ICML
International Conference
On Machine Learning

Reinforcement Learning with Action-Free Pre-Training from Videos

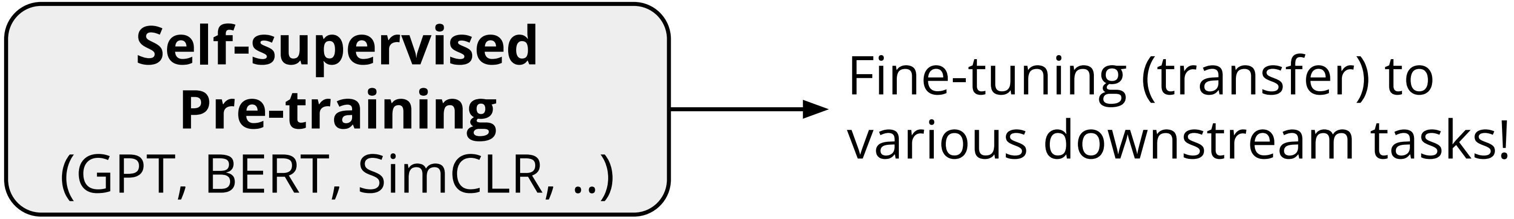
Younggyo Seo, Kimin Lee, Stephen James, Pieter Abbeel

KAIST

THE UNIVERSITY OF CALIFORNIA · BERKELEY · 1868 · LET THERE BE LIGHT

Introduction

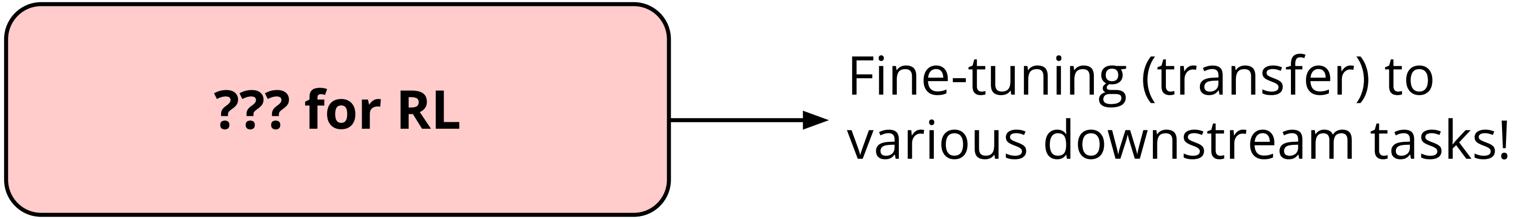
- Unsupervised pre-training is successful in CV / NLP



Self-supervised Pre-training
(GPT, BERT, SimCLR, ..)

Fine-tuning (transfer) to various downstream tasks!

How to do pre-training for RL? 🤷‍♀️ Still an open question!

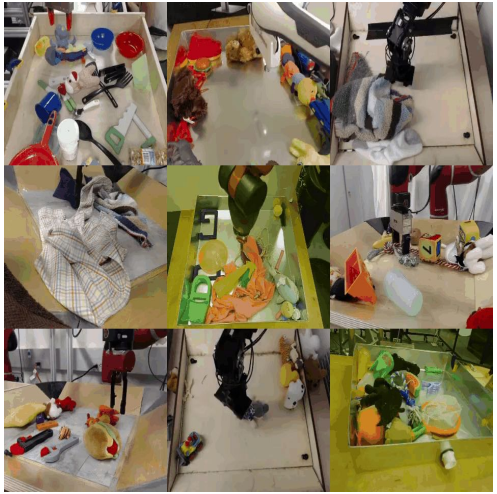


??? for RL

Fine-tuning (transfer) to various downstream tasks!

