Approaches to Imitation Learning and their Applications

Intelligent Robotics Seminar SS25

Sven Schreiber

TAMS Group
Department of Informatics
MIN Faculty
University of Hamburg

May 22, 2025

Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Table of Contents

Motivation

- Unstructured environments are hard to navigate
  - Autonomous driving, household, construction sites,...
- Challenging to specify
  1. a set of rules (manual programming)
  2. a reward function (reinforcement learning)
- Solution: learn directly from demonstrations

Imitation Learning

- Imitation of expert demonstrations
- Learn mapping from observations to actions
- Generalize policies for unstructured environments

Figure: Jiang et al.[1]

[1] Jiang et al. 'DexMimicGen: Automated Data Generation for Bimanual Dexterous Manipulation via Imitation Learning', ICRA'25

Behavioral Cloning (BC)

- Imitation directly from state-action pairs
- **Input**: state information
  - Sensory data (camera, lidar, audio)
  - Positions, TF frames
- **Output**: actions
  - "Move to position X"
  - "Open gipper"

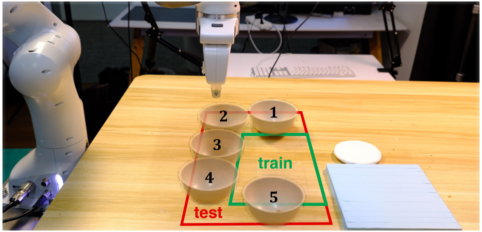


Figure: Ze et al.[1]

[1] Ze et al. '3D Diffusion Policy: Generalizable Visuomotor Policy Learning via Simple 3D Representations', RSS'24

Formal Description

Definition

Let $\mathcal{D} = \{\tau_1, \ldots, \tau_n\}$ be a set of n demonstrations, with $\tau_i = \{(s_1, a_1), \ldots, (s_{N_i}, a_{N_i})\}$ being state-action pair sequences of length N_i. To learn policy π, we minimize the negative log-likelihood for action $a \in \mathcal{A}$ given state $s \in \mathcal{S}$:

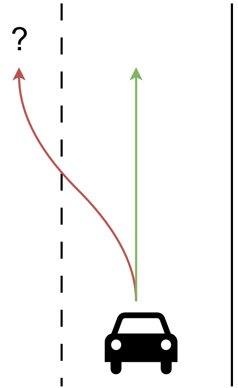
$$\mathcal{L}(\pi) = -\mathbb{E}_{(s,a)\sim p_{\mathcal{D}}}[\log \pi(a \mid s)]$$

Covariate Shift Problem

- Agent only sees expert states during training
- Encounters unseen states during deployment

$\Rightarrow$ Agent doesn't know how to return to demonstrated states

- High risk in safety-critical tasks
  - E.g. autonomous driving



Handling Covariate Shift

- **Dataset Aggregation (DAgger)**[1]
  1. Train policy on dataset
  2. Let policy play out
  3. Expert labels new data
  4. Dataset is aggregated
- **Robot-gated**[2]: robot predicts uncertainty
  - $\Rightarrow$ expert only queried when uncertainty is high
- Still resource and labor intensive

[1] Ross et al. 'A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning', AISTATS'11

[2] Zhang et al. 'Query-Efficient Imitation Learning for End-to-End Autonomous Driving', AAAI'17

DAgger

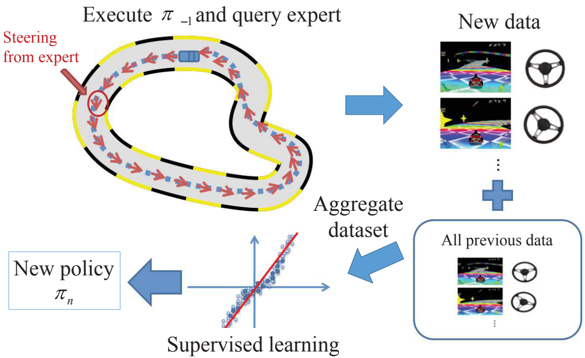


Figure: Osa et al.[1]

