

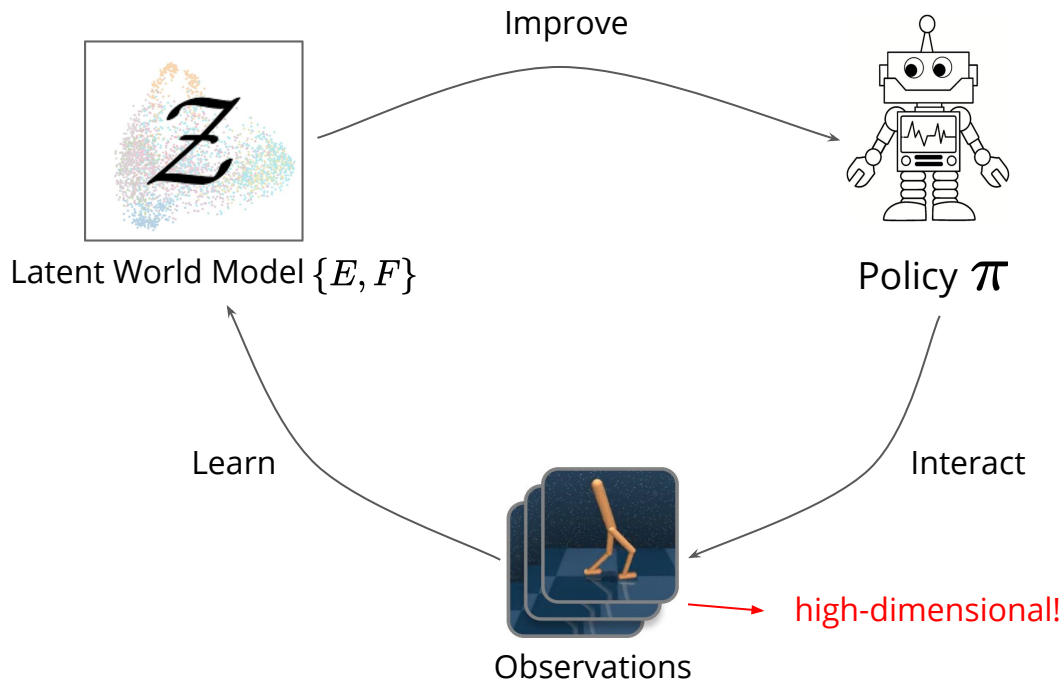
# Temporal Predictive Coding For Model-Based Planning In Latent Space

