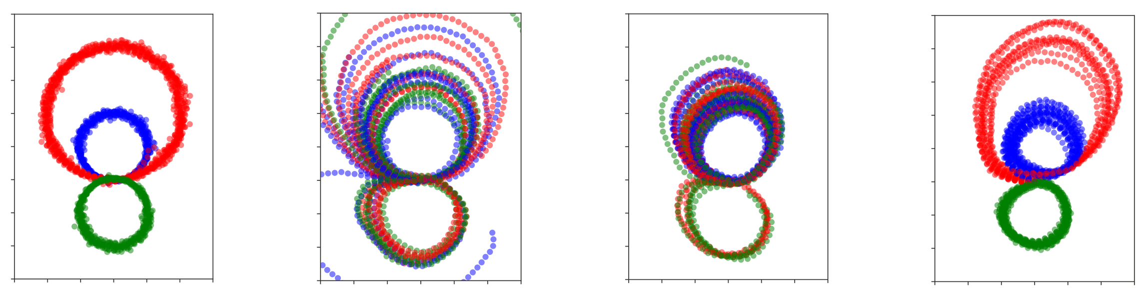
InfoGAIL: Interpretable Imitation Learning from Visual Demonstrations

Yunzhu Li[1], Jiaming Song[2] and Stefano Ermon[2]

[1]MIT Computer Science and Artificial Intelligence Laboratory
[2]Stanford Artificial Intelligence Laboratory

Introduction

- **Imitation Learning** mimic expert behavior without access to an explicit reward signal.
- Expert demonstrations provided by humans, however, often show *significant variability*.



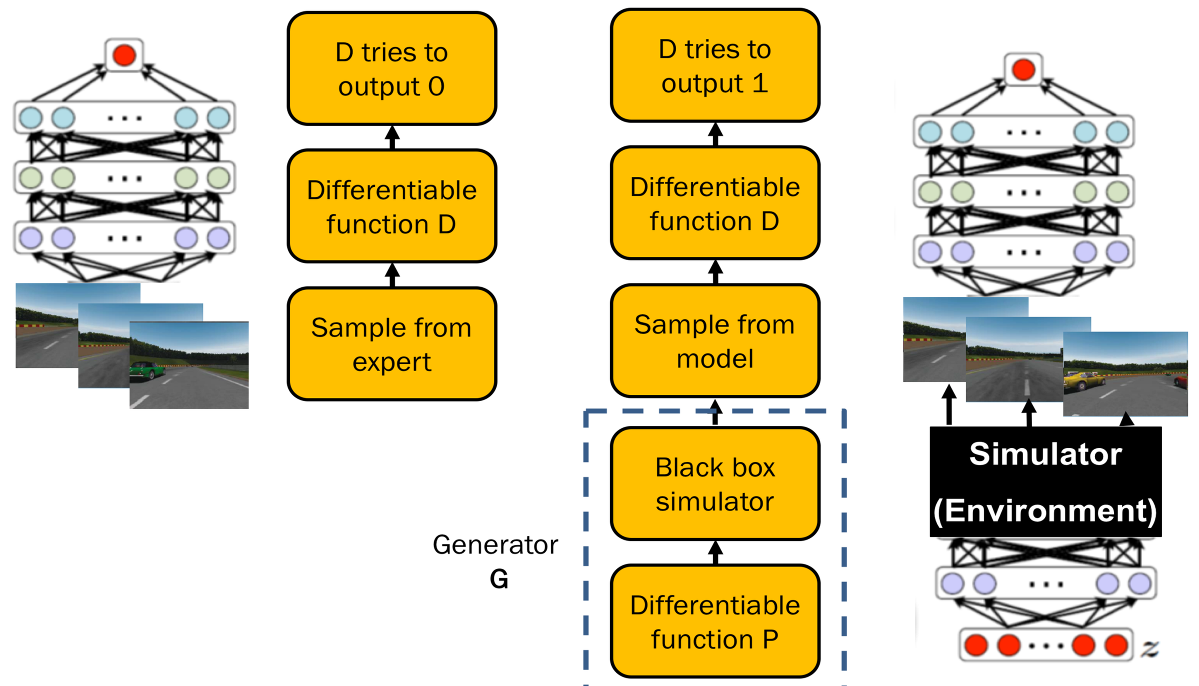
(a) Expert **(b)** BC **(c)** GAIL **(d)** InfoGAIL

- BC - deviates due to compounding errors.
- GAIL - fails to capture the latent structure.

