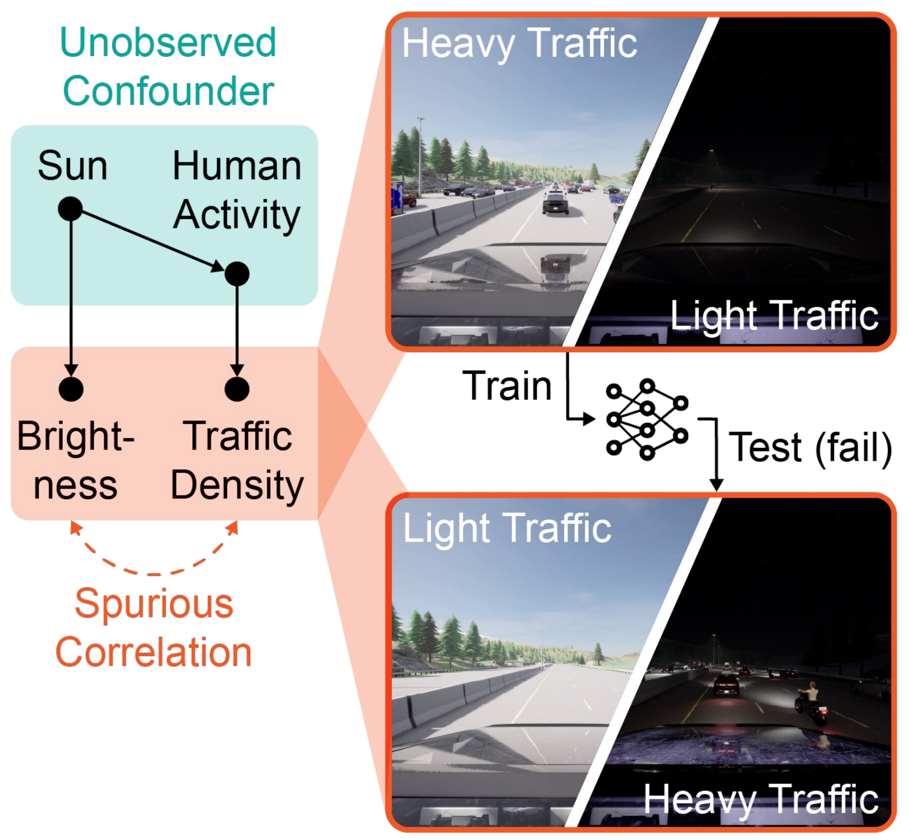
Seeing is not Believing:
Robust Reinforcement Learning against Spurious Correlation

*Wenhao Ding, *Laixi Shi, Yuejie Chi, Ding Zhao

Motivation

▶ **Spurious correlations are ubiquitous in the real world**



In usual cases, there are many cars on the road during the daytime and few cars at night.

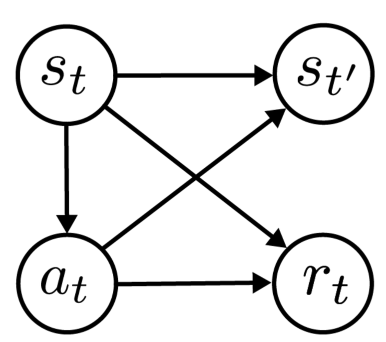
This is mainly caused by the human activity, which is usually not a prior knowledge to autonomous vehicles.

This generates the spurious correlation between traffic density and light.

During testing stage, if this spurious correlation is broken, i.e., a scenario with heavy traffic at night, the autonomous vehicle could fail.

