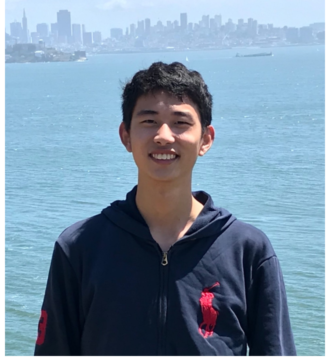
Unsupervised Visual Attention and Invariance for Reinforcement Learning

Xudong Wang*
Long Lian*
Stella Yu

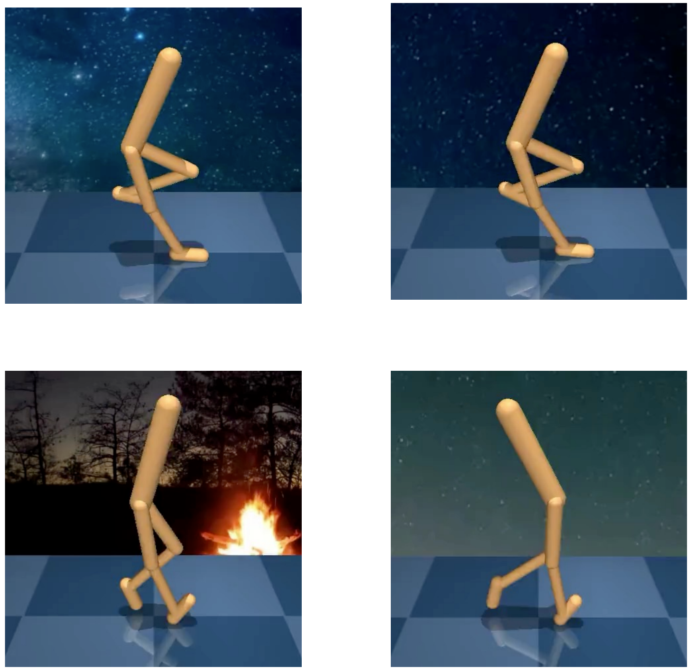

UC Berkeley / ICSI

How to Generalize Vision-based RL to Unknown Test Environments?

Training: environment with fixed background

Testing: unknown test environments

Deploy

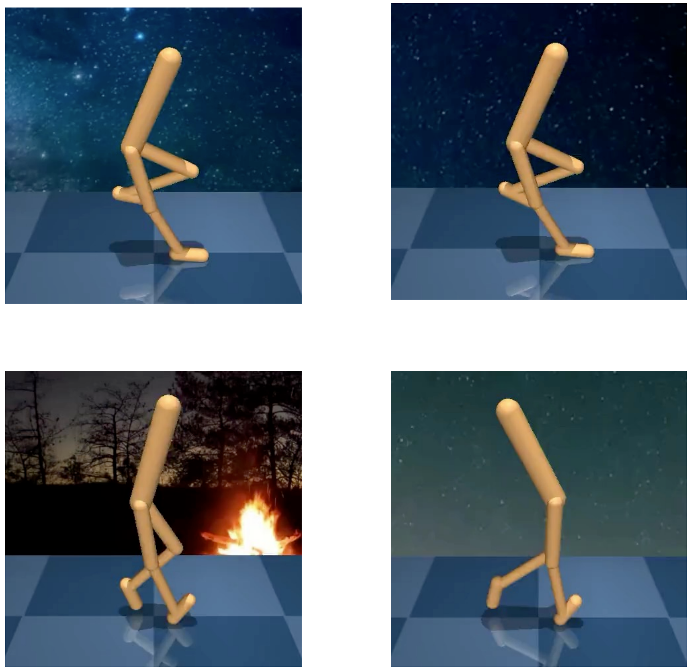


How to Generalize Vision-based RL to Unknown Test Environments?

