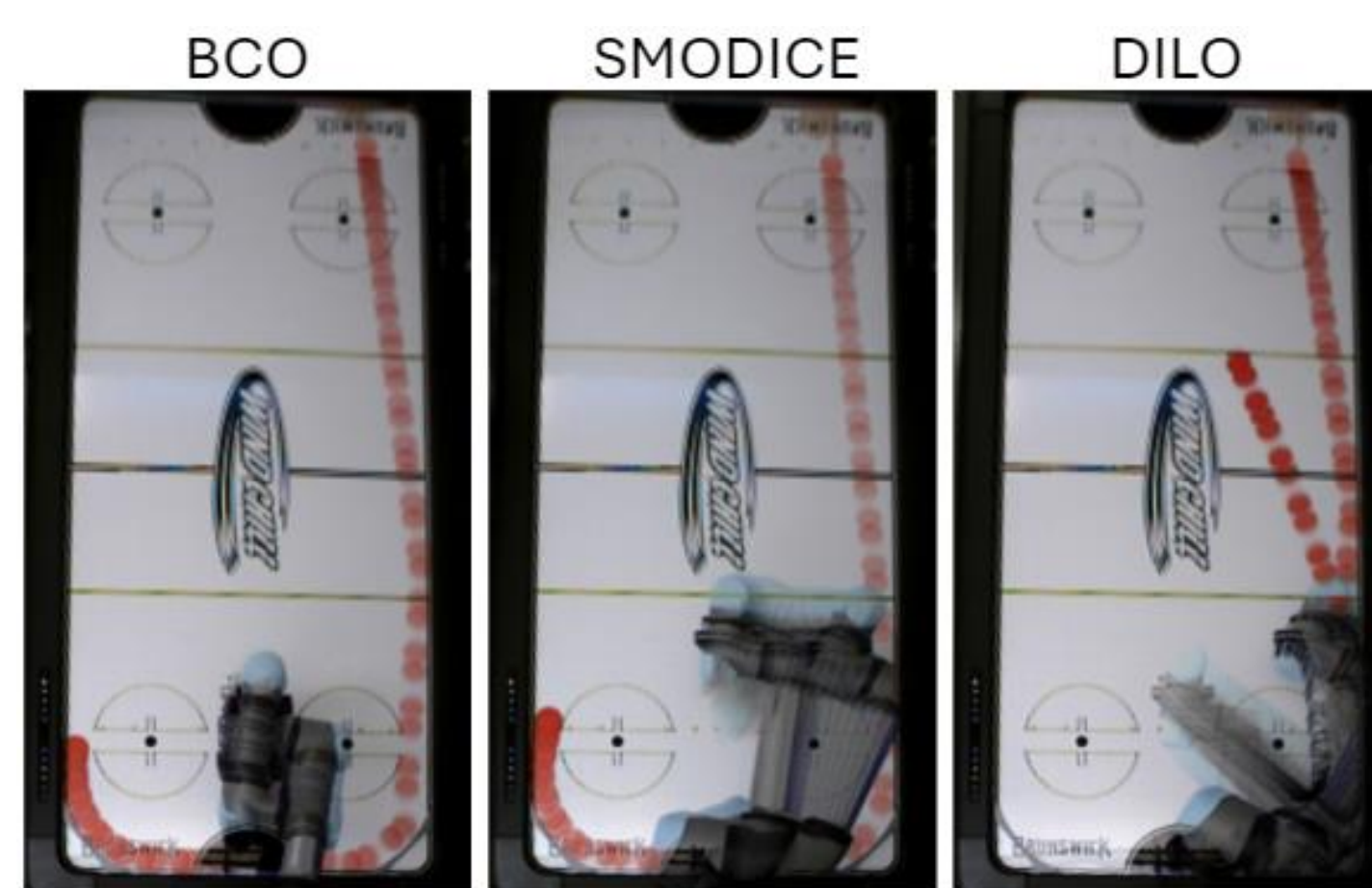
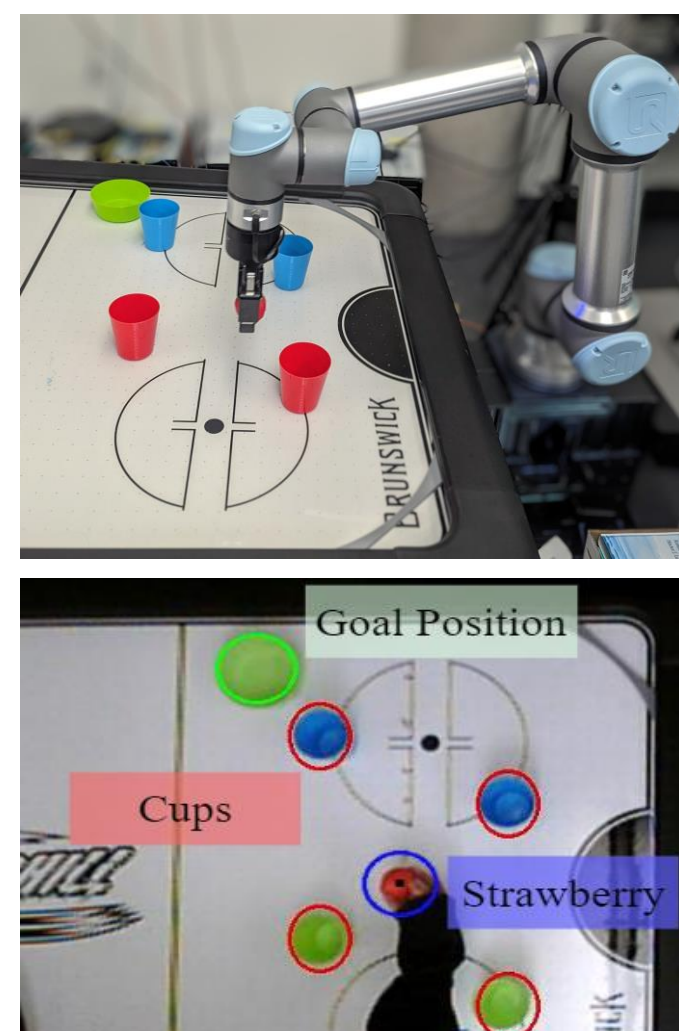
Able to learn from cross-embodied demonstrations
And performs well on complex dynamic tasks