Tung Nguyen\*, Rui Shu\*, Tuan Pham\*, Hung Bui, Stefano Ermon

---

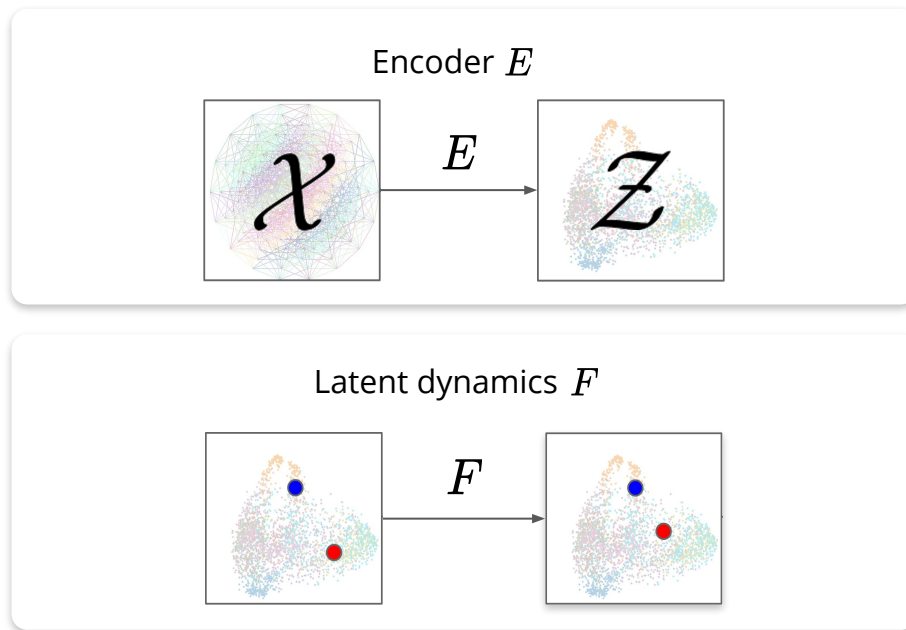


# Model-based planning in Latent Space

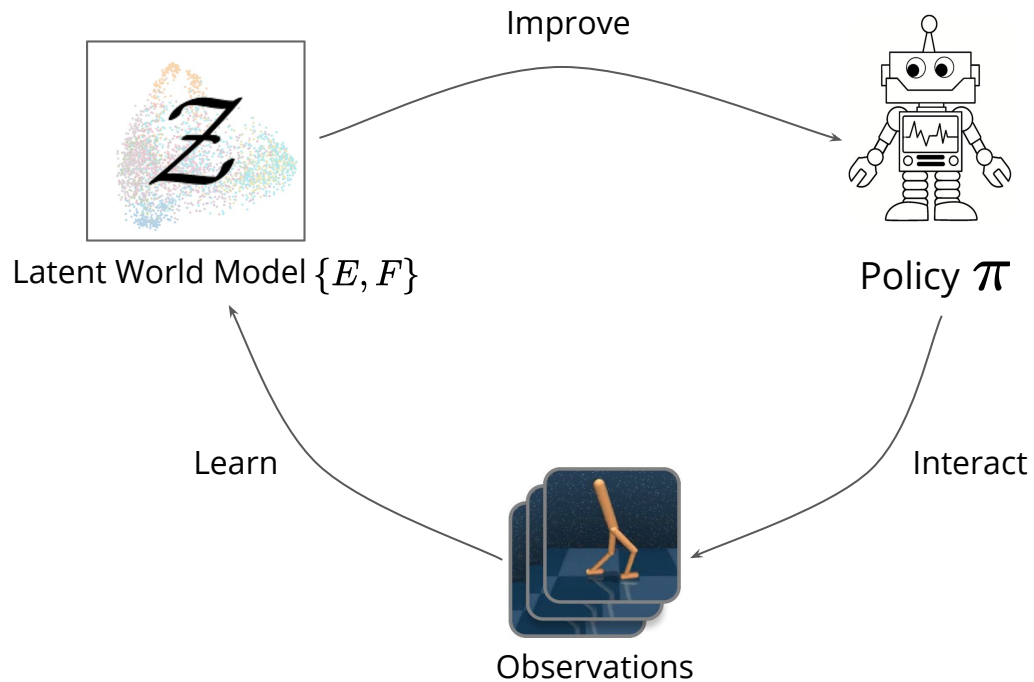


