# Robot Manipulation Learning Using Generative Adversarial Imitation Learning

Mohamed Khalil Jabri

# Robot Manipulation Learning Using Generative Adversarial Imitation Learning*

**Mohamed Khalil Jabri**[1,2,3]

[1]IMT Atlantique, Lab-STICC, UMR 6285, team RAMBO, F-29238 Brest, France
[2]The University of Adelaide, SA, Australia
[3]CROSSING IRL CNRS 2010

mohamed-khalil.jabri@imt-atlantique.fr

## Abstract

Imitation learning allows learning complex behaviors given demonstrations. Early approaches belonging to either Behavior Cloning or Inverse Reinforcement Learning were however of limited scalability to complex environments. A more promising approach termed as Generative Adversarial Imitation Learning tackles the imitation learning problem by drawing a connection with Generative Adversarial Networks. In this work, we advocate the use of this class of methods and investigate possible extensions by endowing them with global temporal consistency, in particular through a contrastive learning based approach.

## 1 Introduction

In [Ho and Ermon, 2016], *Generative Adversarial Imitation Learning* (GAIL) was proposed to tackle the Imitation Learning (IL) problem by drawing a connection with generative adversarial networks (GANs) [Goodfellow *et al.*, 2014]. GAIL allows to recover the demonstrator's behavior without explicitly recovering the reward function by training a policy to produce trajectories resembling the expert's demonstrations and a discriminator to distinguish them.

In this project, we focus on this class of IL algorithms and their application on an assistive robot manipulator to assist frail people. While GAIL provides a natural and direct way for this goal, it gives rise to a number of challenges. We believe we can take use of the advantages brought by this approach, and investigate possible ways to improve it. In the following sections, we discuss a specific limitation of the majority of works on adversarial imitation learning and present how we envision to mitigate it.

## 2 Research Questions

The sequential nature of IL requires learning a good representation that incorporates the underlying temporal structure governing the task. It is important to be able to identify and capture the relevant high-level temporal features that govern the demonstrator behaviour.

*The work is performed in the context of project ROGAN funded by the region of Brittany, France.

Indicatively, consider the task of picking an object. When the end-effector is too far from the object, the robot does not have to exactly match the expert trajectory all the way to the goal. Instead, it should only reason about features that shape the long-term outcome of the actions it takes. If the agent is too focused on temporally local features that govern the immediate outcome of its actions, it would be harder for it to recover when it deviates from the demonstrator's trajectory.

Model-based approaches allows effective long-term planning by learning a model of the agent's environment and using it to plan. However learning such a model is a challenging and data-intensive task. On the other hand, most works on model-free imitation learning do not take into account this aspect in the learning process, and this is even more true with adversarial imitation learning. The idea we are currently working on is a GAIL-based attempt to incorporate learning long-term temporal features within the model-free adversarial imitation learning setup. We envision to achieve that by using a a more complex architecture, while, at the same time, learning better representations.

## 3 Proposed Approach

Instead of the baseline GAIL, we are building our work on GoalGAIL [Ding *et al.*, 2019]. In GoalGAIL, the authors extend GAIL by conditioning the policy on the goal and using "relabeling" as introduced in Hindsight experience Replay (HER) [Andrychowicz *et al.*, 2017] This allows to leverage failed attempts to improve the efficiency and speed up the learning. The learning is done by alternating two things: training the discriminator to distinguish the generated trajectories from real ones, and updating the policy using an off-policy RL algorithm and the discriminator output as the reward.

We seek to extend GoalGAIL with two interrelated components that simultaneously tackle the discussed limitation. These extensions are inspired from recent works on GANs architectures, as well as recent works on contrastive self-supervised learning that has proven effective in learning good data representations.

At this point, we finished reproducing the baseline on which we build our work as well as the first extension, and we are working on the second one and evaluating the performance on a set of simulated tasks.