Our method

- **Can disentangle different behaviors (modes).**
- **Can do imitation learning from raw images.**
- **Can be used to anticipate actions.**

Generative Adversarial Imitation Learning



Ho and Ermon, *Generative Adversarial Imitation Learning*

- Discriminator D tries to distinguish between expert trajectories and ones from policy π.
- Policy π tries to fool the discriminator.

$$\min_{\pi} \max_{D} \mathbb{E}_{\pi}[\log D(s,a)] + \mathbb{E}_{\pi_E}[\log(1 - D(s,a))] \quad (1)$$

Interpretable Imitation Learning

- Introduce a **latent variable/code** c.
- Imitate while maximizing the mutual information (MI) between
  (1) the latent code
  (2) the observed trajectories

$$\min_{\pi,Q} \max_{D} \mathbb{E}_{\pi}[\log D(s,a)] + \mathbb{E}_{\pi_E}[\log(1 - D(s,a))] - \lambda_1 L_I(\pi, Q) \quad (2)$$

- $L_I(\pi, Q)$ is a variational lower bound of MI

$$L_I(\pi, Q) = \mathbb{E}_{c\sim p(c), a\sim\pi(\cdot|s,c)}[\log Q(c|\tau)] + H(c) \quad (3)$$

where Q is an approximation to the posterior.

Algorithm 1 InfoGAIL

Input: Initial parameters of policy, discriminator and posterior approximation $\theta_0, \omega_0, \psi_0$; expert trajectories $\tau_E \sim \pi_E$ containing state-action pairs.
Output: Learned policy π_θ
for $i = 0, 1, 2, \dots$ **do**
  Sample latent codes: $c_i \sim p(c)$
  Sample trajectories: $\tau_i \sim \pi_{\theta_i}(c_i)$.
  Sample state-action pairs $\chi_i \sim \tau_i$ and $\chi_E \sim \tau_E$.
  Update ω_i to ω_{i+1} by ascending with gradients
  $$\Delta_{\omega_i} = \hat{\mathbb{E}}_{\chi_i}[\nabla_{\omega_i} \log D_{\omega_i}(s,a)] + \hat{\mathbb{E}}_{\chi_E}[\nabla_{\omega_i} \log(1 - D_{\omega_i}(s,a))]$$
  Update ψ_i to ψ_{i+1} by descending with gradients
  $$\Delta_{\psi_i} = -\lambda_1 \hat{\mathbb{E}}_{\chi_i}[\nabla_{\psi_i} \log Q_{\psi_i}(c|s,a)]$$
  Update θ_i to θ_{i+1}, using TRPO with the objective:
  $$\hat{\mathbb{E}}_{\chi_i}[\log D_{\omega_{i+1}}(s,a)] - \lambda_1 L_I(\pi_{\theta_i}, Q_{\psi_{i+1}})$$
end for