# Model-based planning in Latent Space

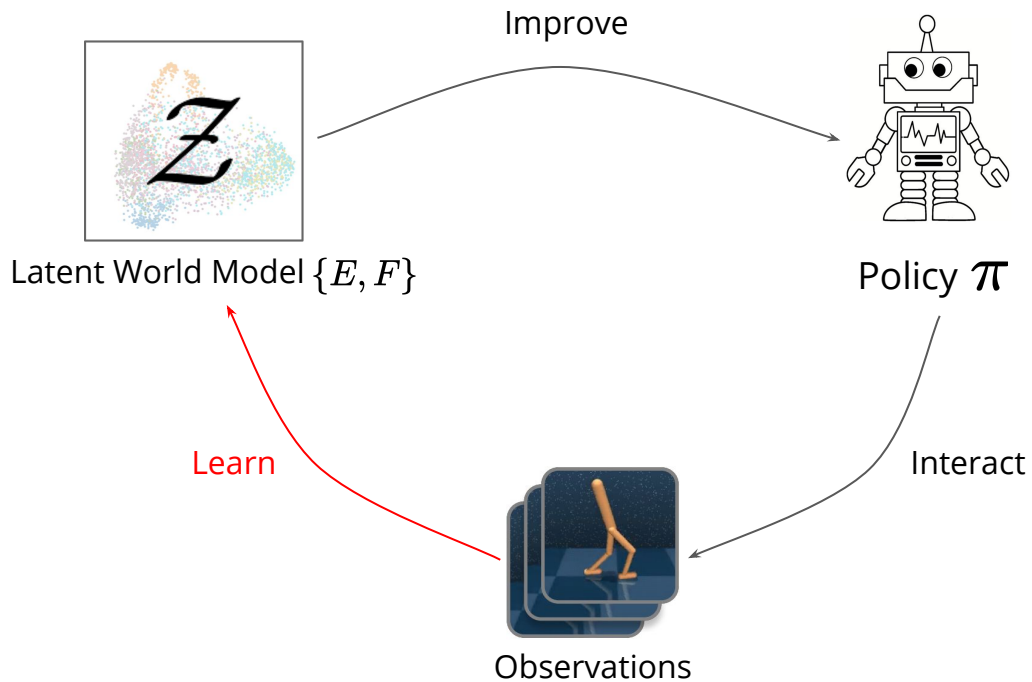
Latent World model



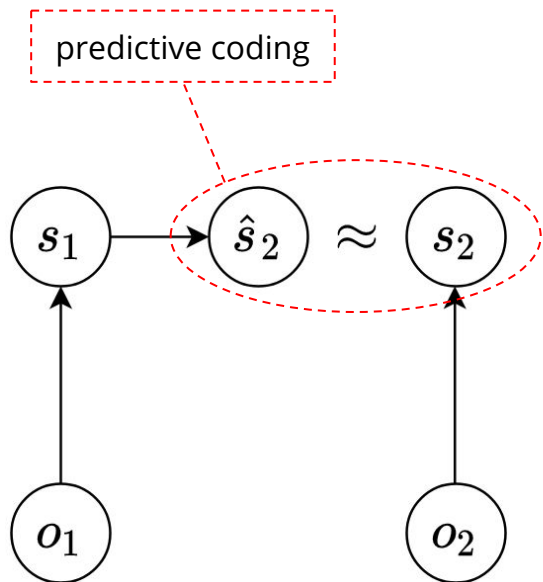
# Model-based planning in Latent Space



# Model-based planning in Latent Space



# Existing frameworks

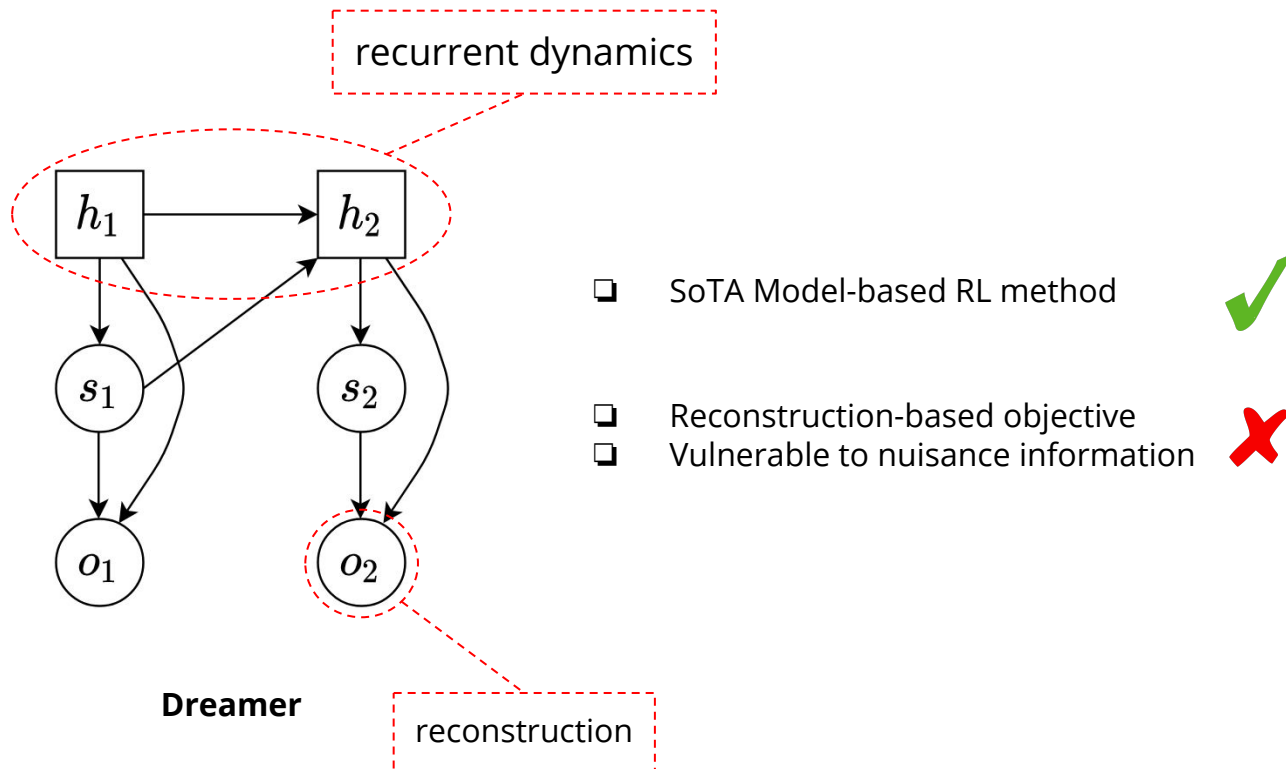