Pre-training from Diverse & Action-Free Datasets

- **Main idea:** Pre-train representations from **Videos**

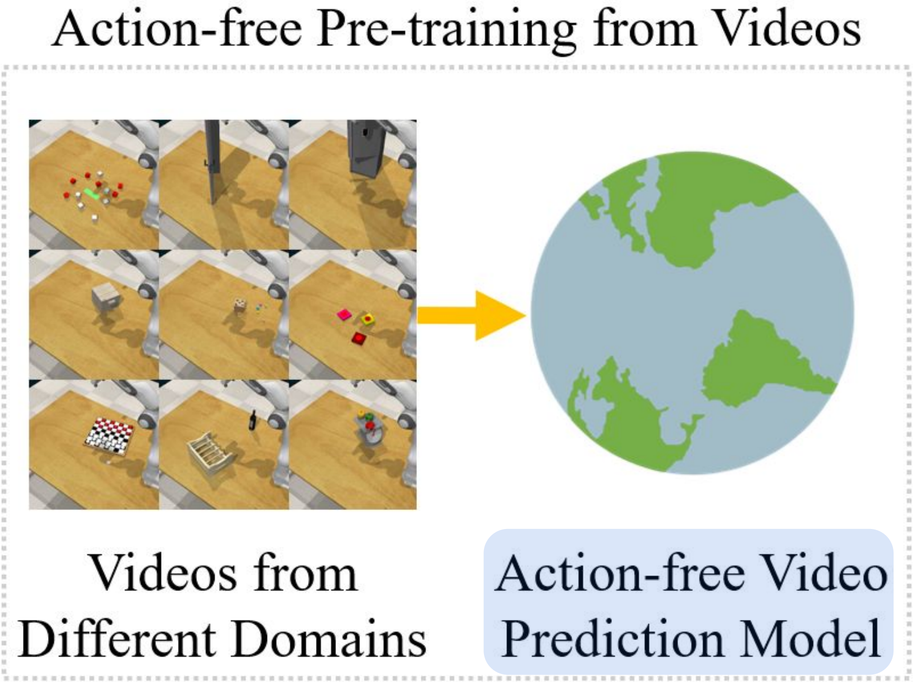


Why?

- Diverse
- Readily available
- Rich visual information
- Temporal information is crucial for sequential decision making problems

Method: APV

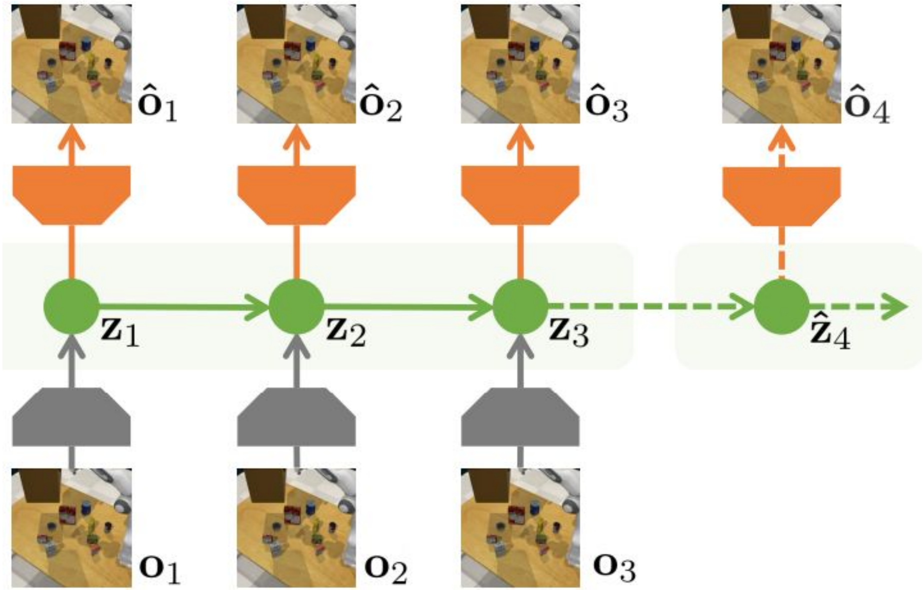
- We present **APV**: **A**ction-free **P**re-training from **V**ideos
  - **Step 1:** Pre-train an action-free video prediction model


Action-free Pre-training from Videos
Videos from Different Domains
Action-free Video Prediction Model

Learns both visual representations and dynamics from videos

Method: APV

- **Step 1:** Pre-train an action-free video prediction model
  - Train an action-free recurrent state-space model [Hafner'19]



