Adaptive World Models: Learning Behaviors by Latent Imagination Under Non-Stationarity

Emiliyan Gospodinov, Vaisakh Shaj, Philipp Becker, Stefan Geyer, Gerhard Neumann

Autonomous Learning Robots Lab

Karlsruhe Institute of Technology

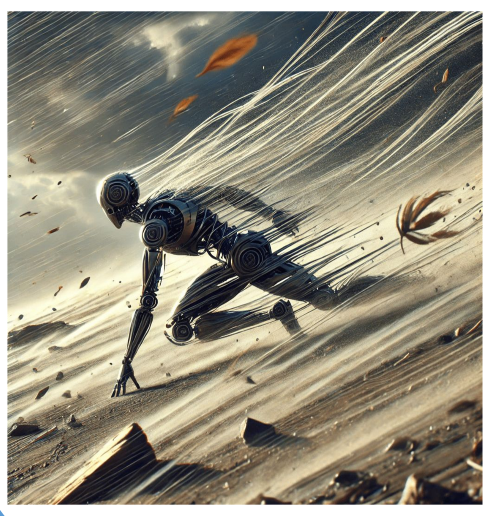
Project Page

Motivation

Dreamer-series world models achieve SOTA-results on **narrow, stationary tasks**

- Can they **model changing environments**?
- Can we use them to **infer adaptive behaviors**?

Dynamics changes: Wind Friction



Dynamics changes: Mass and inertia

Objective changes: Multiple Skills

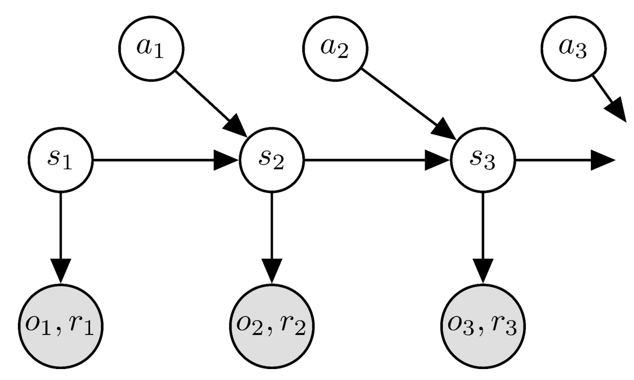
Non-Stationary RL Formalisms

POMDP:

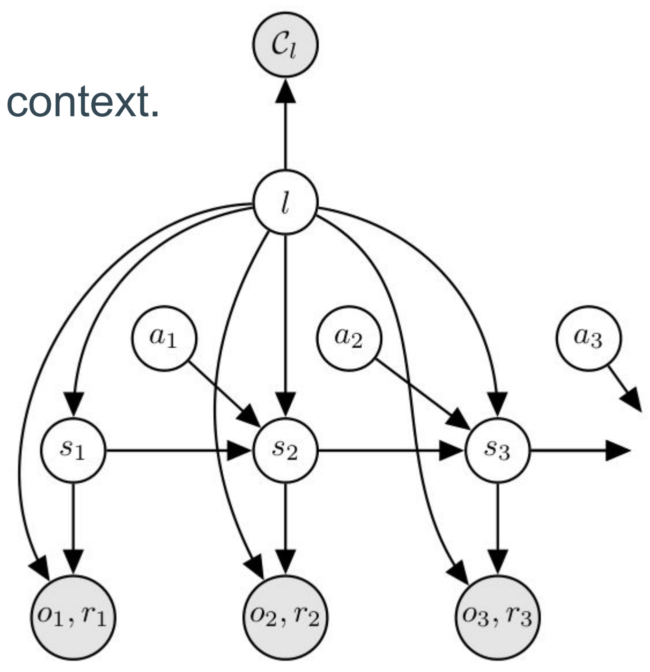
- **Assumption:** Environment is stationary, changes arise due to missing information.
- **Problem**: Joint encoding of state and task in a single latent variable.