**PC3**

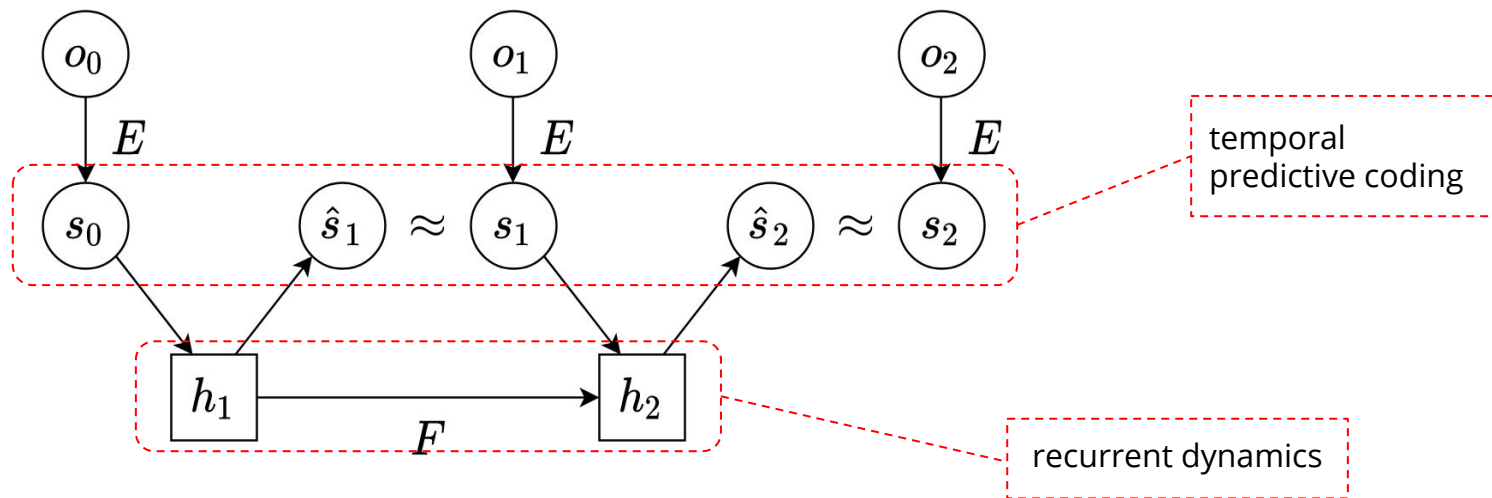
- Superior to reconstruction-based
- Focus on temporally-predictable information
- Optimal-control perspective
- Not applicable for RL problems