Formulation: State-Confounded MDP

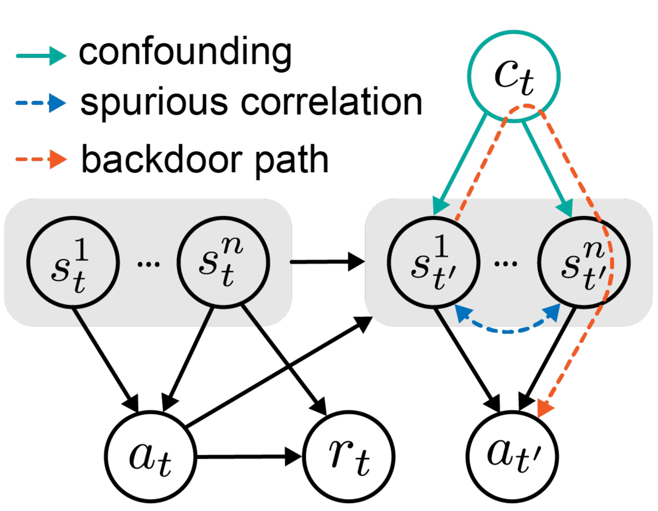
Standard MDP



▶ Standard MDP: the next state $s_{t'}$ is determined by the current state s_t and action a_t:

$$s_{t+1} \sim P_t(\cdot \mid s_t, a_t)$$

Our formulation (SC-MDP)



▶ Our SC-MDP: besides current state s_t and action a_t, there are additional variables -- unobserved confounders -- that determine the next state $s_{t'}$:

$$s^i_{t+1} \sim \mathcal{P}^i_t(\cdot \mid s_t, a_t, c_t),$$
$$\forall i \in 1, 2, \cdots, n$$

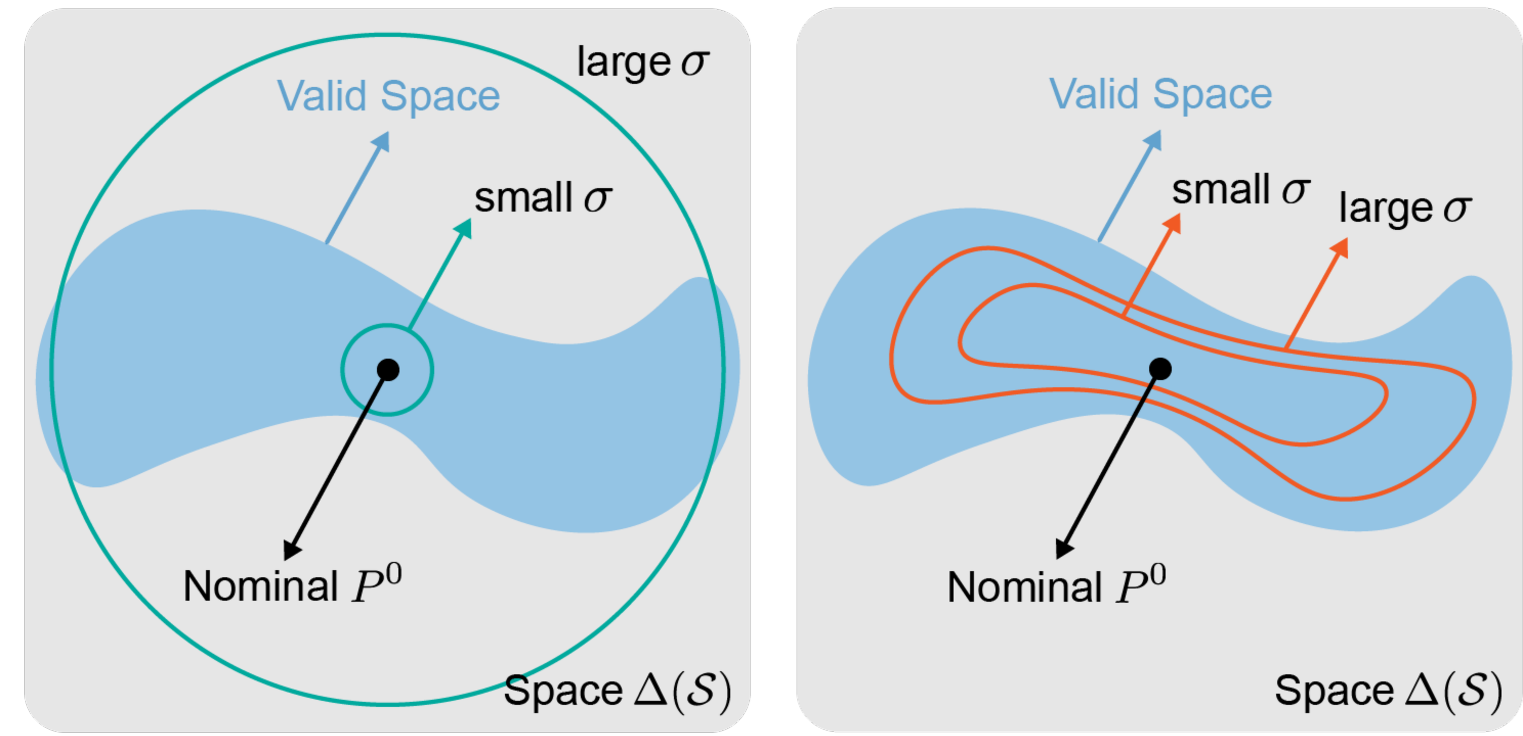
SC-MDP: a nature formulation since the agent can't observe all the importance factors about the real world task.

