



# An Imitation from Observation Approach to Transfer Learning with Dynamics Mismatch

Siddharth Desai  
UT Austin

Ishan Durugkar  
UT Austin

Haresh Karnan  
UT Austin

Garrett Warnell  
Army Research Laboratory  
UT Austin

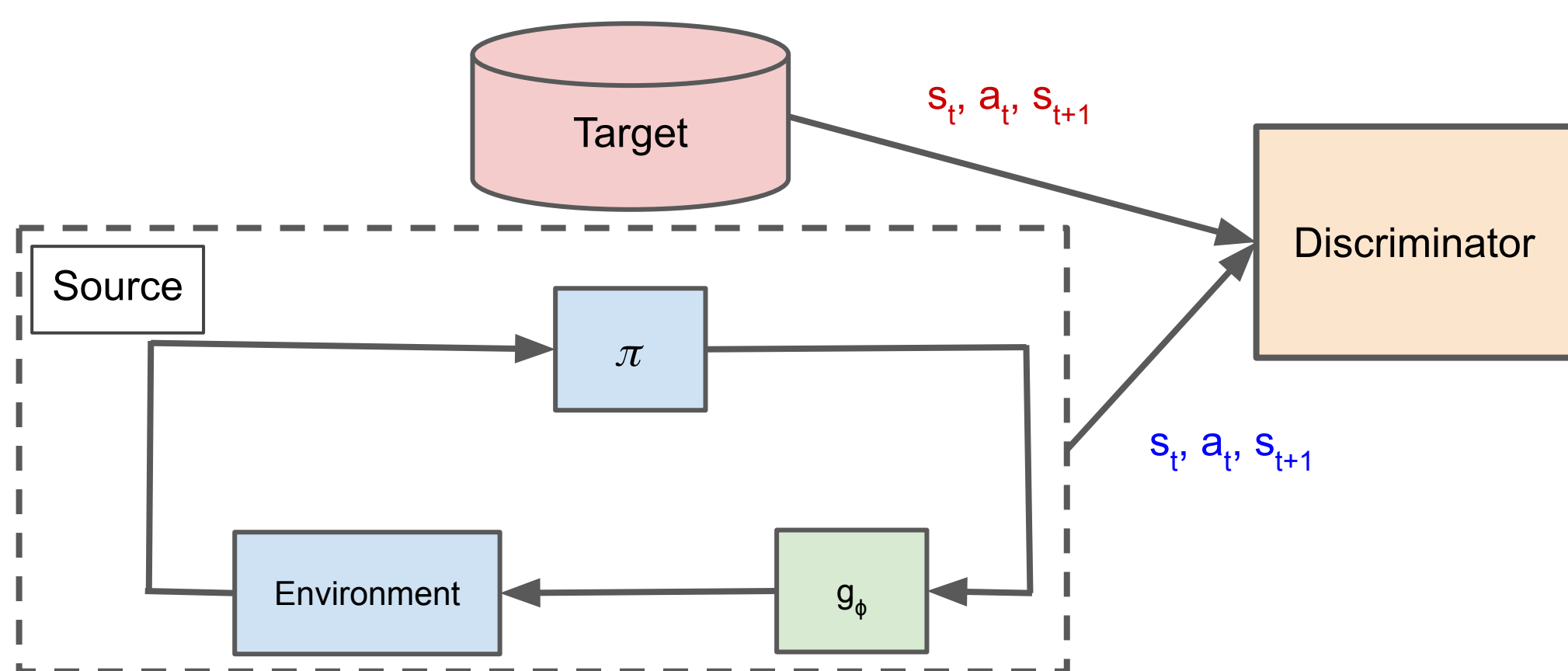
Josiah P Hanna  
School of Informatics  
The University of Edinburgh

Peter Stone  
UT Austin  
Sony AI



## Abstract

- We examine the problem of transferring a policy learned in a source environment to a target environment with **different dynamics**, particularly in the case where it is critical to reduce the amount of interaction with the target environment during learning (e.g **Sim-to-Real** transfer).
- We show that one existing solution to this transfer problem - Grounded Action Transformation (GAT) - is closely related to the problem of **Imitation from Observation** (IfO).
- After establishing this relationship, we hypothesize that recent state-of-the-art approaches from the IfO literature can be effectively repurposed for **grounded transfer learning**.
- To validate our hypothesis, we derive a new algorithm - Generative Adversarial Reinforced Action Transformation (**GARAT**) - based on adversarial imitation from observation techniques.
- We run experiments in several domains with mismatched dynamics, and find that agents trained with **GARAT** achieve relatively higher returns in the target environment.