# Existing frameworks



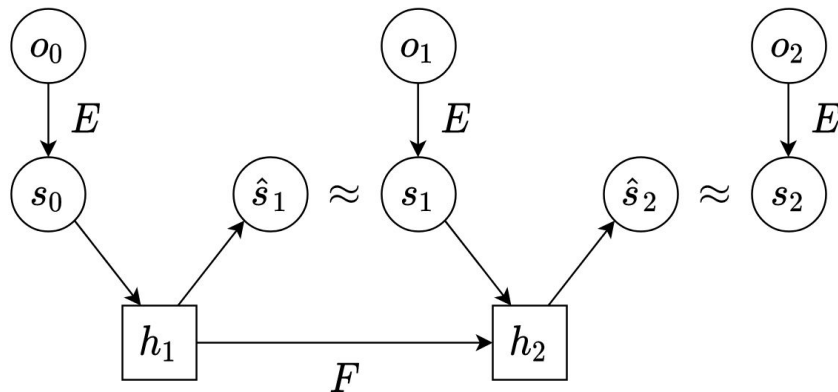
# Temporal Predictive Coding



TPC graphical model

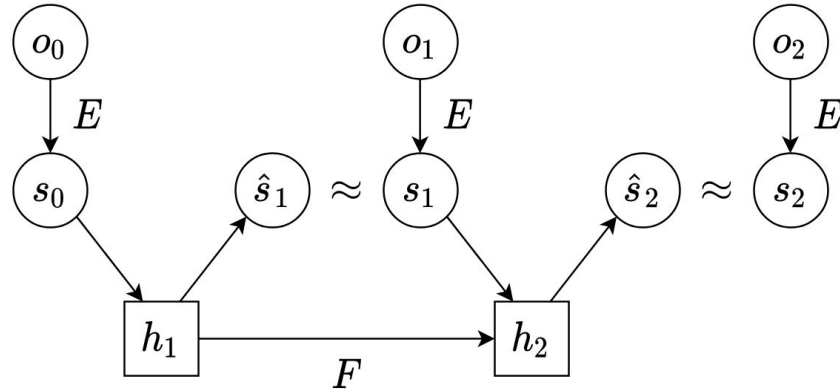


# Temporal Predictive Coding



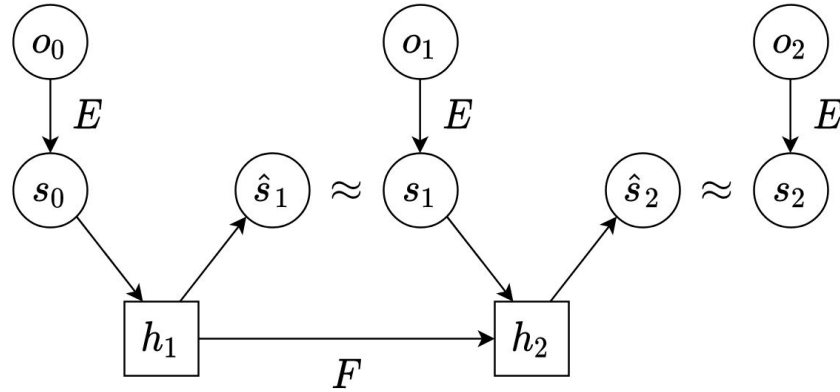
$$\ell_{\text{tpc}}^{(t)}(E, F) = \mathbb{E} \frac{1}{K} \sum_i \ln \frac{F(s_t^{(i)} | s_{<t}^{(i)}, a_{<t}^{(i)})}{\frac{1}{K} \sum_j F(s_t^{(i)} | s_{<t}^{(j)}, a_{<t}^{(j)})} \leq I(E(O_t); E(O_{<t}), A_{<t})$$