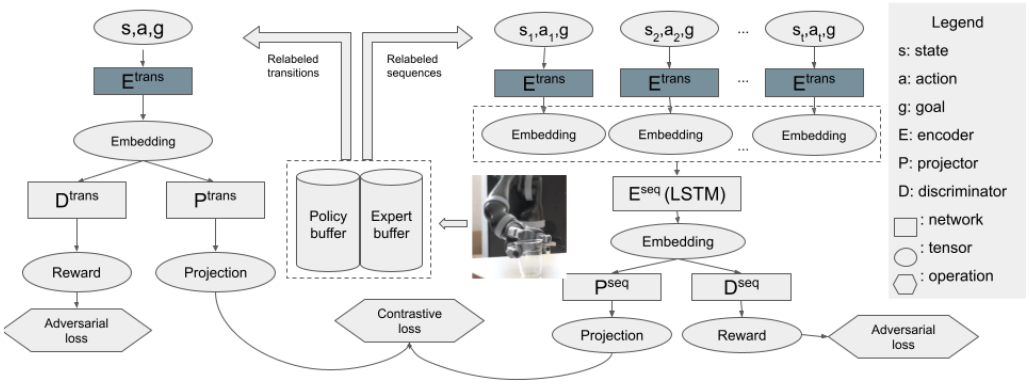


Figure 1: The trajectories produced by both the expert and the policy are stored in a memory from which transtions and sequences are sampled and relabeled to train two discriminators, while their embeddings are projected into a latent space to compute the contrastive loss

## Extension 1: Temporal discriminator

Video generation is a task that shares a particular aspect with the imitation learning problem in the sense they both need the inter-temporal dependencies to be taken into consideration, and in which GANs have proven effective. In [Vougioukas *et al.*, 2019], the authors introduce a GAN architecture capable of generating realistic facial animation by using two separate discriminators: one for the individual frames and another for the generated sequences. In our setting, we similarly consider an architecture composed of a policy network and two discriminators: a transition and a sequence discriminator, denoted $D^{trans}$ and $D^{seq}$ respectively in figure 1. The former impels the policy behaviour to be locally similar to the demonstrator's behavior while the latter encourages the policy trajectories to be globally similar to those of the expert.

For this architecture to be effective, the transitions and sequences representations need to be disentangled, meaning they should capture distinct features. This would allow the discriminators to distinguish between the immediate and the long-term effects of the actions.

## Extension 2: Contrastive learning of temporally local and global representations

InfoMaxGAN [Lee *et al.*, 2021] is a GAN framework that mitigates GANs instability using contrastive learning of local and global visual features. On one hand, maximizing the mutual information between the local and global features helps the discriminator learn long-term representation in such a way as to reduce catastrophic forgetting within the non-stationary training environment. On the other hand, distinguishing local and global features by contrastive learning helps the generator produce more diverse images, which helps the generator avoid mode collapse.

We take inspiration from this work by trying to augment the proposed architecture with a contrastive learning task where the projected embeddings of sequences and transitions lying within them are pushed closer, while the projected representations of transitions lying within distinct sequences are pulled apart. As shown in figure 1, the discriminators $D^{trans}$ and $D^{seq}$ are fed with embeddings produced by two encoders: $E^{trans}$ and $E^{seq}$ respectively. Those embedding are passed through two projectors $P^{trans}$ and $P^{seq}$ that project them into a latent space, in which we compute the contrastive loss.

The contrastive loss has a double aim. First we want the temporally local and distant features learned by the transition and sequence encoders to be disentangled in such a way as to allow each discriminator to focus only on the features that are relevant to its task. Second, in analogy with InfoMaxGAN, we want the contrastive loss to stabilize the adversarial learning.

## References


[Andrychowicz *et al.*, 2017] Marcin Andrychowicz, Filip Wolski, Alex Ray, Jonas Schneider, Rachel Fong, Peter Welinder, Bob McGrew, Josh Tobin, OpenAI Pieter Abbeel, and Wojciech Zaremba. Hindsight experience replay. In *Neural Information Processing Systems*, 2017.

[Ding *et al.*, 2019] Yiming Ding, Carlos Florensa, Pieter Abbeel, and Mariano Phielipp. Goal-conditioned imitation learning. In *Advances in Neural Information Processing Systems*. 2019.

[Goodfellow *et al.*, 2014] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*. 2014.

[Ho and Ermon, 2016] Jonathan Ho and Stefano Ermon. Generative adversarial imitation learning. In *Advances in Neural Information Processing Systems*, 2016.

[Lee *et al.*, 2021] Kwot Sin Lee, Ngoc-Trung Tran, and Ngai-Man Cheung. Infomax-gan: Improved adversarial image generation via information maximization and contrastive learning. In *IEEE Int. Winter Conf. on Applications of Computer Vision*, pages 3942–3952, 2021.

[Vougioukas *et al.*, 2019] Konstantinos Vougioukas, Stavros Petridis, and Maja Pantic. End-to-end speech-driven realistic facial animation with temporal gans. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition Workshops*, 2019.