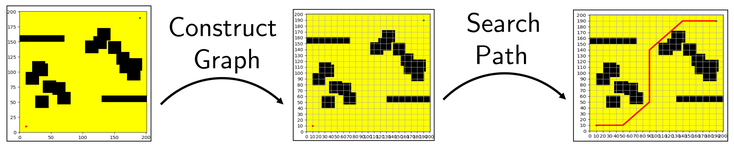
Learning Heuristic Search via Imitation

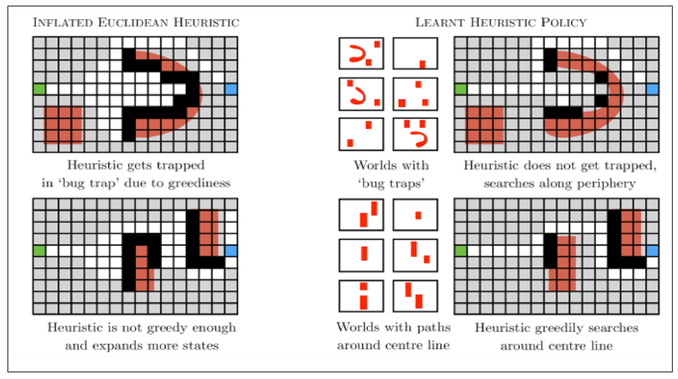
Mohak Bhardwaj, Sanjiban Choudhury, Sebastian Scherer

Motivation

Motion Planning:



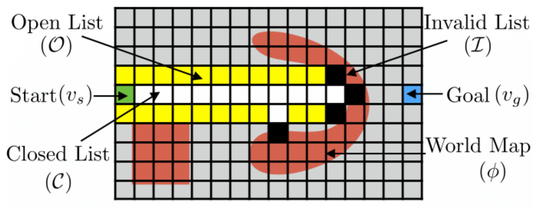
A *heuristic* guides search to minimize number of expansions [1]



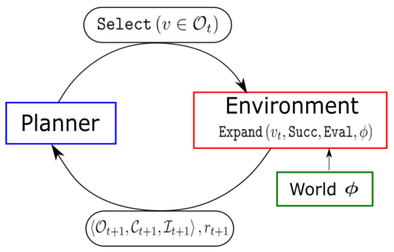
Planning must focus on expected performance on actual distribution using machine learning

Problem Formulation

Recast Search as sequential decision making under uncertainty (over World Map).



Heuristic is a policy from state of search to node to expand



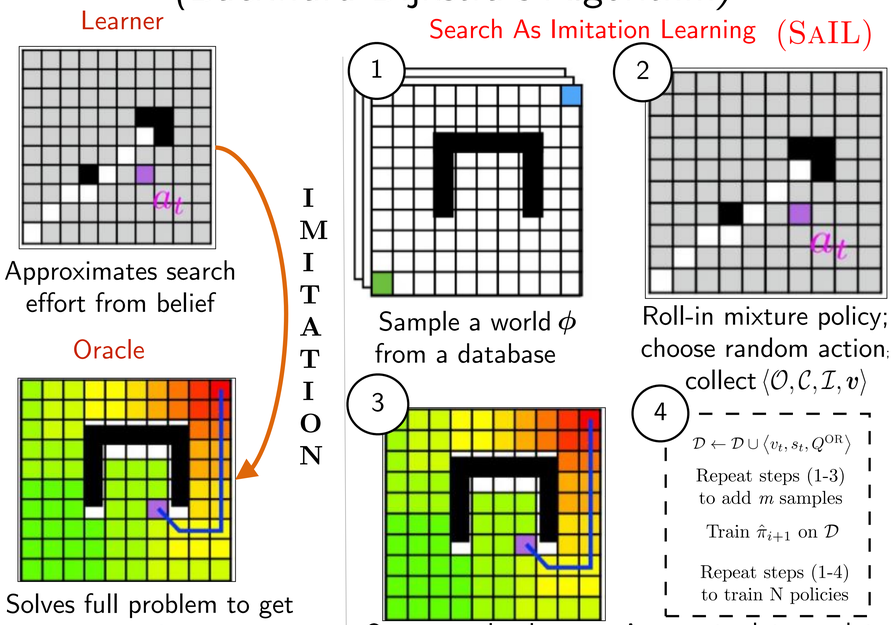
Search MDP

State s_t	$\langle \mathcal{O}_t, \mathcal{C}_t, \mathcal{I}_t \rangle$
Action a_t	Select ($v \in \mathcal{O}_t$)
Reward r_t	0 if $v_g \in \mathcal{O}_t$ -1 otherwise
Transition Model $P(s_{t+1} \mid s_t, a_t)$	Depends on true map ϕ

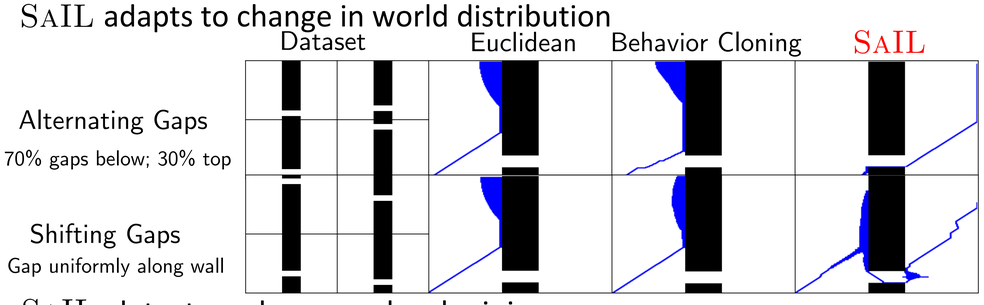
Approach and Algorithm

Imitate clairvoyant oracle planner [2,3]