Representation model: $z_t \sim q_\phi(z_t \mid z_{t-1}, o_t)$

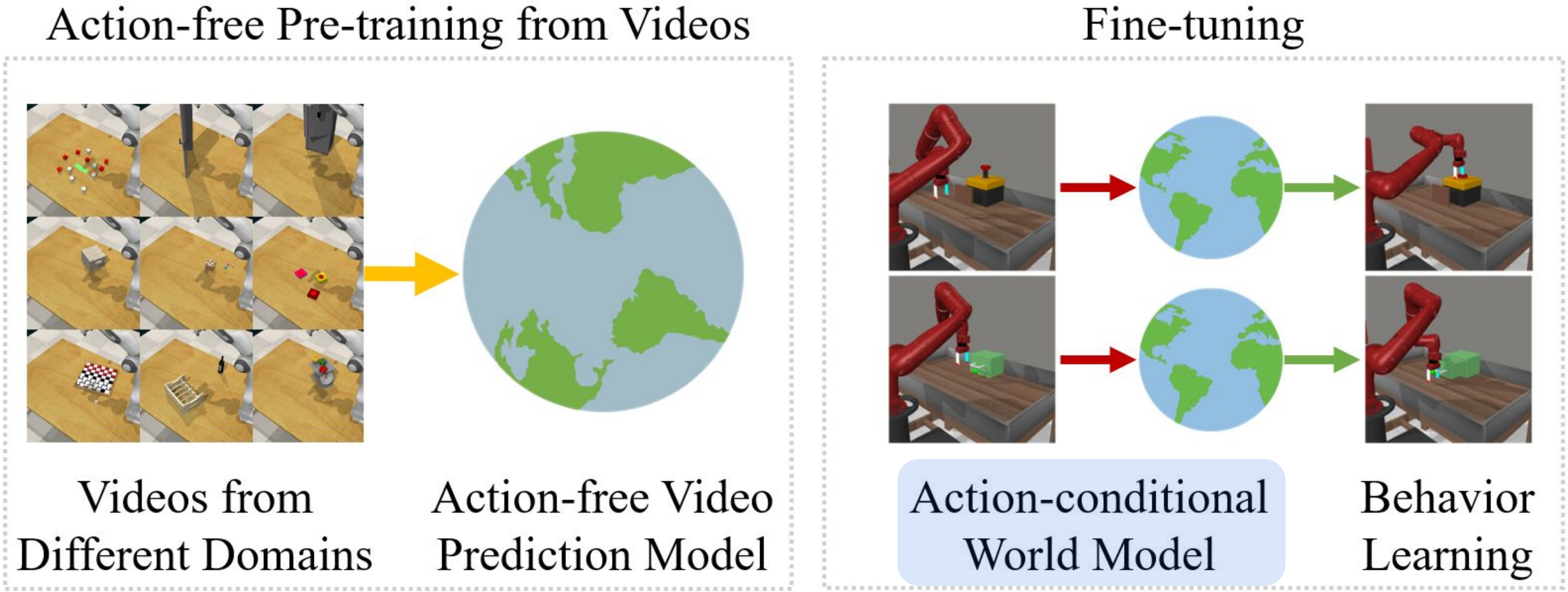
Transition model: $\hat{z}_t \sim p_\phi(\hat{z}_t \mid z_{t-1})$

Image decoder: $\hat{o}_t \sim p_\phi(\hat{o}_t \mid z_t)$

[Hafner'19] Hafner, Danijar, Timothy Lillicrap, Ian Fischer, Ruben Villegas, David Ha, Honglak Lee, and James Davidson. "Learning latent dynamics for planning from pixels." In International conference on machine learning, pp. 2555-2565. PMLR, 2019.

Method: APV

- We present **APV**: **A**ction-free **P**re-training from **V**ideos
  - **Step 2:** Fine-tuning for learning action-conditional world model



Action-free Pre-training from Videos

Fine-tuning

Videos from Different Domains

Action-free Video Prediction Model

Action-conditional World Model

Behavior Learning

Method: APV

- We present **APV**: **A**ction-free **P**re-training from **V**ideos
  - **Step 2:** Fine-tuning for learning action-conditional world model

- **Main challenge:**
  - How to incorporate additional actions to the pre-trained action-free prediction model?