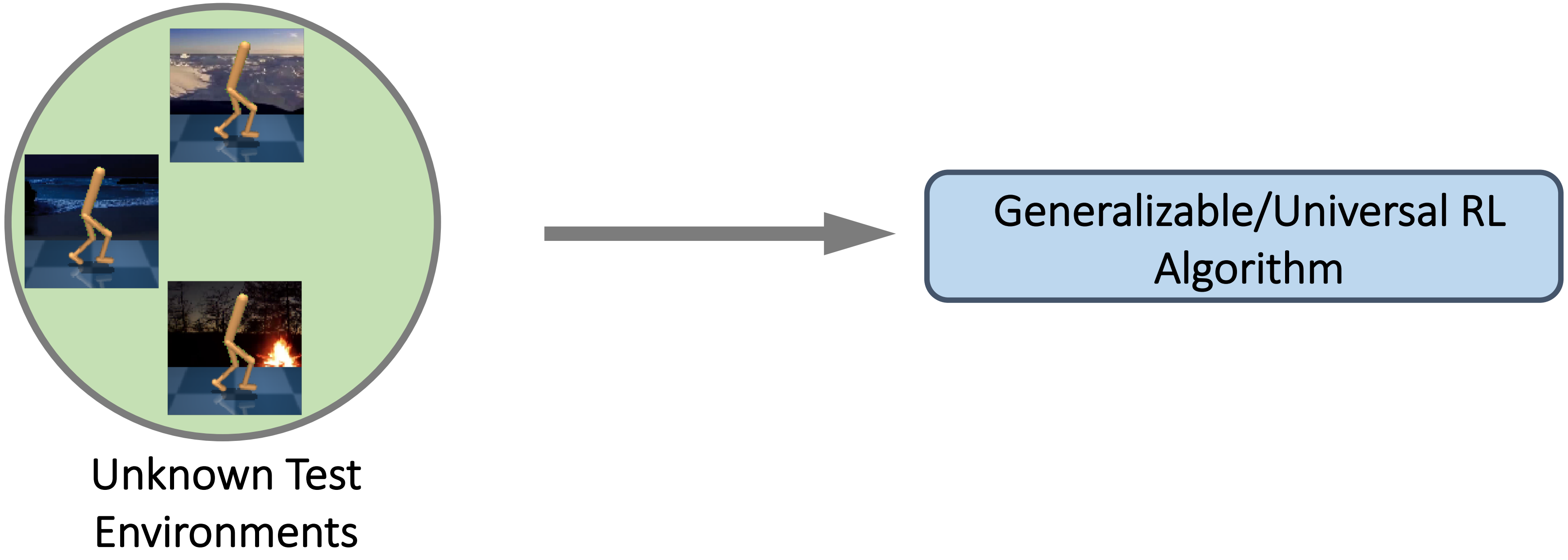
Training: environment with fixed background