# Temporal Predictive Coding



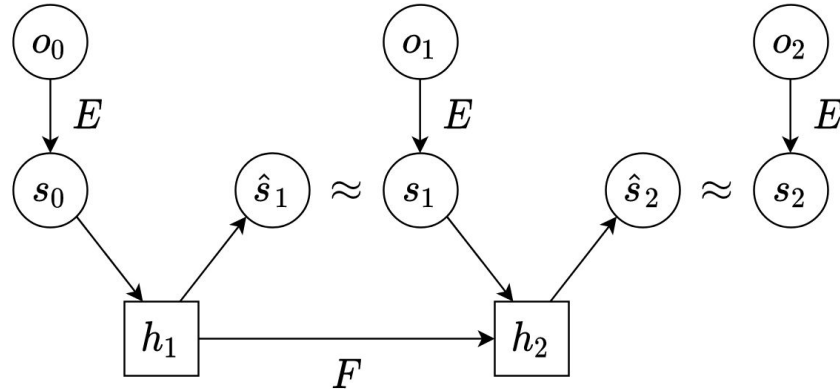
$$\ell_{\text{cons}}^{(t)}(E, F) = \ln F(s_t \mid s_{<t}, a_{<t})$$

# Temporal Predictive Coding



$$\ell_{\text{reward}}(E, R) = \ln R(r_t | s_t)$$

# Temporal Predictive Coding



$$\ell_{\text{overall}} = \max_{E, F, R} \lambda_1 \ell_{\text{tpc}}(E, F) + \lambda_2 \ell_{\text{cons}}(E, F) + \lambda_3 \ell_{\text{reward}}(E, R)$$