Method: APV

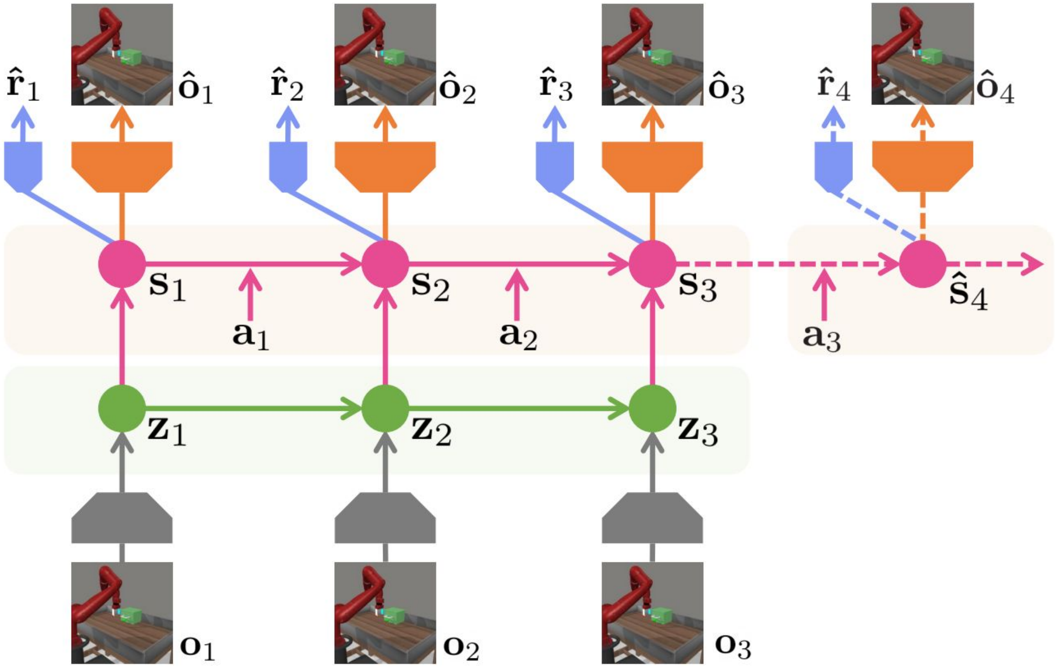
- We present **APV**: **A**ction-free **P**re-training from **V**ideos
  - **Step 2:** Fine-tuning for learning action-conditional world model

- **Main challenge:**
  - How to incorporate additional actions to the pre-trained action-free prediction model?

- We should design a framework for **smooth transition from action-free pre-training to action-conditional fine-tuning!**

Method: APV

- **Step 2:** Fine-tuning for learning action-conditional world model
  - Train a **stacked** latent dynamics model



Action-free

$$\begin{cases} \text{Representation model:} & z_t \sim q_\phi(z_t \mid z_{t-1}, o_t) \\ \text{Transition model:} & \hat{z}_t \sim p_\phi(\hat{z}_t \mid z_{t-1}) \end{cases}$$

Action-conditional

$$\begin{cases} \text{Representation model:} & s_t \sim q_\theta(s_t \mid s_{t-1}, a_{t-1}, z_t) \\ \text{Transition model:} & \hat{s}_t \sim p_\theta(\hat{s}_t \mid s_{t-1}, a_{t-1}) \end{cases}$$

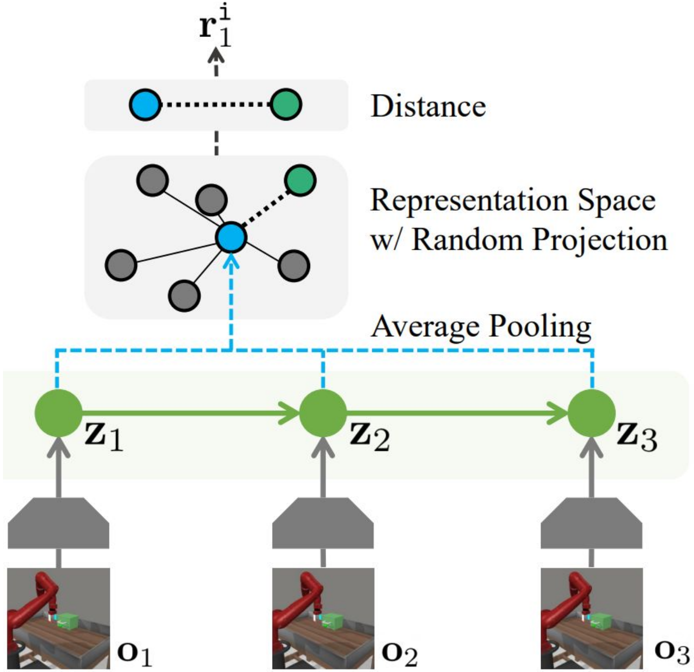
$$\begin{aligned} &\text{Image decoder:} && \hat{o}_t \sim p_\theta(\hat{o}_t \mid s_t) \\ &\text{Reward predictor:} && \hat{r}_t \sim p_\theta(\hat{r}_t \mid s_t), \end{aligned} \qquad (3)$$

Method: APV

- **Additional component:** Video-based Intrinsic Bonus
    - **Motivation:** Utilize pre-trained representations for exploration

Method: APV

- **Additional component:** Video-based Intrinsic Bonus
  - **Motivation:** Utilize pre-trained representations for exploration