Testing: unknown test environments

Deploy

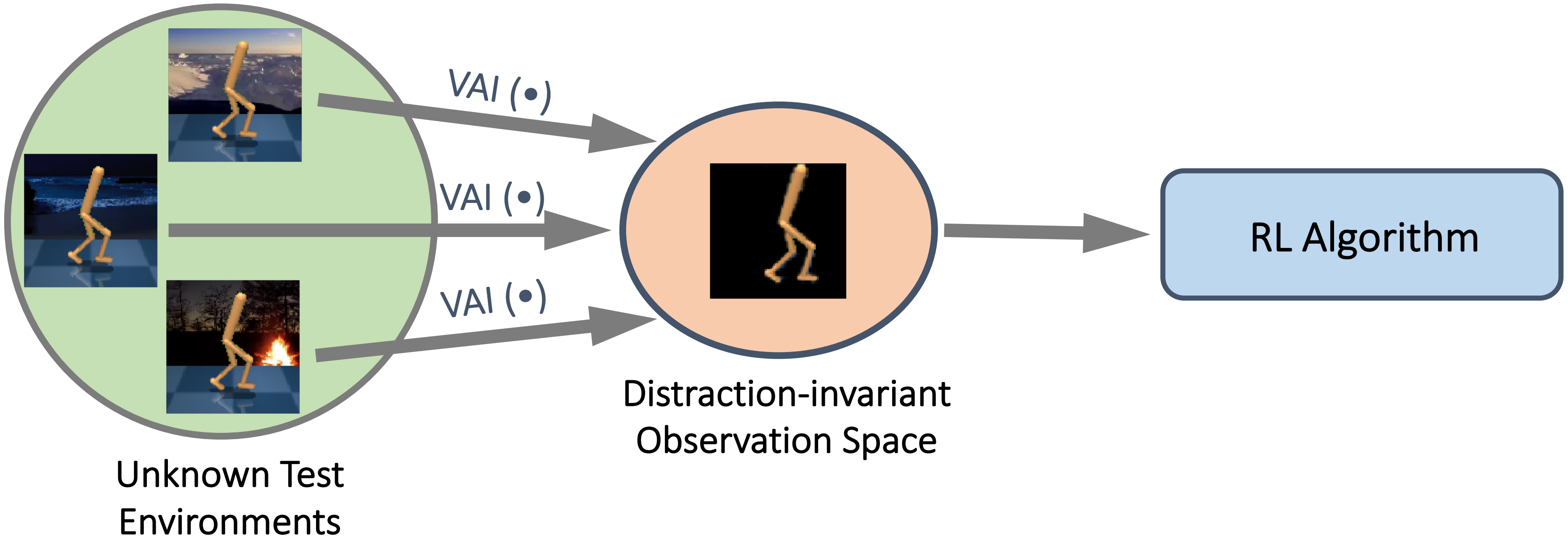


Existing Methods: Universal/Generalizable RL



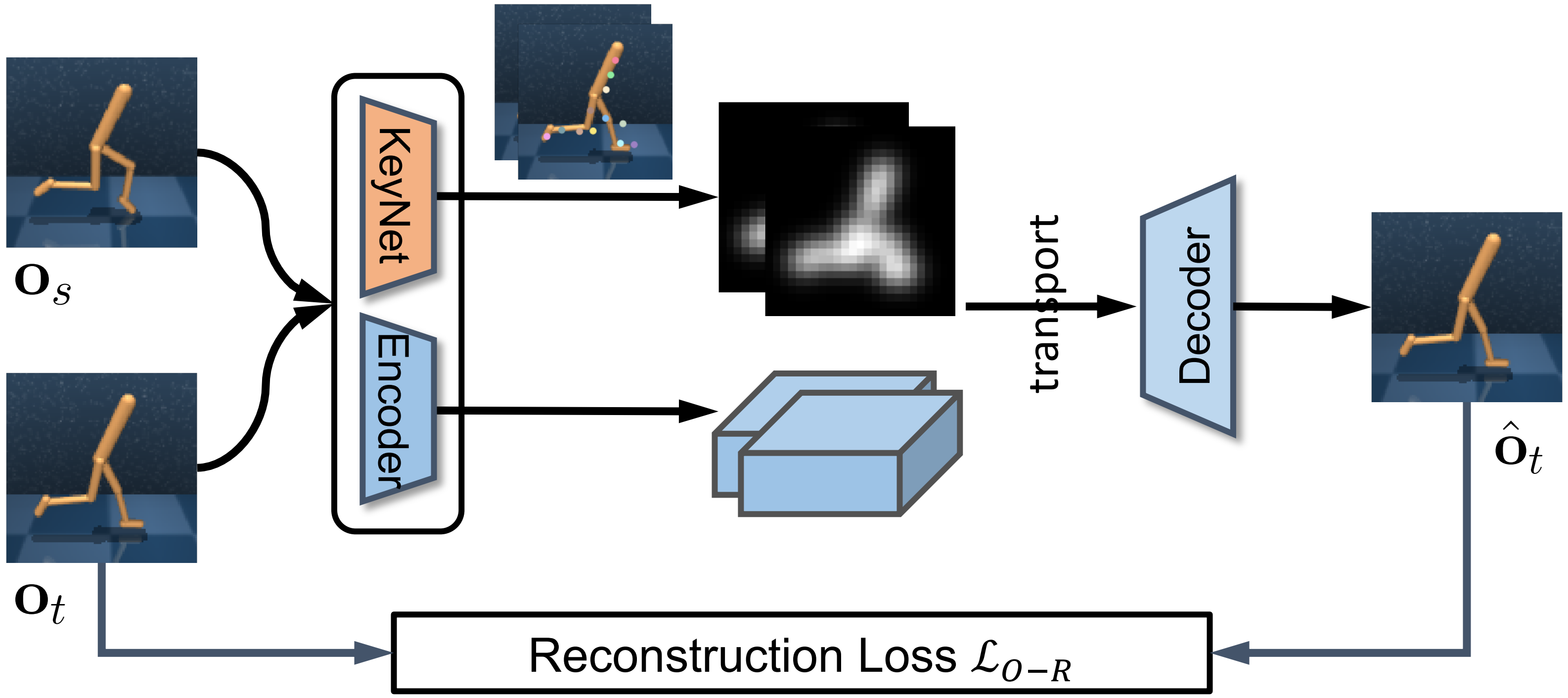
- **Most existing methods:** a **universal RL model.**
  Caveat: often leads to ***instability*** in training since RL algorithms are fragile.
- **Recent works:** *adapt at test time.*
  Caveat: leads to more ***unpredictability*** and ***long latency*** at test time.