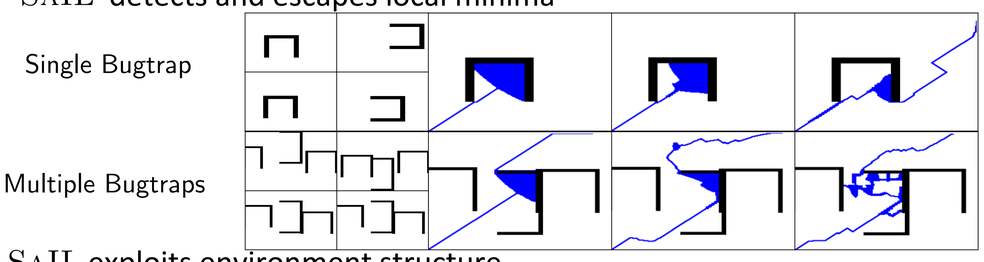
(Backward Dijkstra's Algorithm*)*



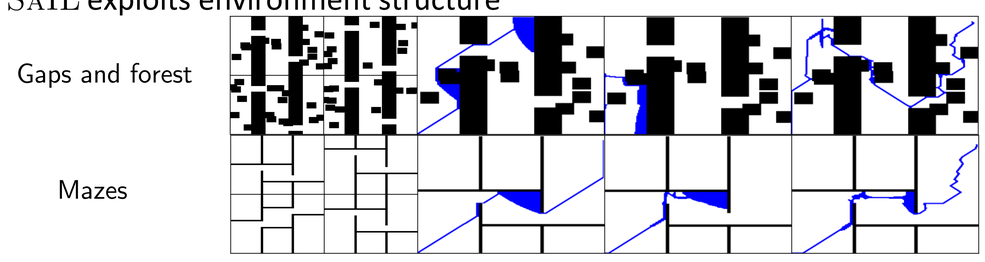
SAIL adapts to change in world distribution



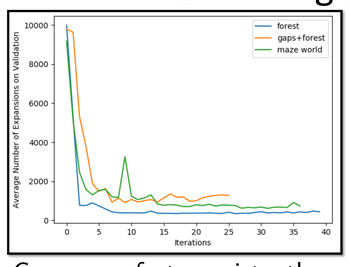
SAIL detects and escapes local minima



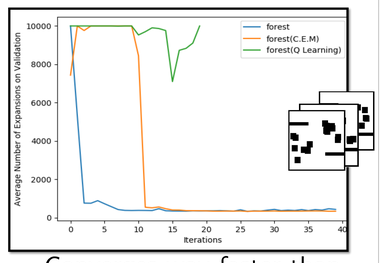
SAIL exploits environment structure



SaIL convergence characteristics



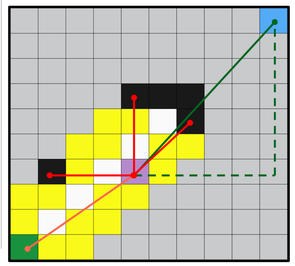
Converges fast consistently across environments



Converges way faster than model free RL

Representing Search State

Compress search state $s_t = \langle \mathcal{O}_t, \mathcal{C}_t, \mathcal{I}_t \rangle$ to get $\boldsymbol{f_t}$ for each $v \in \mathcal{O}_t$



Search based: [x, x_G, g, h_{EUC}, h_{MAN}]

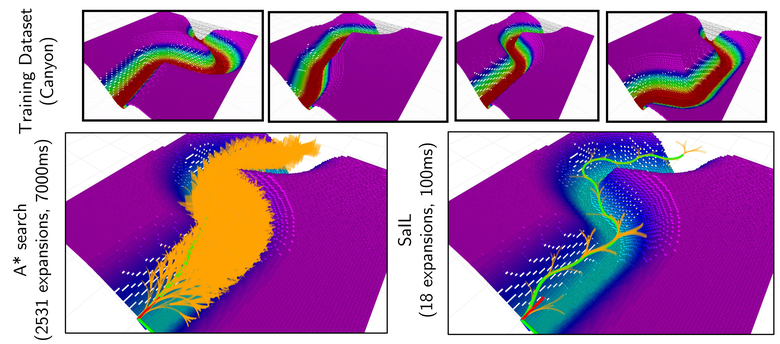
World based: [x_{obs}, d_{obs}, x_{obsx}, d_{obsx}, x_{obsy}, d_{obsy}]

Note: Feature calculation should not expend extra search effort!

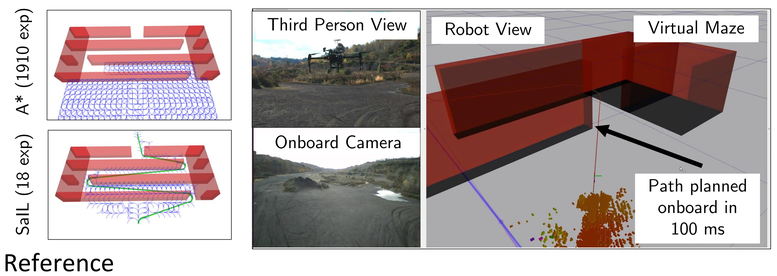
Alternative representations: image patches, distance transforms, number of invalid successors/siblings etc.

Scalability

Helicopter planning in simulation (XYZH with curvature constraints)



Real-world Quadrotor Planning (Trained completely in Sim)



Reference

1. J. Pearl. Heuristics: Intelligent search strategies for computer problem solving, 1984.
2. Ross and Bagnell. "Reinforcement and imitation learning via interactive no-regret learning". arXiv,2014
3. Choudhury, et al. Adaptive information gathering via imitation learning. RSS, 2017.