Our goal: without knowing or observing the confounder, we learn a model with robustness against spurious correlation.

Our method: using the distributionally robust optimization (DRO) framework, we mimic the uncertainty set around unknown confounder resorting to a causal graph.

Robust SC-MDP

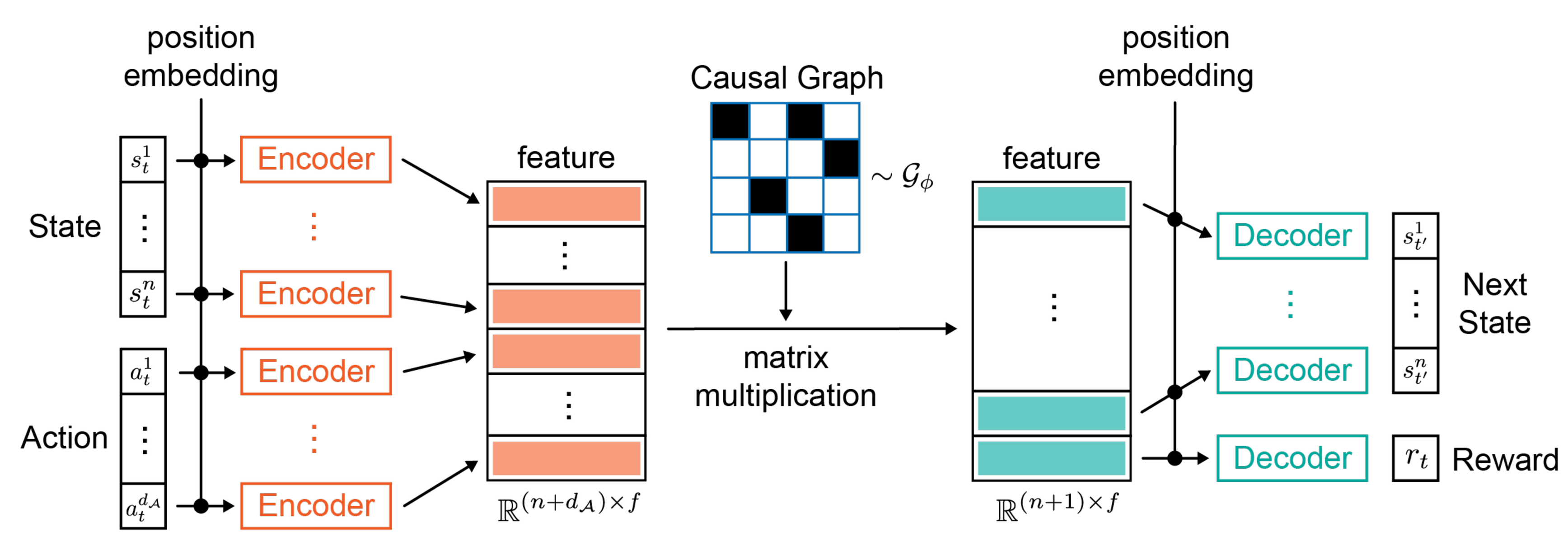
1. Comparison between Robust RL (DRO) and RSC-MDP