HiP-POMDP:

- **Assumption:** Environmental components evolve over time.
- **Solution:**
  - **Introduce inductive bias**. Separate latent variables for task and state.
  - **Two-stage inference**:
    1) Infer a task representation from data context.
    2) Infer latent state conditioned on task.

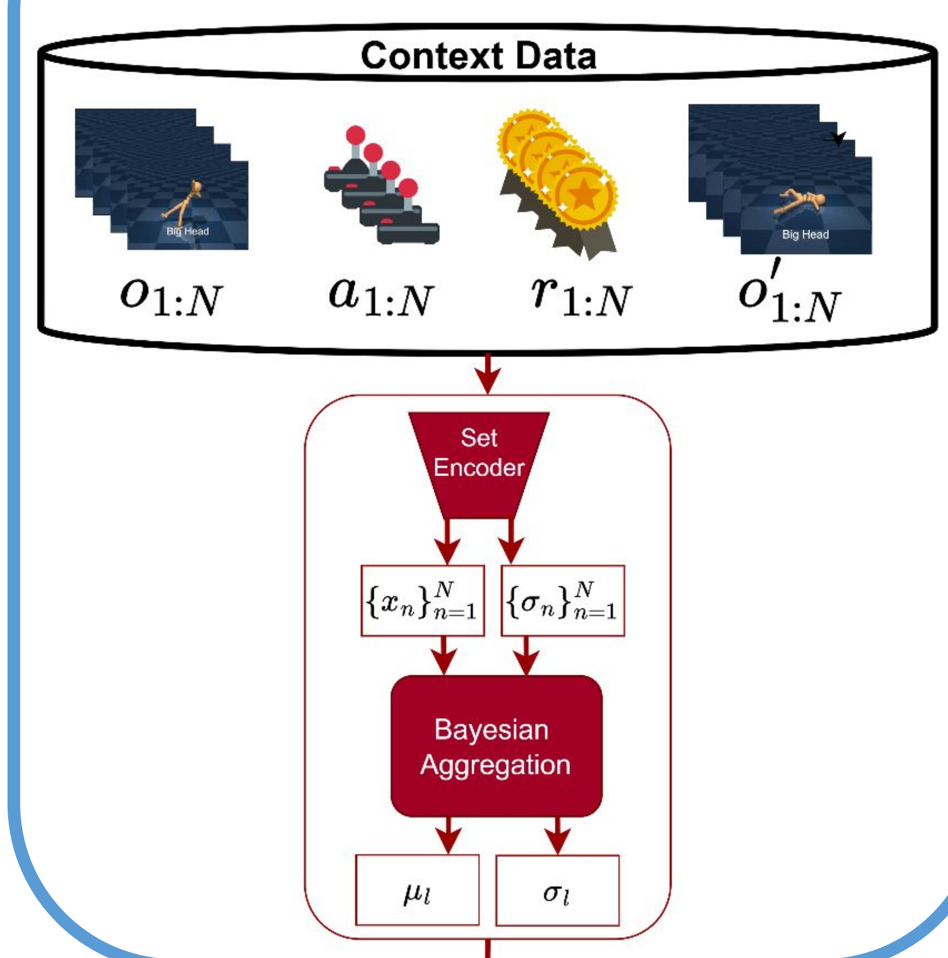


POMDP

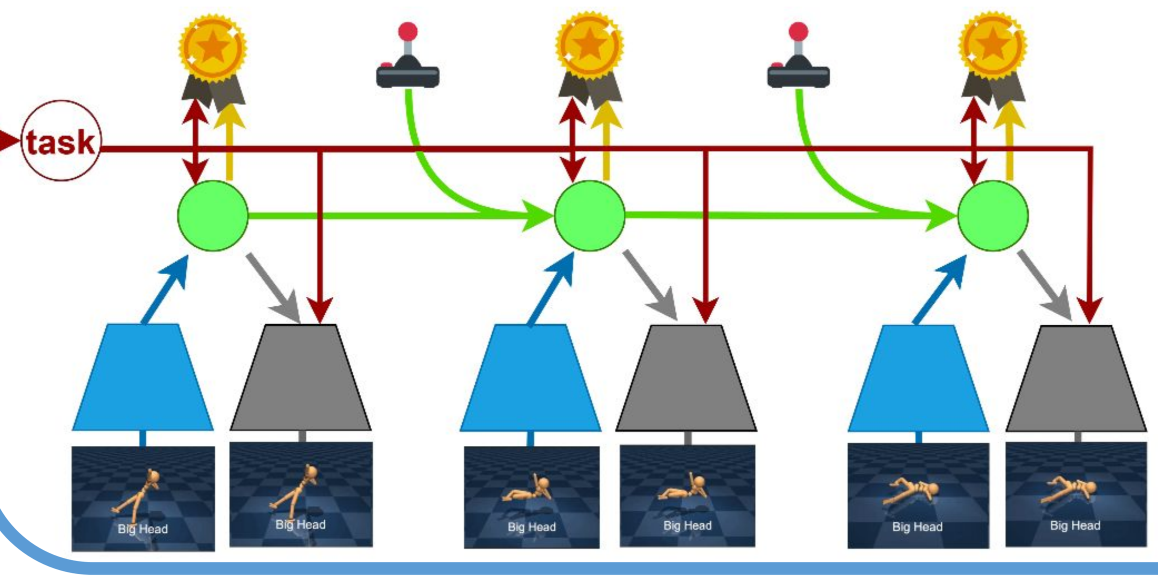


HiP-POMDP

Infer Task Belief



Learning Adaptive Representations

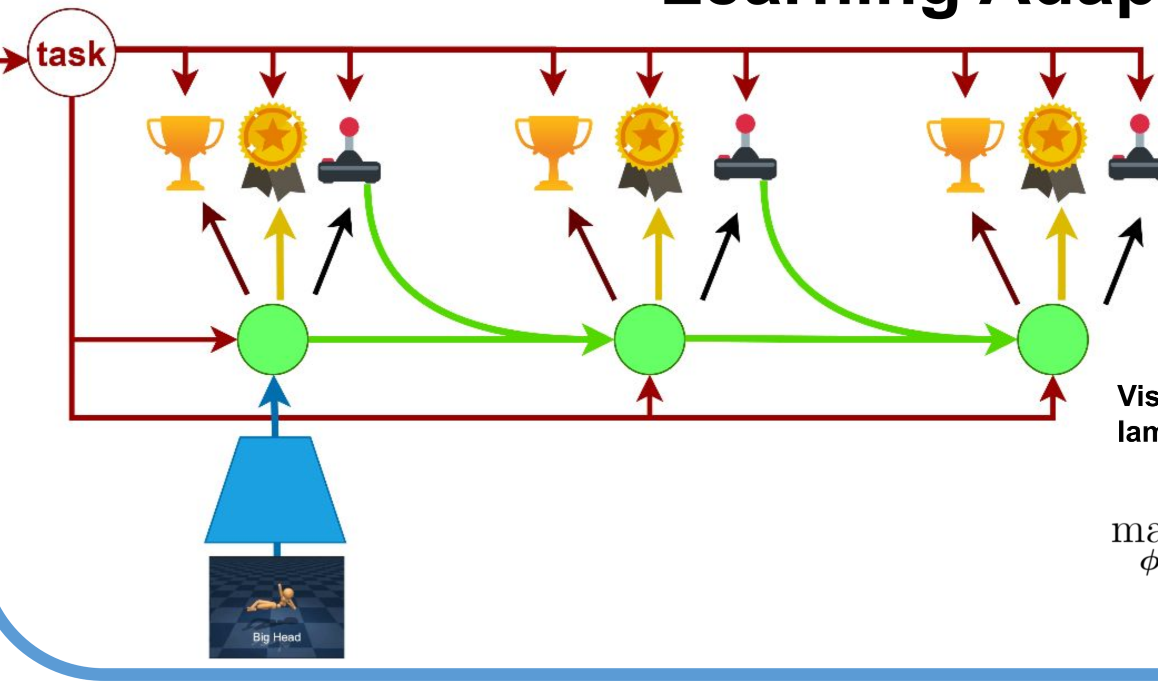


Task-conditioned ELBO:

$$\ln p(\boldsymbol{o}_{1:T}, r_{1:T} \mid \boldsymbol{a}_{1:T}, \mathcal{C}_l) \geq \sum_{t=1}^{T} \underbrace{\mathbb{E}_{p(\boldsymbol{l}|\mathcal{C}_l)q(\boldsymbol{s}_t|\boldsymbol{o}_{\leq t},\boldsymbol{a}_{<t},\boldsymbol{l})}\left[\ln p(\boldsymbol{o}_t, r_t \mid \boldsymbol{s}_t, \boldsymbol{l})\right]}_{\text{Reconstruction Term}}$$

$$+ \underbrace{\mathbb{E}_{p(\boldsymbol{l}|\mathcal{C}_l)q(\boldsymbol{s}_{t-1}|\boldsymbol{o}_{\leq t-1},\boldsymbol{a}_{<t-1},\boldsymbol{l})}\left[\mathrm{D}_{\mathrm{KL}}\left(q(\boldsymbol{s}_t \mid \boldsymbol{o}_{\leq t}, \boldsymbol{a}_{<t}, \boldsymbol{l}) \,\|\, p(\boldsymbol{s}_t \mid \boldsymbol{s}_{t-1}, \boldsymbol{a}_{t-1}, \boldsymbol{l})\right)\right]}_{\text{Regularization Term}}$$

Learning Adaptive Behaviors



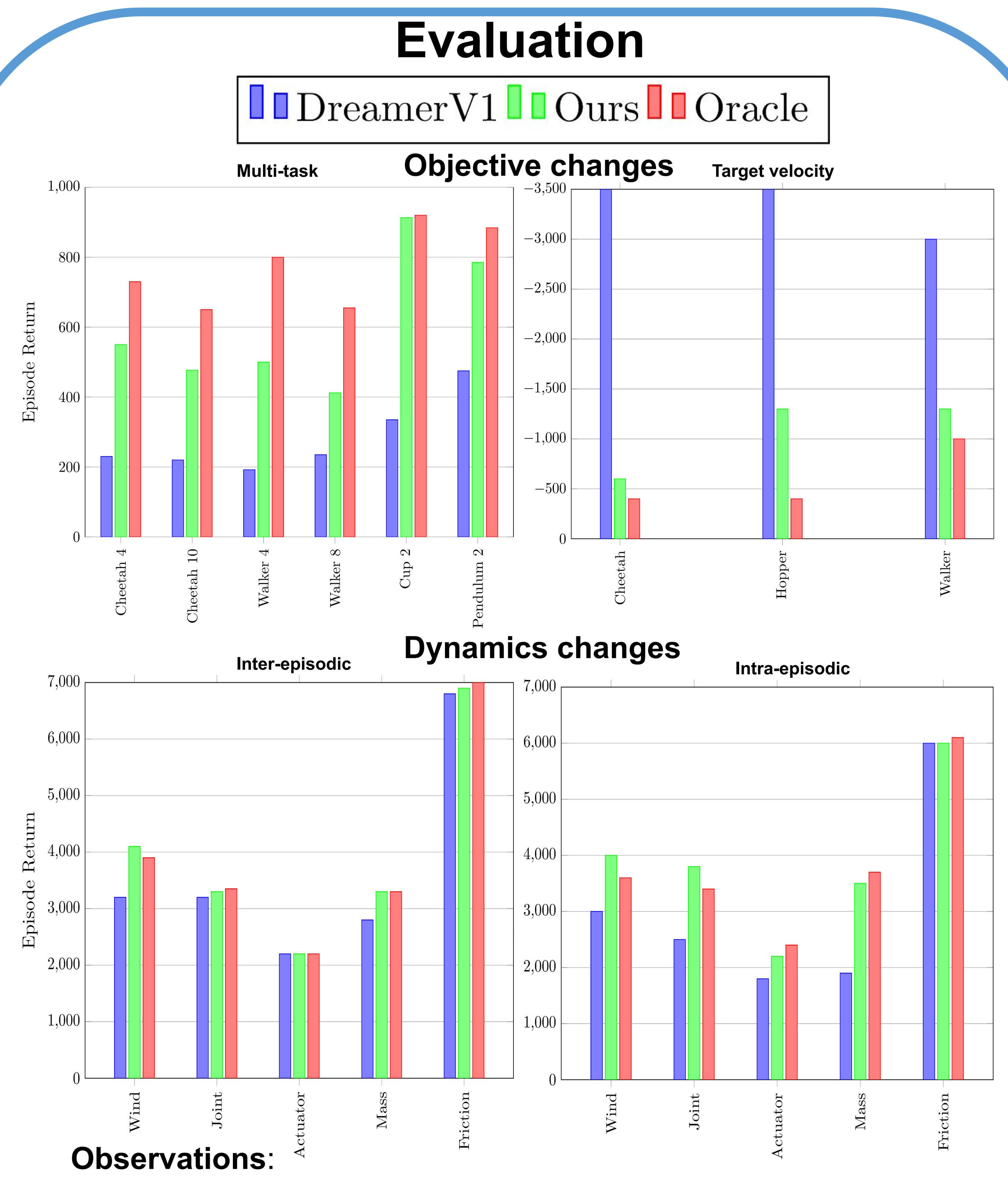
Visit states with high task-conditioned lambda returns:

$$\max_{\phi} \ \mathbb{E}_{q_{\theta},\pi_{\phi}}\left(\sum_{\tau=t}^{t+H} \mathrm{V}_{\lambda}(\boldsymbol{s}_{\tau}, \boldsymbol{l})\right)$$

Bellman consistency: Regress **task-conditioned** lambda returns

$$\min_{\psi} \ \mathbb{E}_{q_{\theta},\pi_{\phi}}\left(\sum_{\tau=t}^{t+H} \frac{1}{2}\left\|v_{\psi}(\boldsymbol{s}_{\tau}, \boldsymbol{l}) - \mathrm{V}_{\lambda}(\boldsymbol{s}_{\tau}, \boldsymbol{l})\right\|^2\right)$$

Evaluation



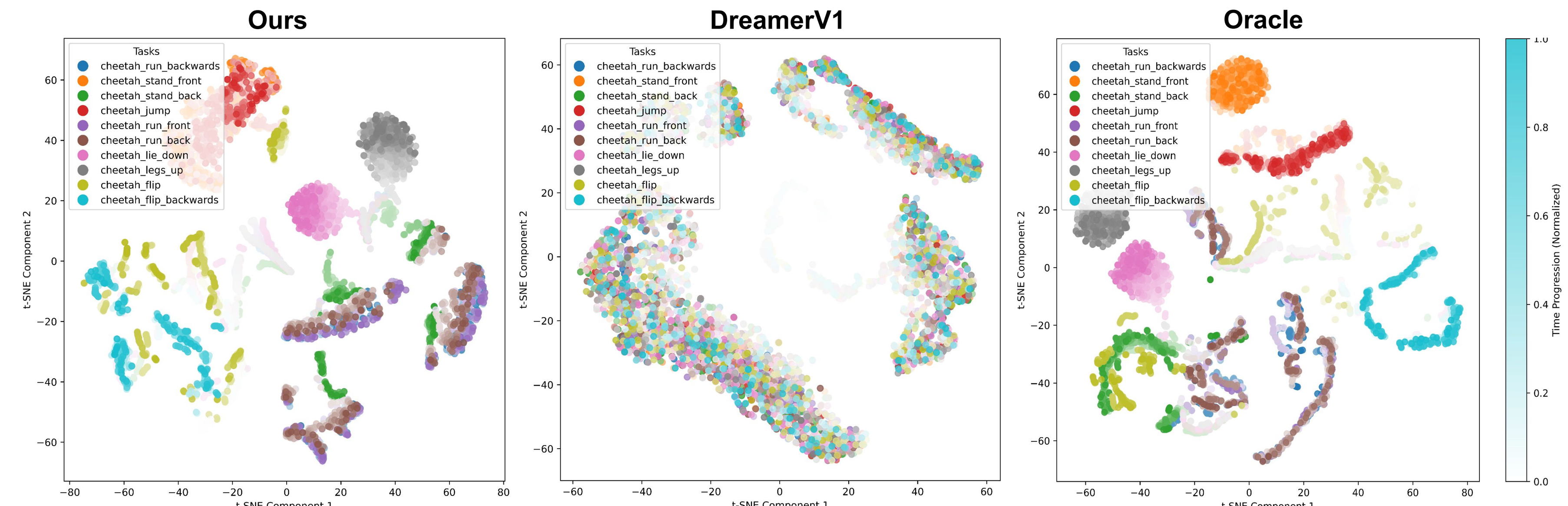
Observations:

1. All agents adapt under dynamics changing scenarios.
2. DreamerV1 fails under all objective changes.

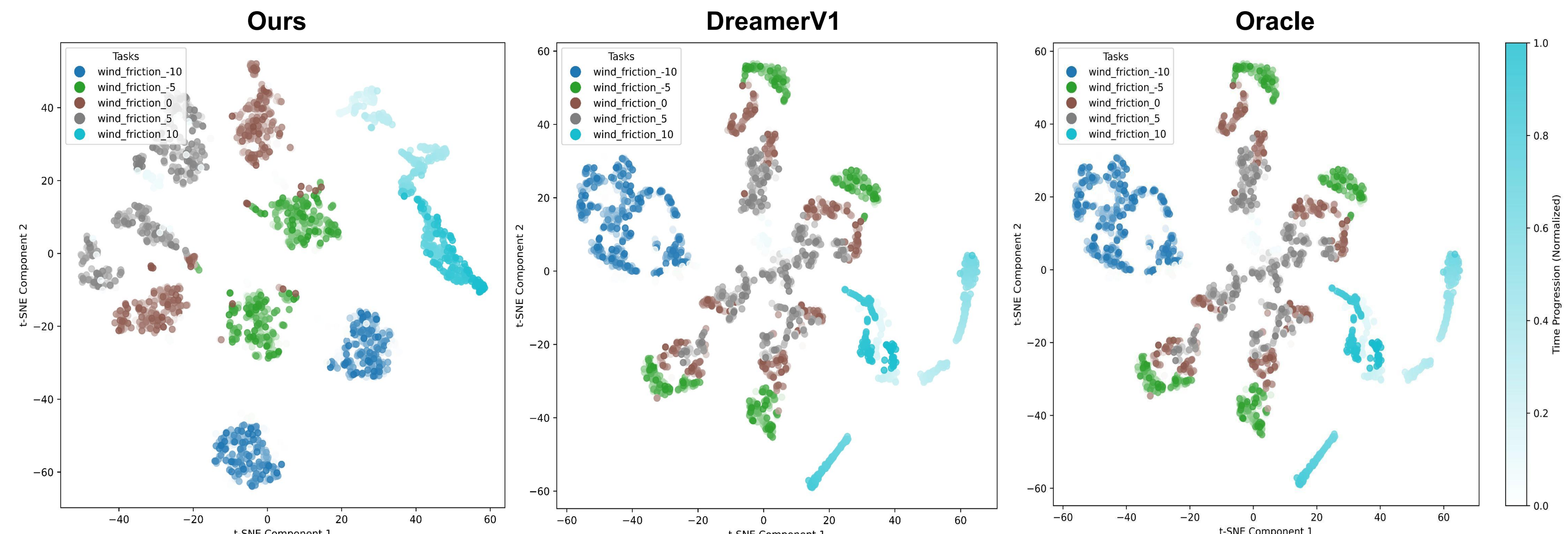
Takeaway: Additional inductive bias aids agent adaptation under all environmental changes.

2D Latent State Space Projections

Objective changes



Dynamics changes



Observations:

1. Latent space is task-aware clustered across all agents under dynamics changes.
2. DreamerV1 fails to organize its latent state space by task under objective changes.

Takeaway: Take-awareness in the latent space improves agent performance.