Control First, Robustness Next

Decoupled Representation Learning for Visual RL Generalization

Chanyong Heo, Hyelyn Jeong, Jongchan Park, Seungjun Oh, and Yusung Kim
CSI Lab. https://csi-skku.github.io

SUNG KYUN KWAN UNIVERSITY 1398

Visit our project page!

NEURAL INFORMATION PROCESSING SYSTEMS

1 Problem: Visual Shifts Disrupt Control

The task is unchanged, yet visual shifts can cause incorrect actions and lower performance.

DMC-GB



DrQ avg. return: 826.8 → 459.9 (−44.4%)

Robosuite



DrQ avg. return: 315.5 → 98.3 (−68.8%)

EXISTING APPROACHES: COUPLED LEARNING

Coupled learning jointly optimizes RL and auxiliary objectives through a shared visual encoder.



2 Motivation: Preserve Control While Learning Robustness

Coupled methods show greater feature drift and higher TD error than DrQ, on average.

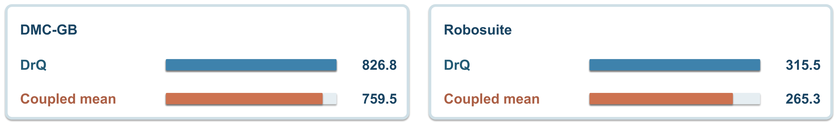
Drift measures how much the same observation's features change during training.

DrQ baseline = 1.0× — **Ratios = Coupled mean / DrQ**

BENCHMARK	Feature change over training	Value-target mismatch
DMC-GB	1.61×	1.51×
Robosuite	3.15×	3.47×

Coupled mean: SVEA, SimGRL, SMG — Above 1.0× = greater drift / TD error

Original-environment return: DrQ baseline vs. coupled mean

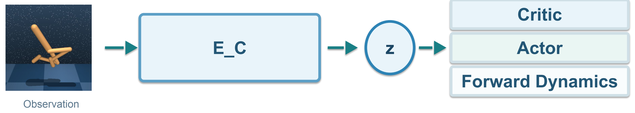


3 Method

STAGE 1: ONLINE RL

Learn a control-oriented reference

Collect source-environment experience with DrQ-style RL.

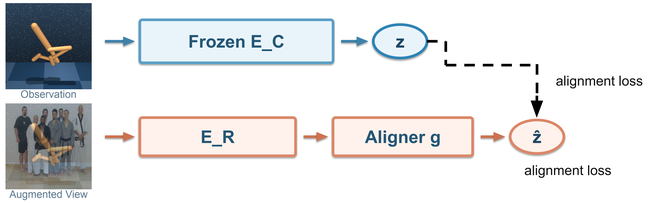


A trained actor for control and an encoder (E_C) to guide Stage 2.

STAGE 2: OFFLINE REPRESENTATION LEARNING

Post-Representation Alignment (PRA)

Initialize E_R with E_C and reuse Stage-1 replay data to learn visual robustness offline.



No policy retraining. No additional environment interactions.

INFERENCE

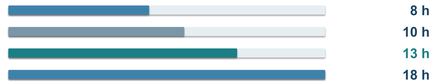
Aligned features drive the pretrained actor



Use g(E_R(o)) as the actor input under visual shifts.

TRAINING TIME

DMControl · 500K online steps · RTX 4090

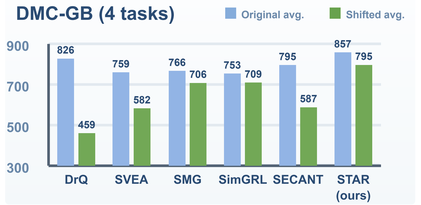


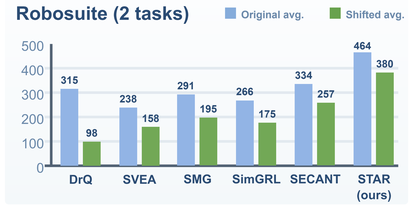
STAR = 11.5 h control + 1.5 h robustness.

One offline alignment stage keeps robustification simple, with 13 h total vs. SMG's 18 h.

4 Evaluation

A Main results on RL-ViGen





Shifted avg gain (DMC-GB)
+12.2%
over best baseline (709 → 795)

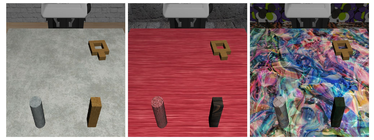
Shifted avg gain (Robosuite)
+48.1%
over best baseline (257 → 380)

B Ablation: Stage 2 Drives Visual Robustness

Benchmark	Forward Dynamics in Stage 1	Before Stage 2 (shifted avg)	After Stage 2 (shifted avg)
DMC-GB	X	399.9	**773.4**
	O	399.3	**795.7**
Robosuite	X	96.0	**262.0**
	O	103.8	**380.5**

C Beyond Online RL: Robustifying Diffusion Policies

Square environment



Mean success over Easy, Medium, and Hard (%)

Task	Pretrained DP	DP + PRA
Can	25.1	**58.4**
Square	28.9	**72.7**

PRA updates visual encoders and aligner offline.

- Apply Post-Representation Alignment to demonstration-trained Diffusion Policies for contact-rich manipulation.

5 Conclusion

- STAR separates **control learning** from **robustness learning** through **offline representation alignment** to a fixed control reference.
- It robustifies pretrained **visual RL and Diffusion Policies** using existing data, **without additional environment interaction or policy retraining**.
- Its effectiveness depends on the **pretrained control representation** and the **visual shifts** addressed by Stage-2 augmentation.