[1] Osa et al. 'An Algorithmic Perspective on Imitation Learning', Foundations and Trends® in Robotics 2018

Reinforcement Learning (RL)

- Agent interacts with environment to maximize cumulative reward
- Actions bringing agent closer to goal get rewarded
- Uses trial-and-error feedback

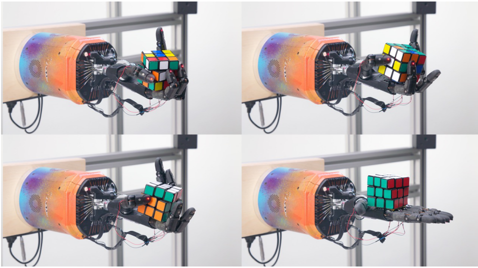


Figure: Ilge et al.[1]

[1] Ilge et al. 'Solving Rubik's Cube with a Robot Hand', arXiv preprint'19

Reinforcement Learning (cont.)

Definition

Let $R(s_t, a_t)$ be a reward function. Learn policy π that maximizes expected return:

$$\mathbb{E}\left[\sum_t \gamma^t R(s_t, a_t)\right],$$

where t is the time step and $\gamma \in [0; 1]$ a discount factor.

- Explorative training mitigates covariate shift
- But: RL needs reward function

Inverse Reinforcement Learning (IRL)

- Find reward function that explains the expert behavior
- **Input**: feature vectors $\phi(s_t, a_t)$
- **Output**: reward function

Definition

Let $R(s_t, a_t)$ be a reward function learned by linearly combining feature vectors $\phi(s_t, a_t)$ with weights w:

$$R(s_t, a_t) = w^\top \phi(s_t, a_t)$$

- Use reward function to train policy via reinforcement learning

Reward Ambiguity

- Infinite reward functions explain the same behavior

$\Rightarrow$ How do we select the best one?

$$\mu(\pi) = \mathbb{E}\left[\sum_t \gamma^t \phi(s_t)\right]$$

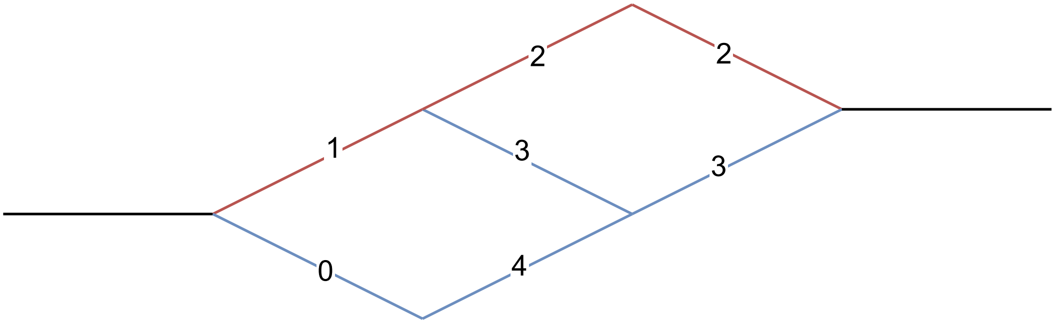
$$\text{Find } w \text{ s.t. } w^\top \mu(\pi_E) \geq w^\top \mu(\pi)\ \forall \pi$$

- **Max-margin**[1]: ensure function fits best by a margin
- **Max-entropy**[2]: optimize for highest entropy
  - $\Rightarrow$ encourages diverse behavior $\rightarrow$ robustness

[1] Ratcliff et al. 'Maximum Margin Planning', ICML'06

[2] Ziebart et al. 'Maximum Entropy Inverse Reinforcement Learning', AAAI'08

Max-entropy



Generative Adversarial Imitation Learning (GAIL)