DRO: task-agnostic uncertainty set

RSC-MDP: task-specific uncertainty set shaped by causal features

2. Causal Dynamic Model Learning

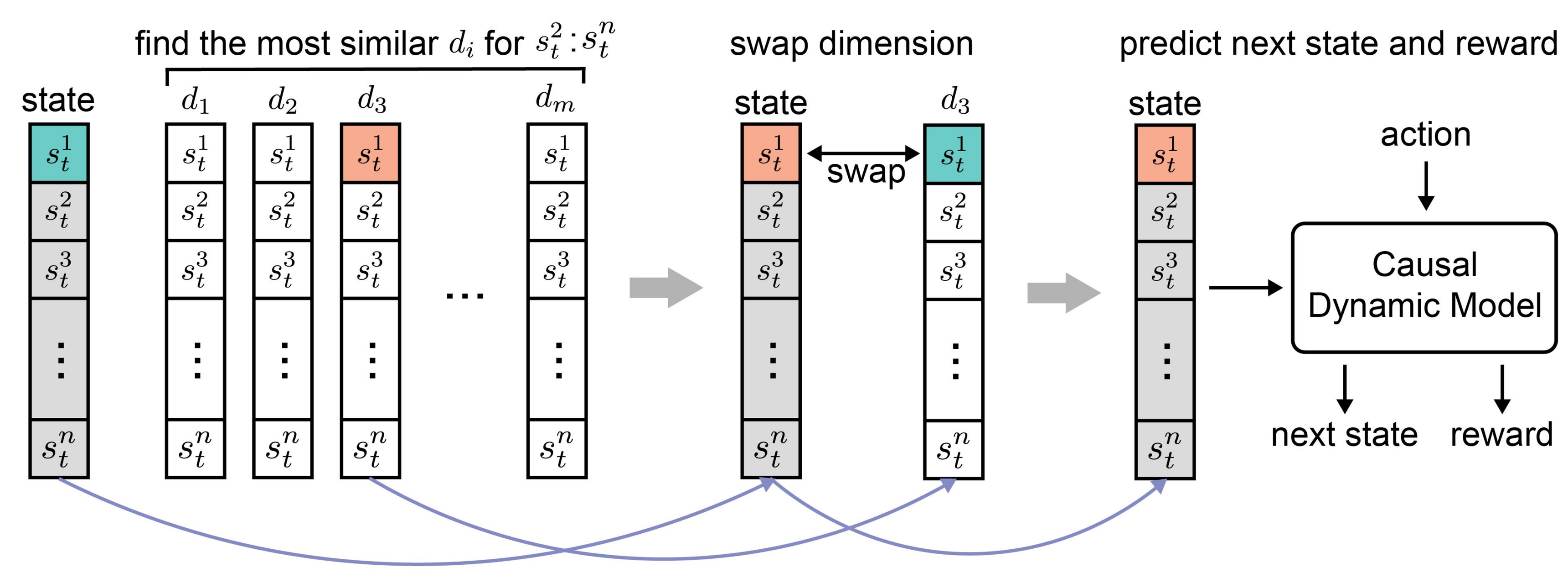


We estimate the causal graph between state + action and next state.

The graph is used to generate new data without spurious correlation.

3. Confounder Perturbation

Empirically, we propose an algorithm based on the SAC agent.