## Method

- We ground the source domain to the target domain by introducing an Action Transformation function,  $g_\phi$ .
- While in prior work, this function was learned through supervised learning, we show that learning  $g_\phi$  can be seen as an Imitation from Observation problem.
- Using this insight, **GARAT** adapts techniques from adversarial imitation from observation to learn the action transformation.
- A policy learned in the new updated source domain transfers well to the target domain.

**GARAT** overcomes dynamics mismatch in transfer learning using adversarial imitation from observation

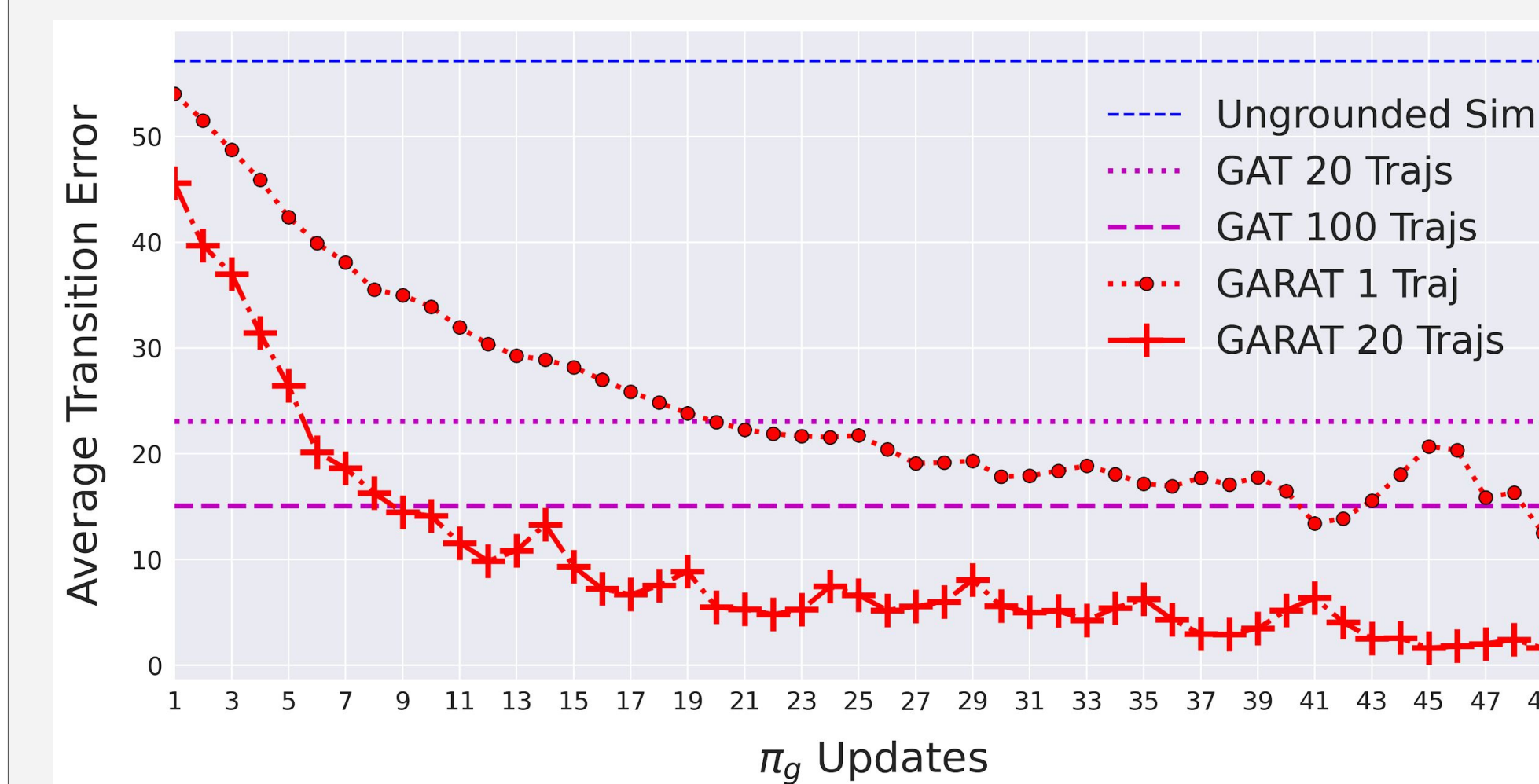