- Increase the diversity of visited **trajectories** instead of single states

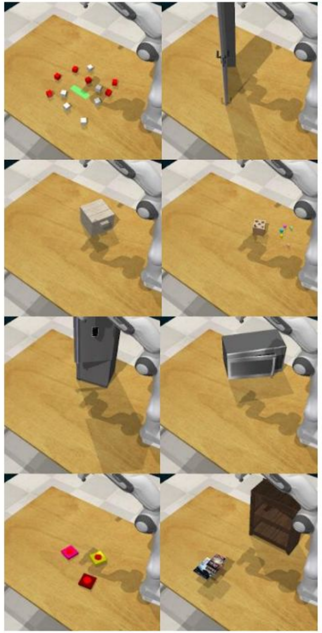
$$y_t = \mathrm{Avg}(z_{t:t+\tau})$$

$$r_t^{\mathtt{int}} \doteq ||\psi(y_t) - \psi(y_t^k)||_2$$

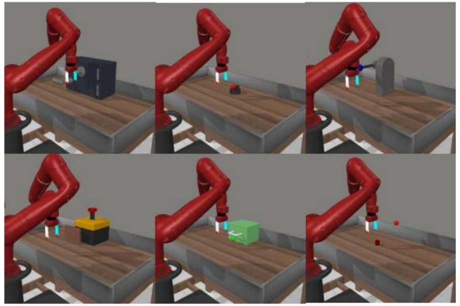
ψ is a random projection

Experimental Setup

- For behavior learning, we utilize DreamerV2 [Hafner’21]
- We consider two **transfer** setups