Image Inputs

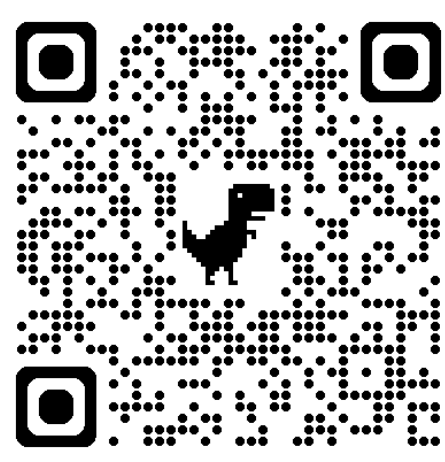
Learns from image observations
With minimal hyperparameter tuning

	Safe Object Manipulation			Puck-Striking	
	Few Trajectories	Fixed Start	Few Uniform	20 expert	10 expert
BCO	3/10	7/10	6/10	7/11	4/11
SMODICE	2/10	1/10	0/10	5/11	4/11
DILO	8/10	9/10	5/10	8/11	5/11

	Safe Object Manipulation (Cross-Embodiment)			Dynamic Puck Hitting	
	Few Trajectories	Fixed Start	Few Uniform	400 expert (touch)	400 expert (hitting)
BCO	6/10	6/10	8/10	2/10	0/10
SMODICE	1/10	1/10	1/10	6/10	4/10
DILO	8/10	9/10	8/10	10/10	9/10



State	50 Demos	Image Inputs			
		Lift-MG	Lift-MH	Can-MG	Can-MH
State	BCO	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
	SMODICE	0.41 ± 0.02	0.46 ± 0.1	0.54 ± 0.01	0.28 ± 0.01
	DILO	0.59 ± 0.03	0.97 ± 0.02	0.53 ± 0.02	0.64 ± 0.03
Image	BCO	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
	SMODICE	0.21 ± 0.02	0.40 ± 0.12	0.10 ± 0.04	0.02 ± 0.01
	DILO	0.76 ± 0.08	0.94 ± 0.02	0.25 ± 0.02	0.15 ± 0.01



Code

Paper