<https://arxiv.org/abs/2008.01594>

## Results

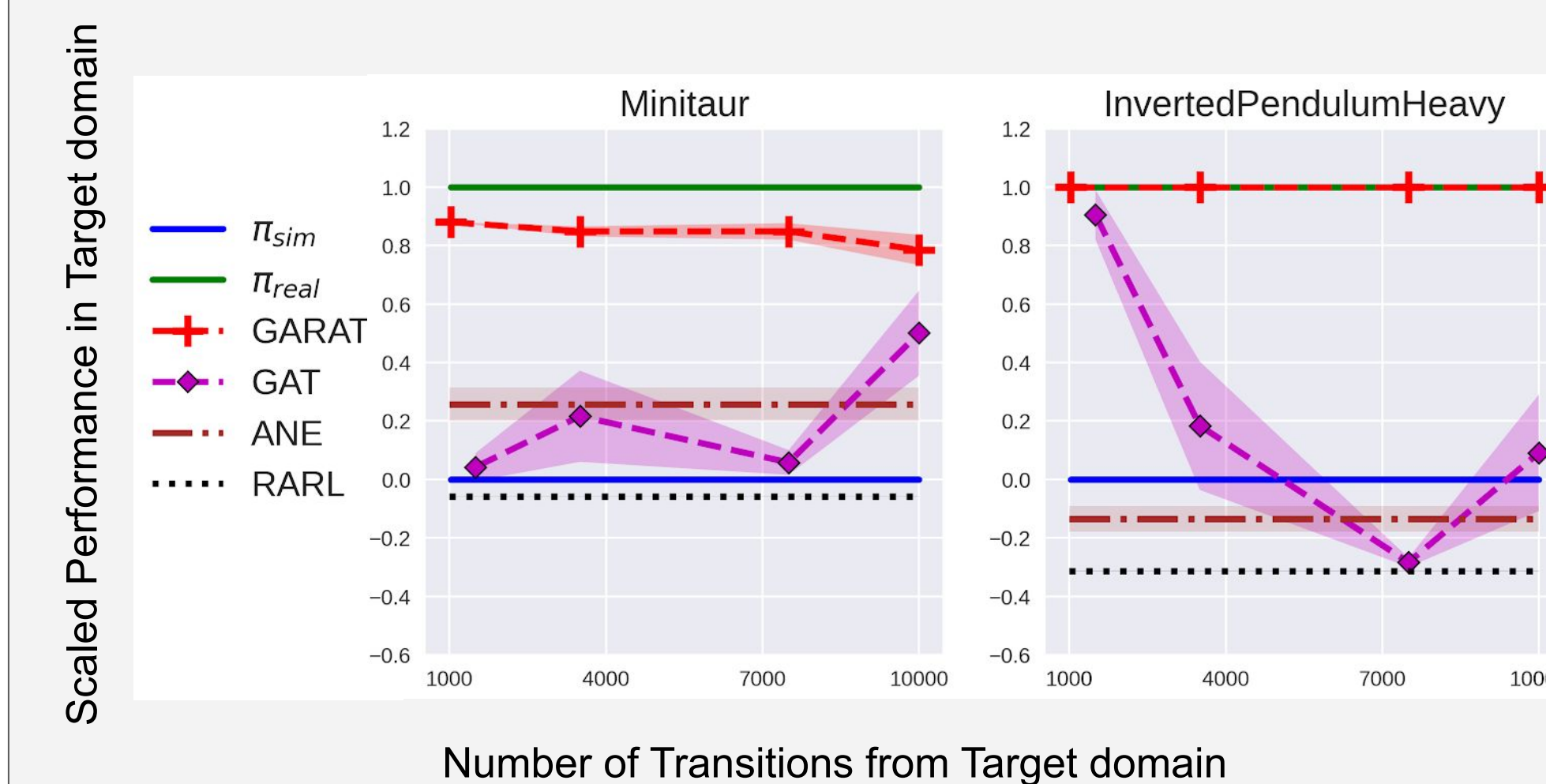
- We empirically validate our results on various Mujoco Gym tasks.
- In most domains, our method outperforms other similar algorithms and almost performs as well as training in the target environment directly.

### Minimizing Transition Error



**GARAT** reduces the mismatch in transition dynamics between source and target environments

### Experimental Results



Our method (**GARAT**) achieves higher returns than the baseline on several Mujoco Domains (full results in paper)