Our Approach: Feeding "Clean" and Invariant Vision to RL



We try to **transform the input data to a distraction-invariant observation space**, and then ask the RL algorithm to perform in such a space without distractions.

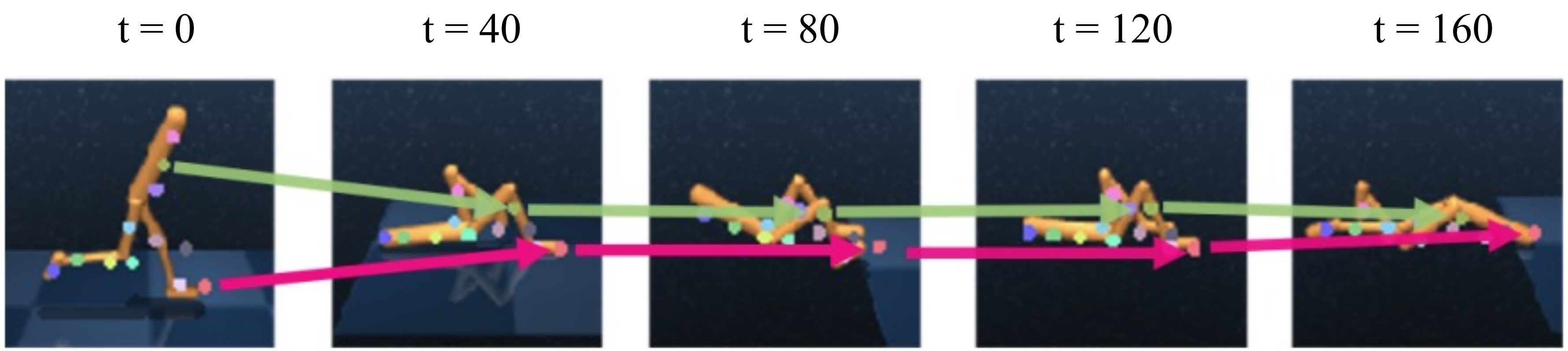
Unsupervised Keypoint Detection (Stage 1 of VAI)



[1] *Tomas Jakab, et al. Unsupervised learning of object landmarks through conditional image generation. NeurIPS 2018.*
[2] *Tejas D Kulkarni, et al. Unsupervised learning of object keypoints for perception and control. NeurIPS 2019.*