# Ignoring Unpredictable Information

**Lemma 1.** Consider an optimal encoder and reward predictor pair  $(E^*, R^*)$  where

$$\operatorname{argmax}_E I(E(O_t); E(O_{<t}), A_{<t}) = E^*$$

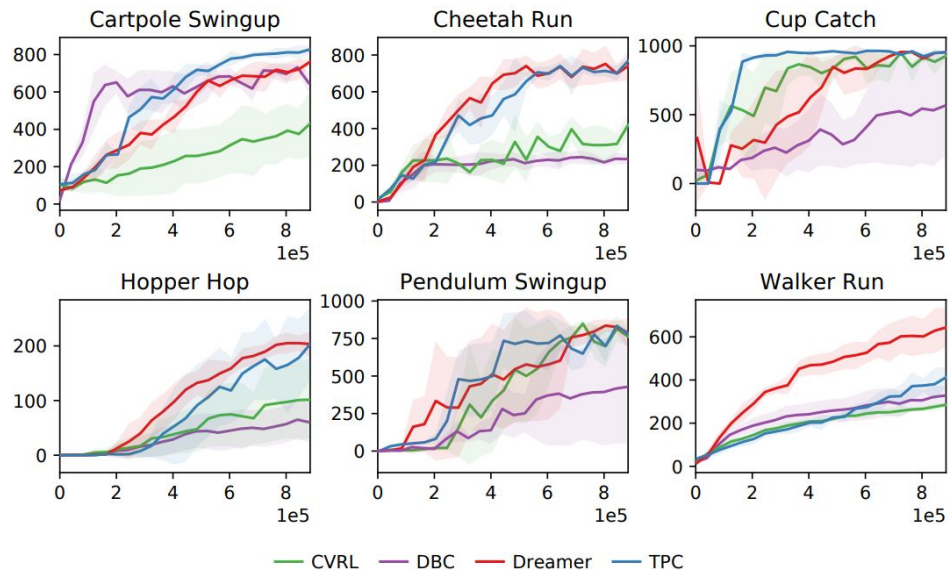
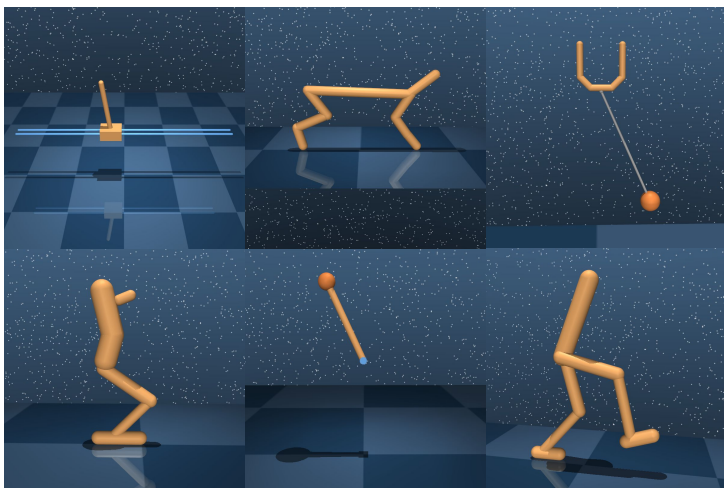
$$D_{KL}(p(r_t | o_t) || R^*(r_t | E^*(o_t))) = 0$$

Let  $\pi(a_t | E^*(o_{\leq t}), a_{<t})$  denote an  $E^*$ -restricted policy whose access to the observations  $O_{<t}$  is restricted by  $E^*$ . Then there exists no encoder  $E'$  where the optimal  $E^*$ -restricted policy underperforms the optimal  $(E^*, E')$ -restricted policy  $\pi_{\text{aux}}(a_t | E^*(o_{\leq t}), E'(o_{\leq t}), a_{<t})$ .

***In other words, any information discarded by an optimal encoder under TPC objective is not helpful for control, or provably task-irrelevant.***