- Inspired by Generative Adversarial Networks (GANs)
  - Discriminator: distinguish expert vs. agent behavior
  - Agent: deceive discriminator by imitating expert
- **Input**: state information
- **Output**: actions
- More robust than BC
- More sample efficient than IRL, but also more unstable

Mode Collapse

- Discriminator becomes too powerful too quickly
- Agent "collapses" on small range of actions

$\Rightarrow$ Insufficient exploration $\rightarrow$ stuck on local optimum

- To mitigate: Wasserstein distance, PacGAN[1],...

[1] Lin et al. 'PacGAN: The power of two samples in generative adversarial networks', NIPS'18

Imitation from Observation (IfO)

- Learn from observations only (no action labels)
- Tries to address scarce data availability
- Extension of BC, IRL, GAIL approaches
- More human-like learning
- Bypasses action space mismatch
  - ⇒ Allows learning from agents with different hardware

Context Translation

- Translate observation to robot context
  - ⇒ E.g. from third person to first person view
- Self-supervised representation learning
  - Encoder/Decoder architecture
  - Learn shared embeddings for different contexts

Context Translation (cont.)

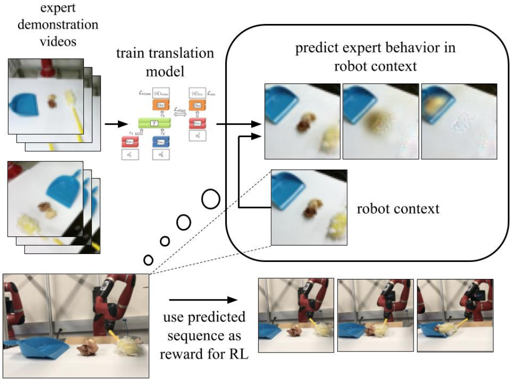


Figure: Liu et al.[1]

[1] Liu et al. 'Imitation from Observation: Learning to Imitate Behaviors from Raw Video via Context Translation', ICRA'18

DexMimicGen[1]

- Real-to-sim-to-real approach
- Data generation for dexterous manipulation
- Behavioral cloning on bimanual robots

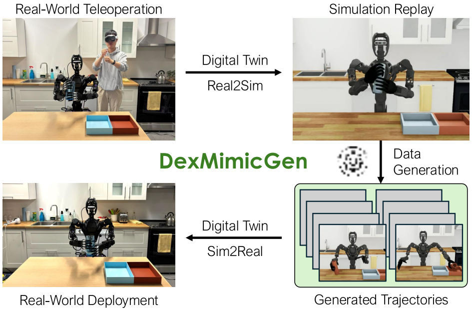


Figure: Jiang et al.[1]

[1] Jiang et al. 'DexMimicGen: Automated Data Generation for Bimanual Dexterous Manipulation via Imitation Learning', ICRA'25

Data Generation

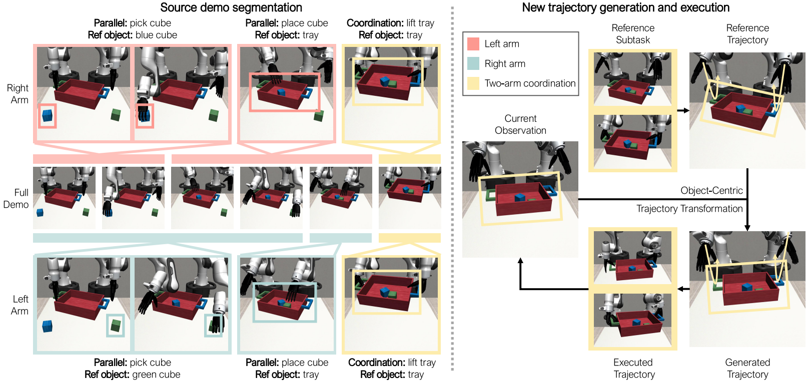


Figure: Jiang et al.[1]