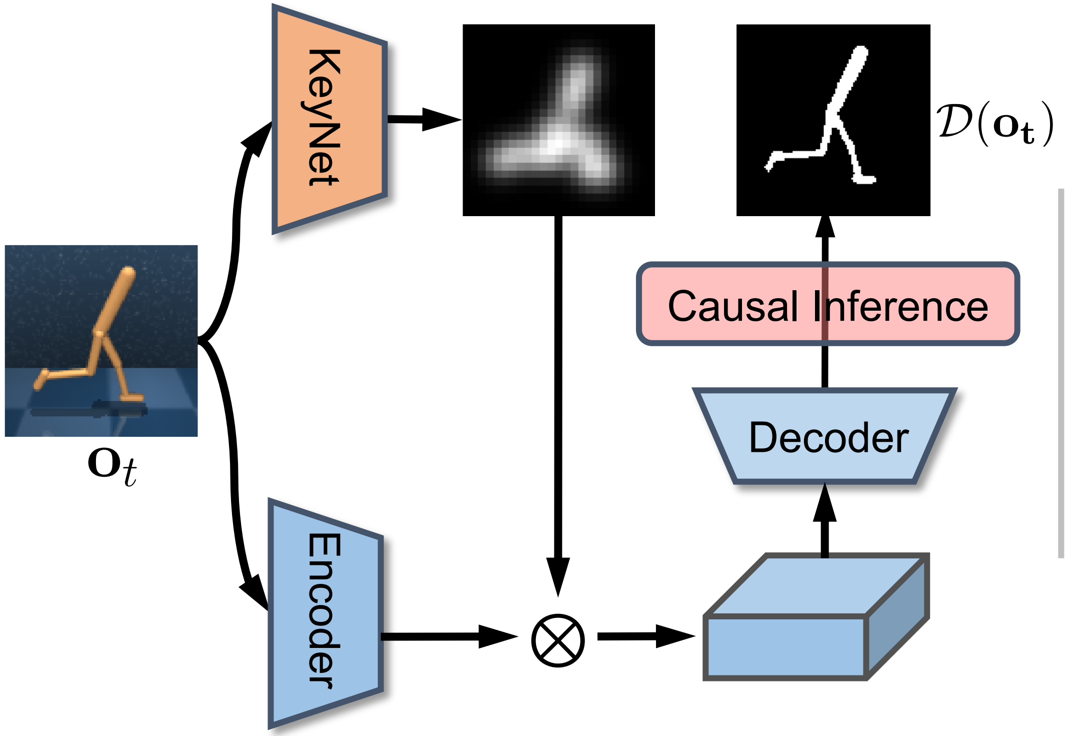
Keypoint Location as an Invariant Visual Representation?

Due to occlusion, symmetry, and lacking visual distinctions, it is often impossible to track keypoints consistently across frames.



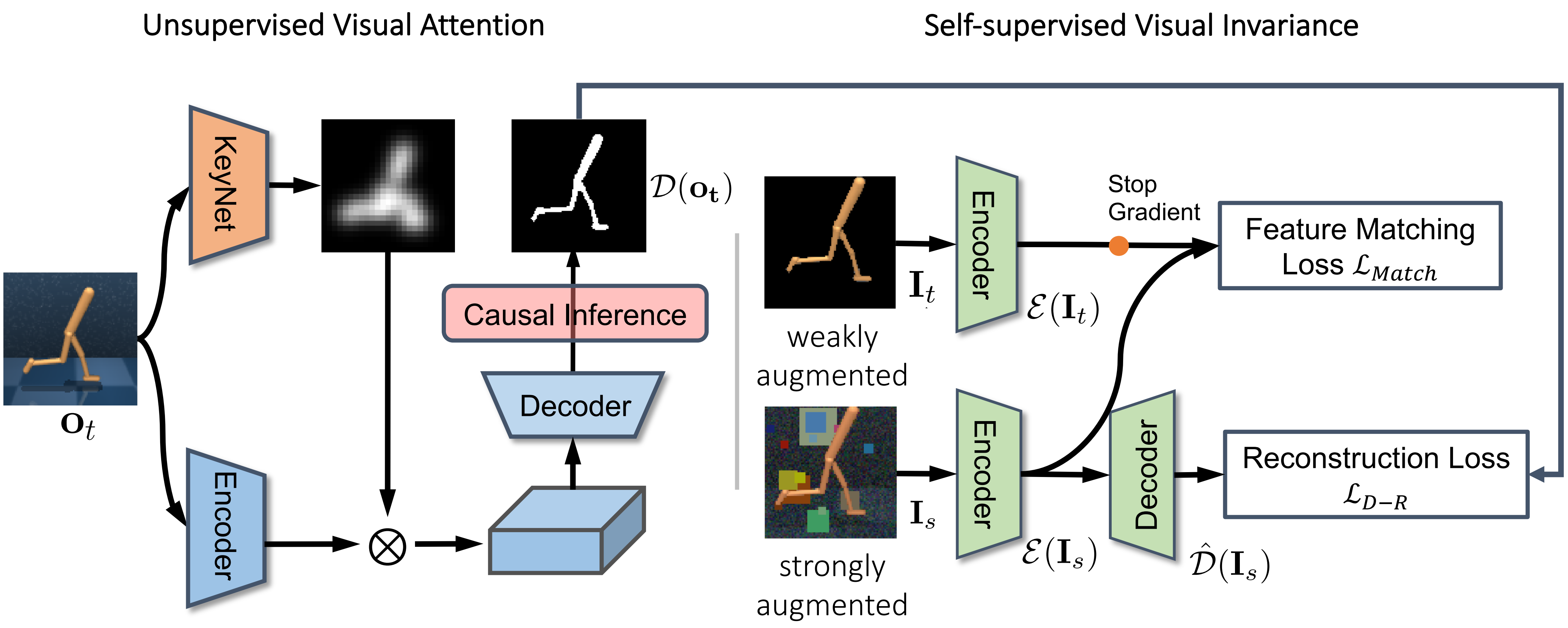
Unsupervised Visual Attention and Invariance (Stage 2&3 of VAI)