Pre-training Videos from RLBench

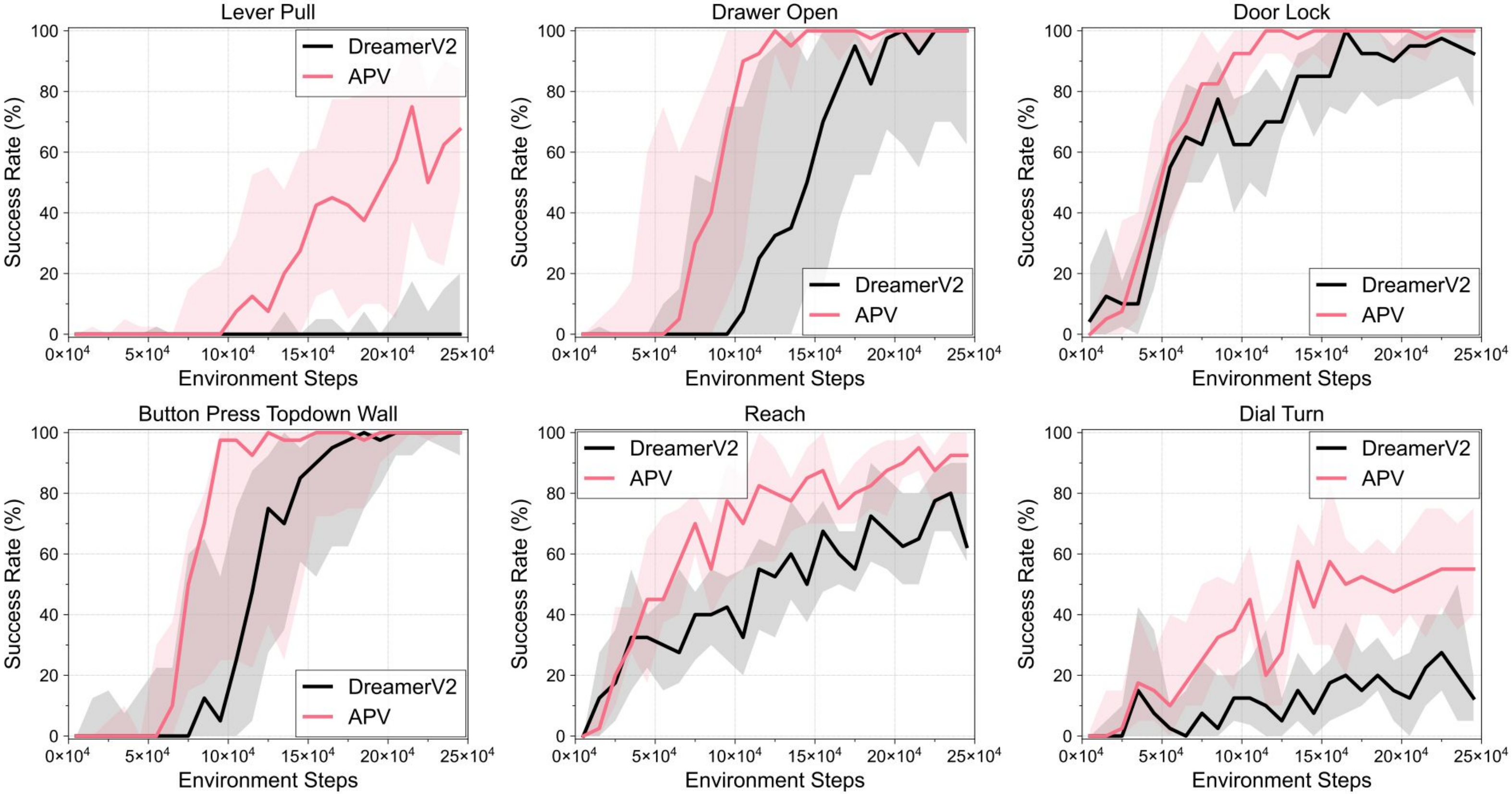


Robotic Manipulation Tasks from Meta-world

Robotic Locomotion Tasks from DeepMind Control Suite

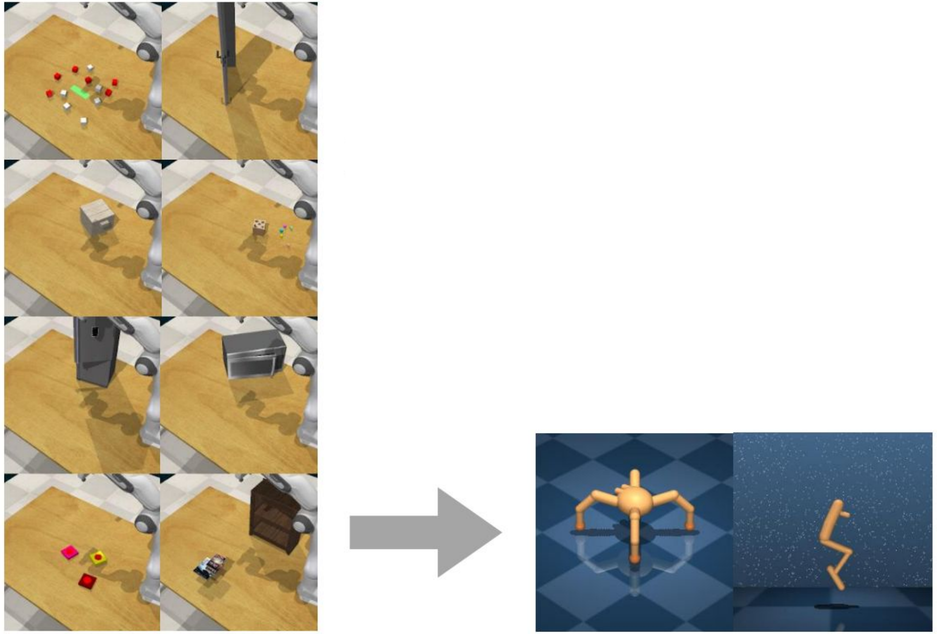
[Hafner’21] Hafner, Danijar, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. "Mastering atari with discrete world models." International Conference on Learning Representations, 2021

Results on Visual Manipulation Tasks



Experimental Setup

- For behavior learning, we utilize DreamerV2 [Hafner'21]
- We consider two **transfer** setups