[1] Jiang et al. 'DexMimicGen: Automated Data Generation for Bimanual Dexterous Manipulation via Imitation Learning', ICRA'25

RialTo[1]

- Assumption: Household stays mostly the same

⇒ Make training for specific environments easy

- 3D scene reconstruction
- Finetuned-RL in simulation

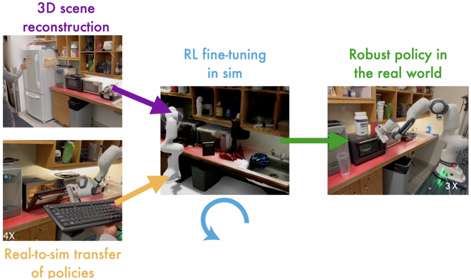


Figure: Torne et al.[1]

[1] Torne et al. 'Reconciling Reality Through Simulation: A Real-to-Sim-to-Real Approach for Robust Manipulation', RSS'24

3D Scene Reconstruction

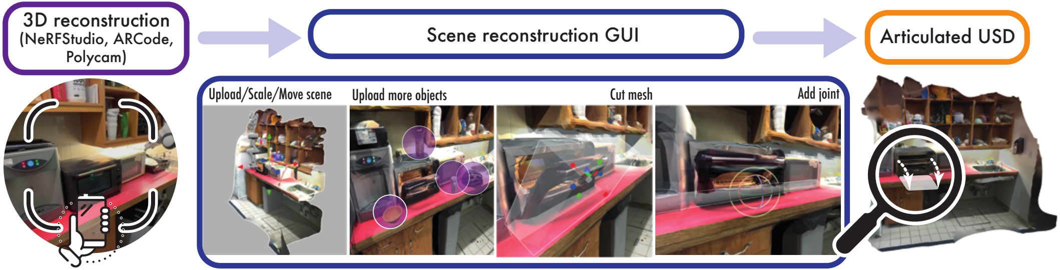
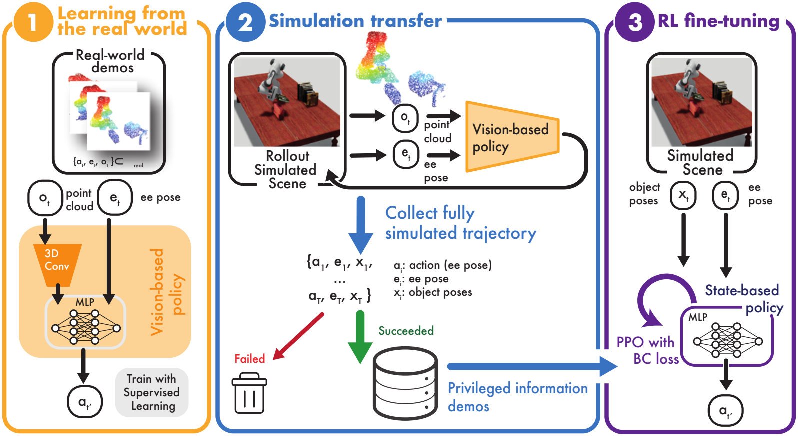


Figure: Torne et al.[1]

[1] Torne et al. 'Reconciling Reality Through Simulation: A Real-to-Sim-to-Real Approach for Robust Manipulation', RSS'24

System Overview[1]



[1]Torne et al. 'Reconciling Reality Through Simulation: A Real-to-Sim-to-Real Approach for Robust Manipulation', RSS'24

Conclusion

- Imitation Learning provides ways to learn from expert demonstrations
- Each approach addresses and has different challenges:
  - BC: simple but suffers from covariate shift
  - IRL: infers reward but ambiguous and resource intensive
  - GAIL: sample efficient but can be unstable
  - IfO: data availability, human-like, hardware-agnostic