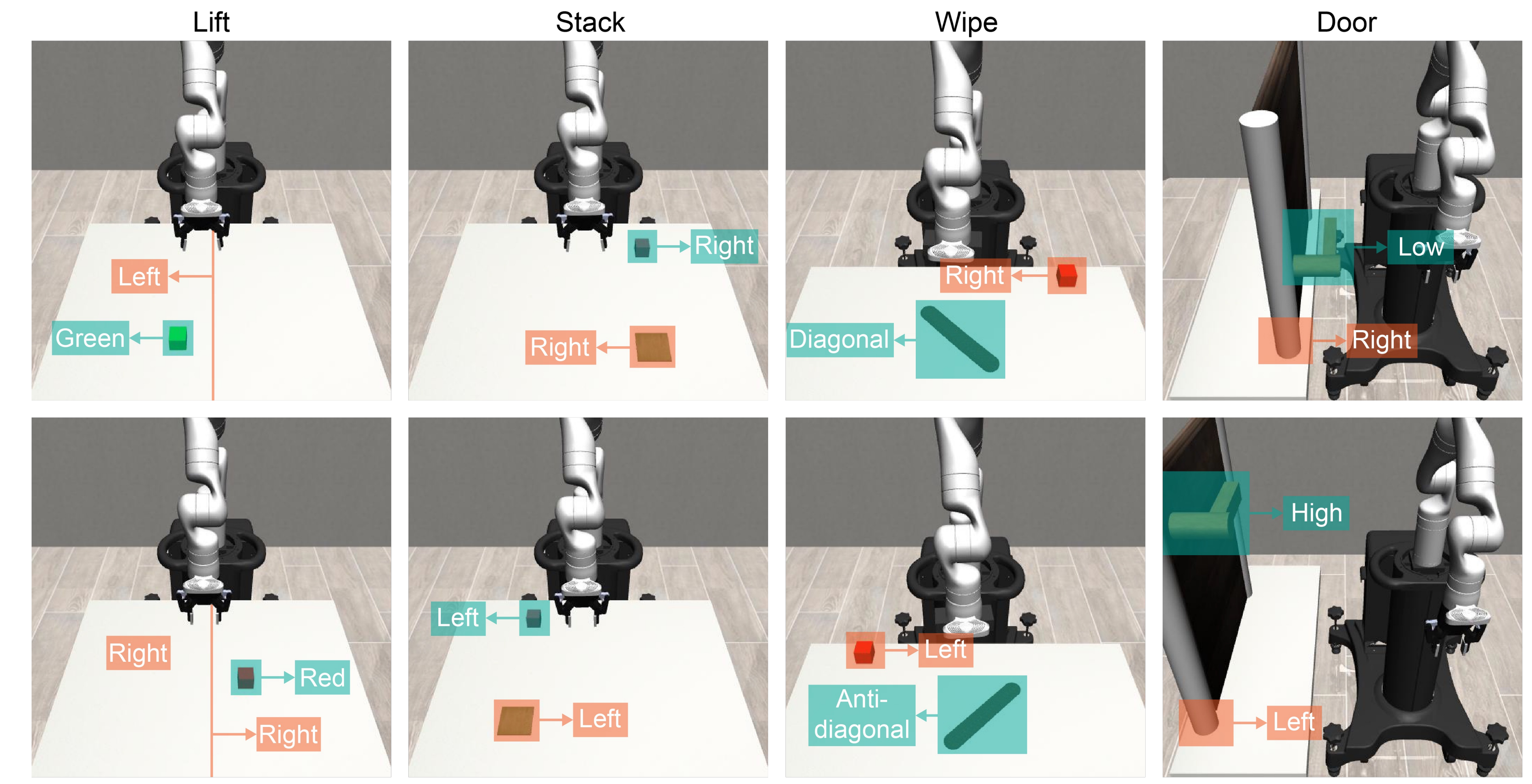


We generate new training data by perturbing some dimensions of the state and use learned dynamic model to generate the next state.