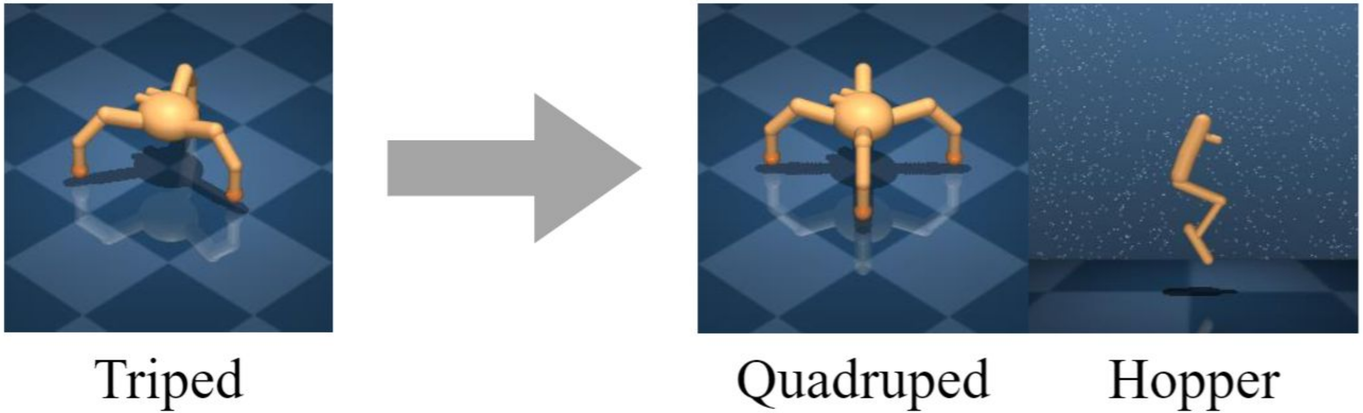


Pre-training Videos from RLBench

Robotic Locomotion Tasks from DeepMind Control Suite

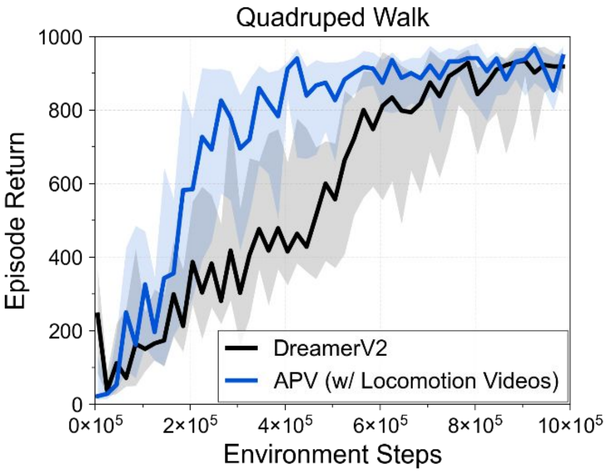
Experimental Setup

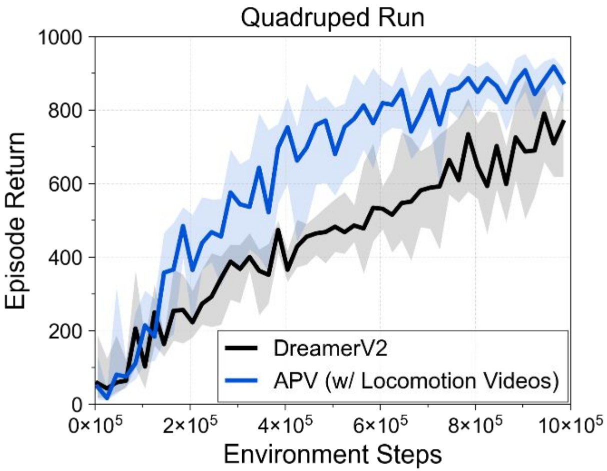
- We also consider videos from a different task but with very similar visuals

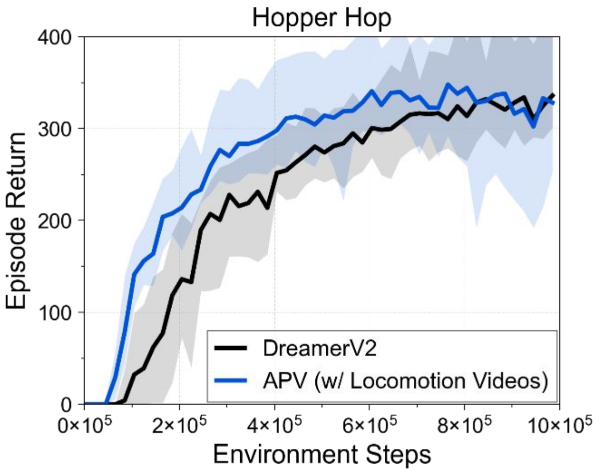


Results on Visual Locomotion Tasks

- Pre-training with in-domain locomotion videos improves performance on locomotion tasks