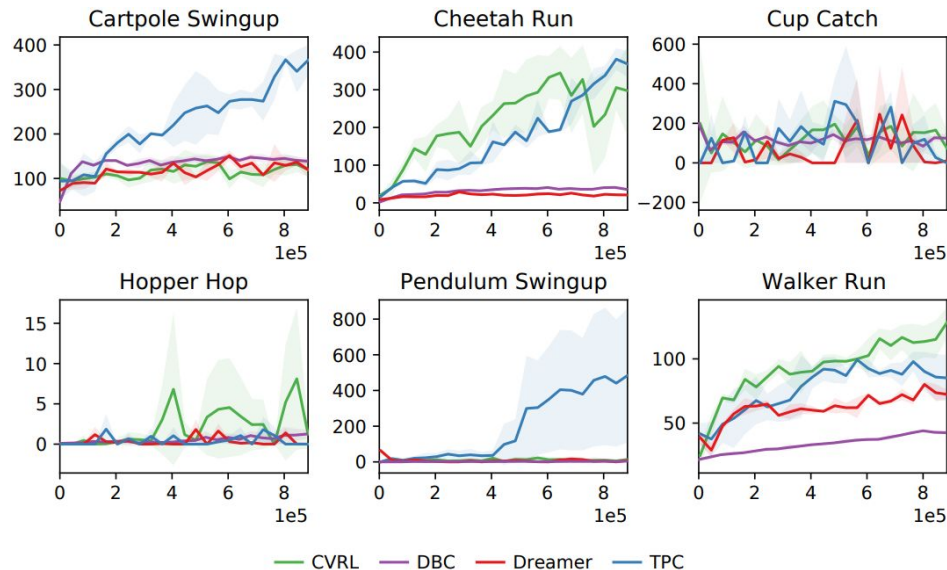
# Experimental results

## □ Standard Deepmind control tasks

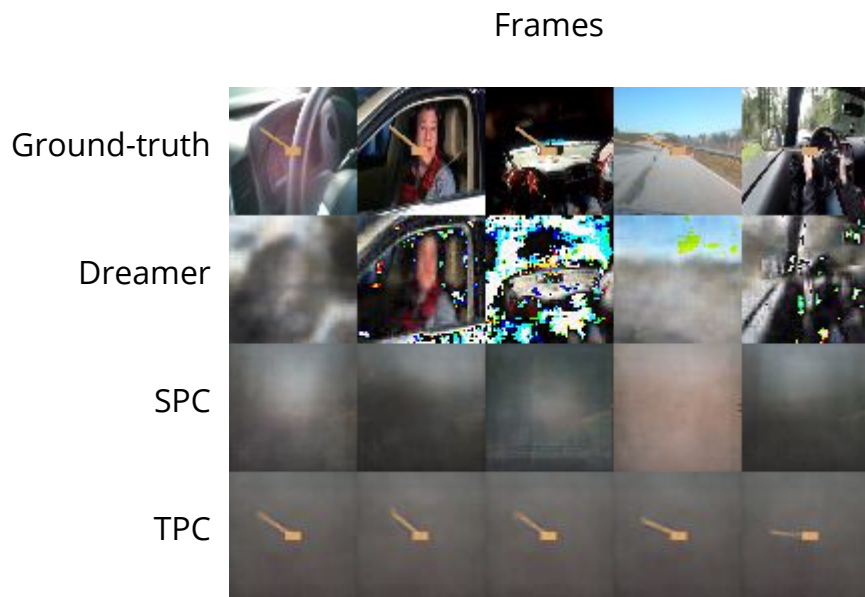


# Experimental results

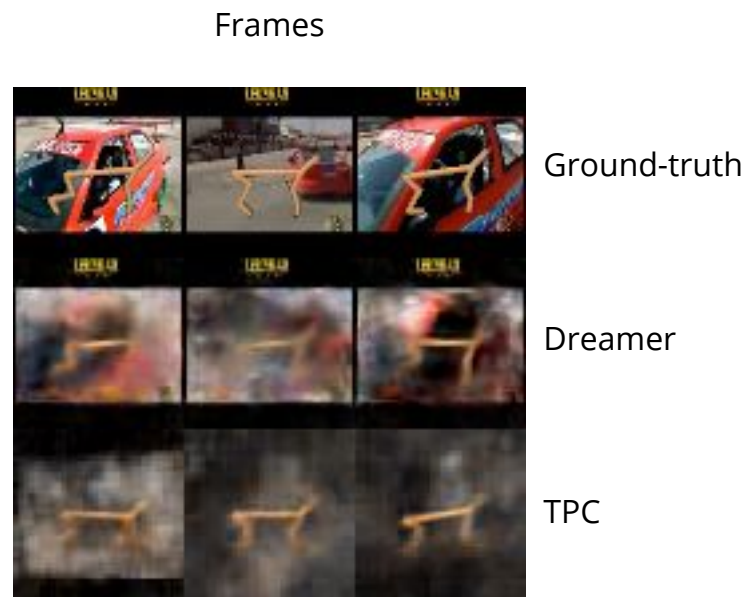
- Deepmind control tasks in natural background settings



# Experimental results



Reconstruction in the  
random background setting



Reconstruction in the  
natural background setting





Paper Link

Tung: [v.tungnd13@vinai.io](mailto:v.tungnd13@vinai.io) @tungnd\_13  
Rui: [ruishu@stanford.edu](mailto:ruishu@stanford.edu) @\_smileyball  
Tuan: [v.tuanpa36@vinai.io](mailto:v.tuanpa36@vinai.io)