Environment with spurious correlation

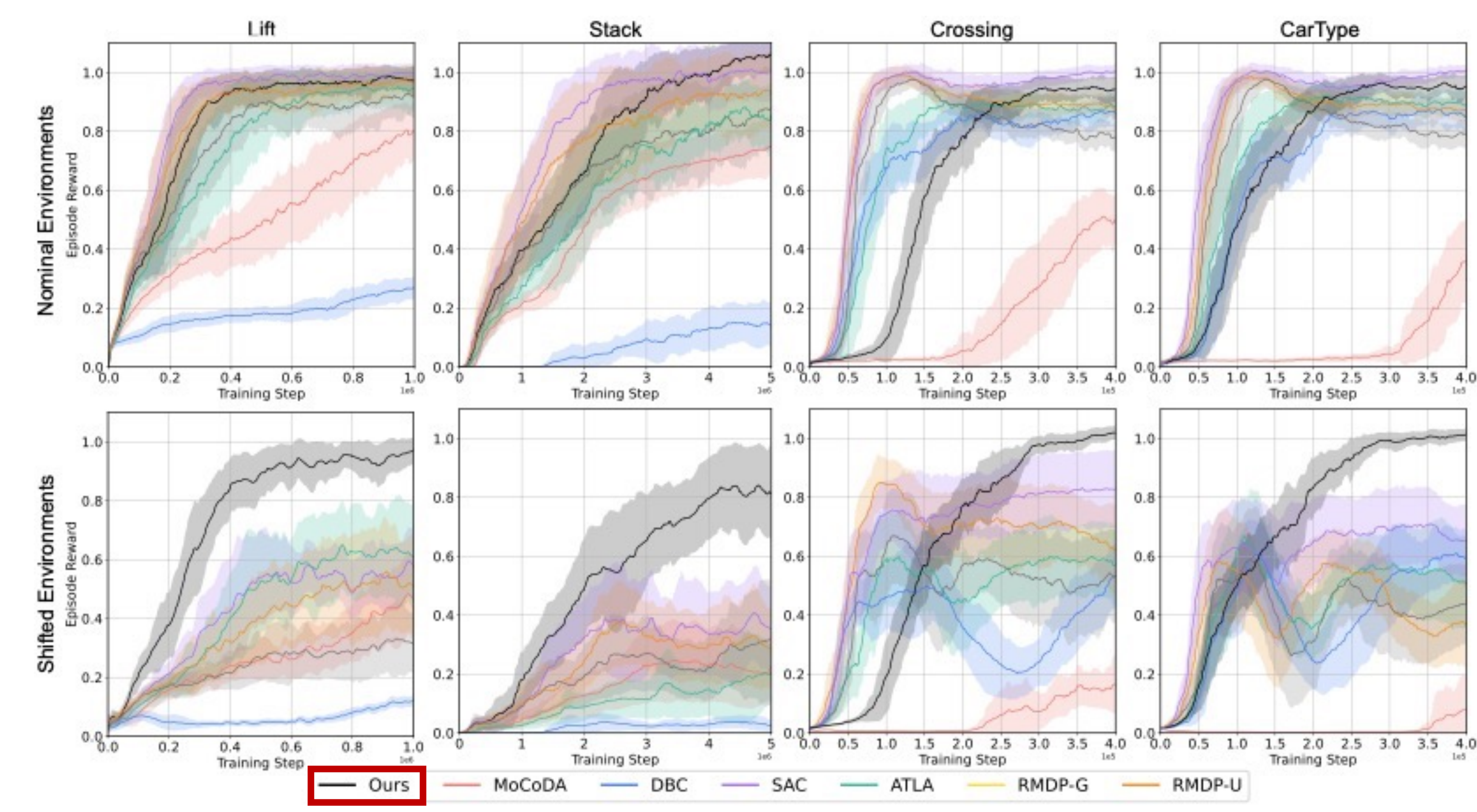
Distraction correlation is between task-relevant and task-irrelevant portions of the state, where the task-irrelevant portion is distraction to the policy

Composition correlation is between two task-relevant portions of the state. Both portions are important for completing the task.



Distraction correlation | Composition correlation | Distraction correlation | Composition correlation

Experimental results



In nominal environments, our algorithm has neglectable drop compared to SAC.
In shifted environments, RSC-SAC is robust to spurious correlations.

Find more results and theoretical analysis in our paper and websites!

wenhaod@andrew.cmu.edu
laixis@caltech.edu