Unsupervised Visual Attention



Self-supervised Visual Invariance

Unsupervised Visual Attention and Invariance (Stage 2&3 of VAI)



DeepMind Control Benchmark

vanilla | randomized colors | video backgrounds | distractions

- Training environment:
  - vanilla environment without domain distractions
- Testing environments:
  - randomized background colors
  - non-stationary videos
  - distracting objects.

[1] Hansen, Nicklas, et al. "Self-supervised policy adaptation during deployment." ICLR 2021.
[2] Yu, Tianhe, et al. "Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning." In Conference on Robot Learning, 2020.

Our Proposed DrawerWorld Benchmark

vanilla

realistic textures

- Training environment:
  - vanilla environment without domain distractions
- Testing environments:
  - realistic textures: such as marble, metal and wood, as background
- DrawerWorld is harder since CNN is very sensitive to texture changes

[1] Hansen, Nicklas, et al. "Self-supervised policy adaptation during deployment." ICLR 2021.

[2] Yu, Tianhe, et al. "Meta-world: A benchmark and evaluation for multi-task and meta reinforcement learning." In Conference on Robot Learning, 2020.

VAI Outperforms Current SOTA by 33~53% on Deepmind Control

Random colors	SAC	DR	PAD	SODA+P	VAI	VAI+P	Δ
Walker, walk	414 ±74	594 ±104	468 ±47	692 ±68	819 ±11	**918 ±6**	+226 (↑ 33%)
Walker, stand	719 ±74	715 ±96	797 ±46	893 ±12	**964 ±2**	**968 ±3**	+75 (↑ 8%)
Cartpole, swingup	592 ±50	647 ±48	630 ±63	805 ±28	**830 ±10**	819 ±6	+14 (↑ 2%)
Cartpole, balance	857 ±60	867 ±37	848 ±29	-	**990 ±4**	957 ±9	+142 (↑ 17%)
Ball in cup, catch	411 ±183	470 ±252	563 ±50	949 ±19	886 ±33	**960 ±8**	+11 (↑ 1%)
Finger, spin	626 ±163	465 ±314	803 ±72	793 ±128	932 ±3	**968 ±6**	+165 (↑ 21%)
Finger, turn_easy	270 ±43	167 ±26	304 ±46	-	**445 ±36**	**455 ±48**	+151 (↑ 50%)
Cheetah, run	154 ±41	145 ±29	159 ±28	-	**337 ±1**	**334 ±2**	+178 (↑ 112%)
Reacher, easy	163 ±45	105 ±37	214 ±44	-	**934 ±22**	**936 ±19**	+722 (↑ 337%)
average	*467*	*464*	*531*	-	*793*	***812***	+281 (↑ 53%)