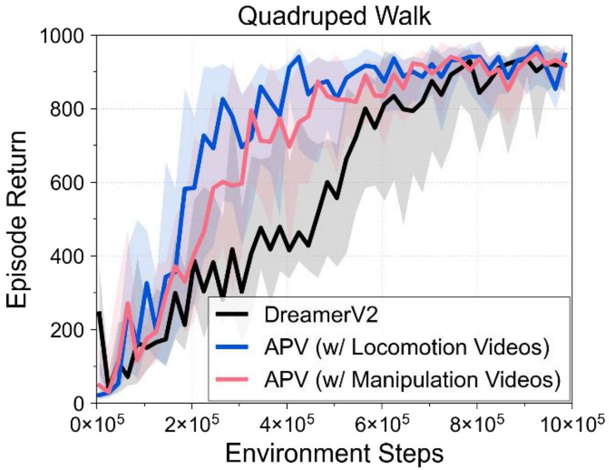


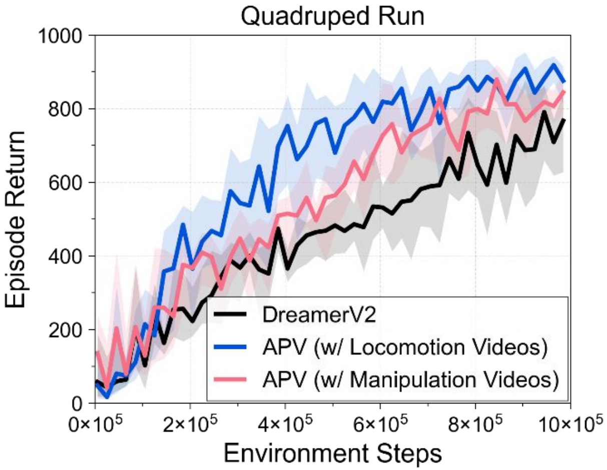


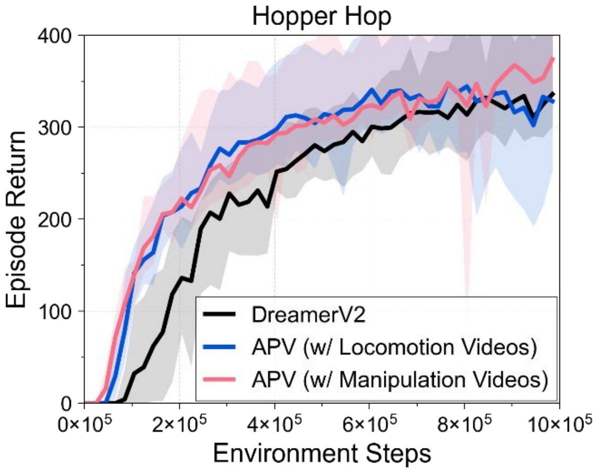


Results on Visual Locomotion Tasks

- Pre-training with in-domain locomotion videos improves performance on locomotion tasks
- Interestingly, pre-training with out-of-domain manipulation videos can also improve performance

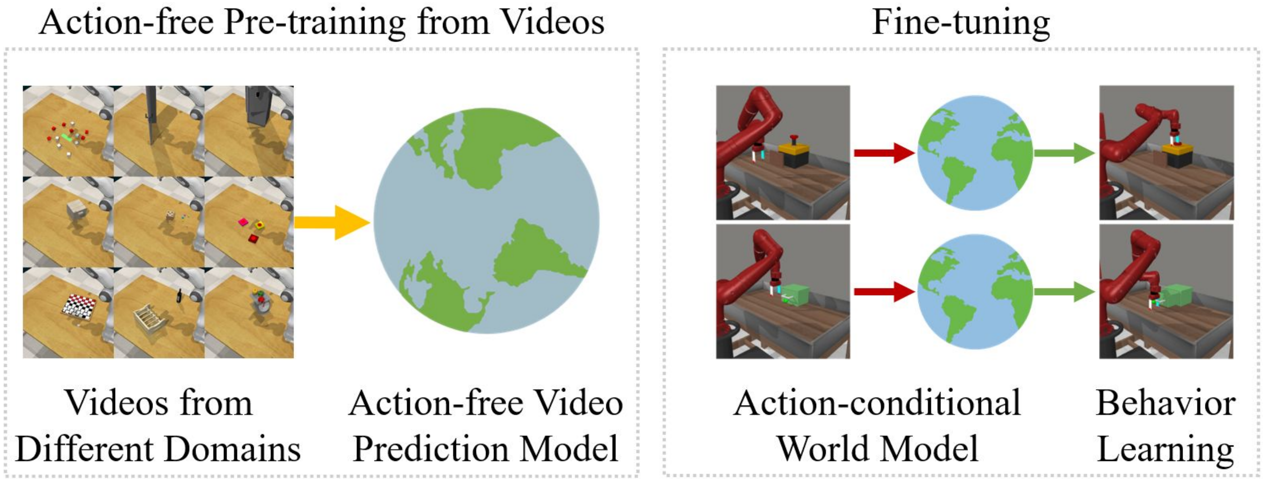






Conclusion

- We introduce APV, a visual model-based RL framework that can leverage diverse, action-free videos for pre-training
- More experimental results and analysis are available in paper
  - Please visit **Hall E #916**


Action-free Pre-training from Videos
Fine-tuning
Videos from Different Domains
Action-free Video Prediction Model
Action-conditional World Model
Behavior Learning