cumulative rewards when tested on
randomized colors

Video background	SAC	DR	PAD	SODA	SODA+P	VAI	VAI+P	Δ
Walker, walk	616 ±80	655 ±55	717 ±79	635 ±48	768 ±38	870 ±21	**917 ±8**	+149 (↑ 19%)
Walker, stand	899 ±53	869 ±60	935 ±20	903 ±56	955 ±13	**966 ±4**	**968 ±2**	+13 (↑ 1%)
Cartpole, swingup	375 ±90	485 ±67	521 ±76	474 ±143	**758 ±62**	624 ±146	**761 ±127**	+3 (↑ 0%)
Cartpole, balance	693 ±109	766 ±92	687 ±58	-	-	**869 ±189**	847 ±205	+182 (↑ 26%)
Ball in cup, catch	393 ±175	271 ±189	436 ±55	539 ±111	**875 ±56**	790 ±249	846 ±229	-29 (↓ 3%)
Finger, spin	447 ±102	338 ±207	691 ±80	363 ±185	695 ±97	569 ±366	**953 ±28**	+258 (↑ 37%)
Finger, turn_easy	355 ±108	223 ±91	362 ±101	-	-	419 ±50	**442 ±33**	+80 (↑ 22%)
Cheetah, run	194 ±30	150 ±34	206 ±34	-	-	**322 ±35**	**325 ±31**	+119 (↑ 58%)
average	*497*	*470*	*569*	-	-	*678*	***757***	+188 (↑ 33%)

cumulative rewards when tested on
video background

VAI Outperforms Current SOTA by 61~229% on DrawerWorld

success %	DrawerOpen				DrawerClose			
	SAC	PAD	VAI	Δ	SAC	PAD	VAI	Δ
Grid	98 ±2	84 ±7	**100 ±0**	+2 (↑ 2%)	**100 ±0**	95 ±3	99 ±1	-1 (↓ 1%)
Black	95 ±2	95 ±3	**100 ±1**	+5 (↑ 5%)	75 ±4	64 ±9	**100 ±0**	+25 (↑ 33%)
Blanket	28 ±8	54 ±6	**86 ±6**	+32 (↑ 59%)	0 ±0	0 ±0	**85 ±8**	+85 (↑ ∞%)
Fabric	2 ±1	20 ±6	**99 ±1**	+79 (↑ 395%)	0 ±0	0 ±0	**74 ±8**	+74 (↑ ∞%)
Metal	35 ±7	81 ±3	**98 ±2**	+17 (↑ 21%)	0 ±0	2 ±2	**98 ±3**	+96 (↑ 4800%)
Marble	3 ±1	3 ±1	**43 ±7**	+40 (↑ 1333%)	0 ±0	0 ±0	**49 ±13**	+49 (↑ ∞%)
Wood	18 ±5	39 ±9	**94 ±4**	+55 (↑ 141%)	0 ±0	12 ±2	**70 ±6**	+58 (↑ 483%)
average	*40*	*54*	***87***	+33 (↑ *61%*)	*25*	*25*	***82***	+57 (↑ *228%*)

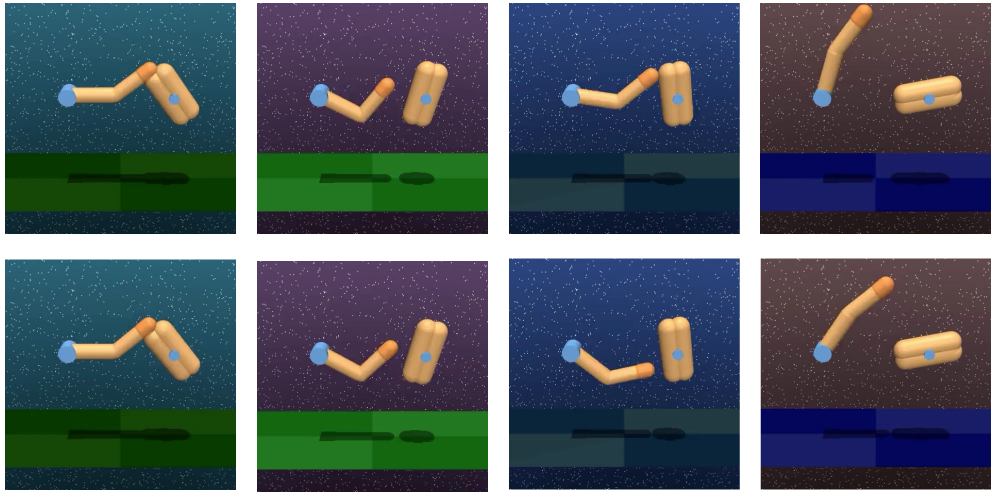
success rate when tested on
realistic textures

Demo

Task: Finger, spin; **Test env.**: randomized color

Baseline

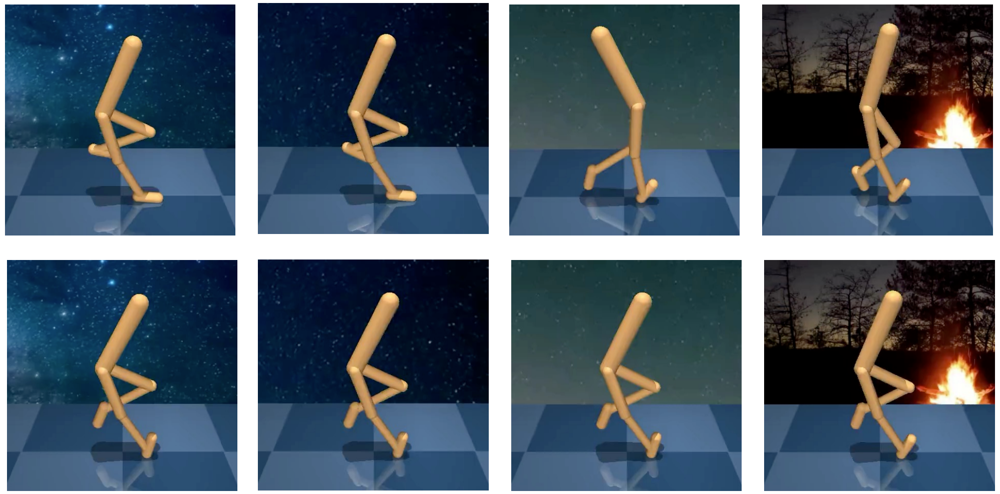
VAI



Task: Walker, walk; **Test env.**: video background

Baseline

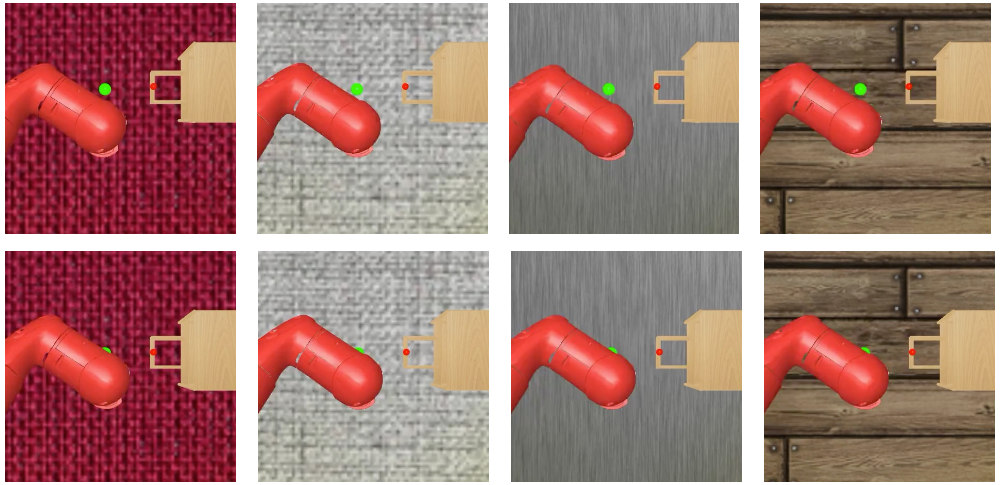
VAI



Task: DrawerOpen; **Test env.**: realistic textures

Baseline

VAI



Takeaway

Adapt the vision, not RL!