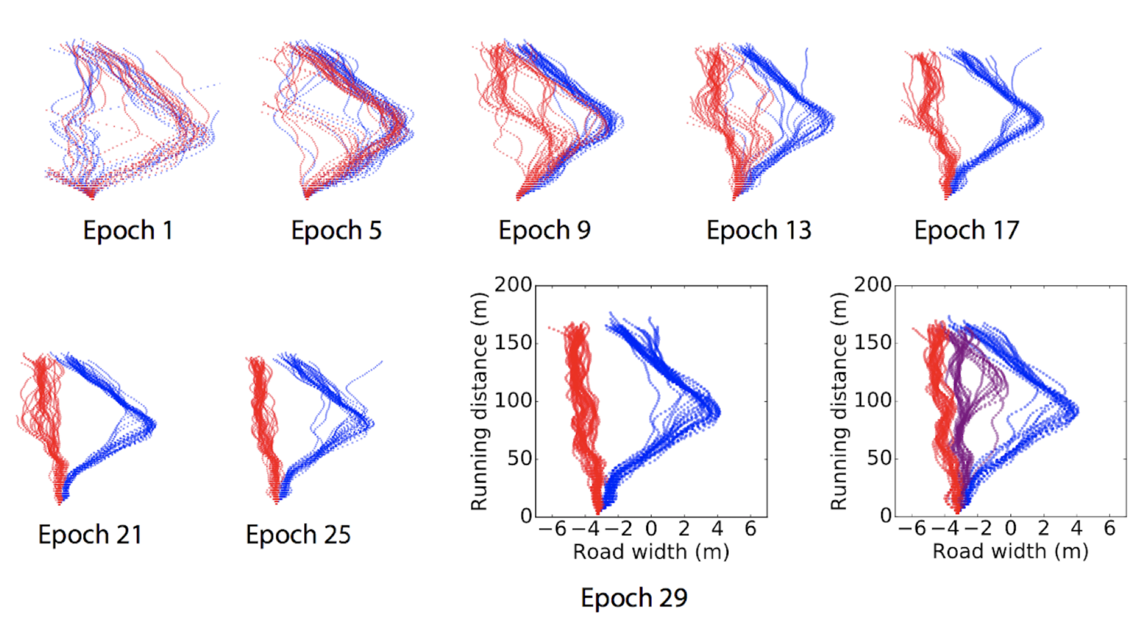


Figure: Visualization of different training stages.

InfoGAIL Training

- **Reward Augmentation**
  Incorporate prior knowledge by adding state-based incentives $\eta(\pi_\theta) = \mathbb{E}_{s\sim\pi_\theta}[r(s)]$.
- **Improved Objective**
  Using WGAN to alleviate the problems of
  (1) vanishing gradient
  (2) mode collapse

$$\min_{\theta,\psi} \max_{\omega} \mathbb{E}_{\pi_\theta}[D_\omega(s,a)] - \mathbb{E}_{\pi_E}[D_\omega(s,a)] - \lambda_0 \eta(\pi_\theta) - \lambda_1 L_I(\pi_\theta, Q_\psi) \quad (4)$$

- **Variance Reduction**
  Baselines, Replay Buffers, etc.
- **Transfer Learning for Visual Inputs**
  Extract features from a pre-trained network.

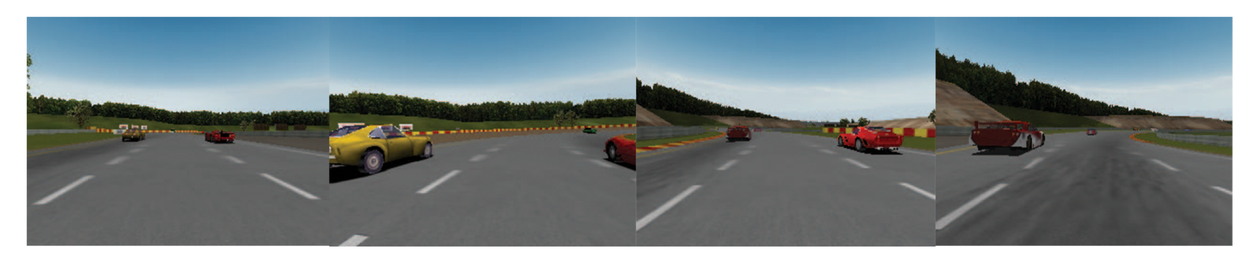


Figure: Transfer learning for handling visual inputs.

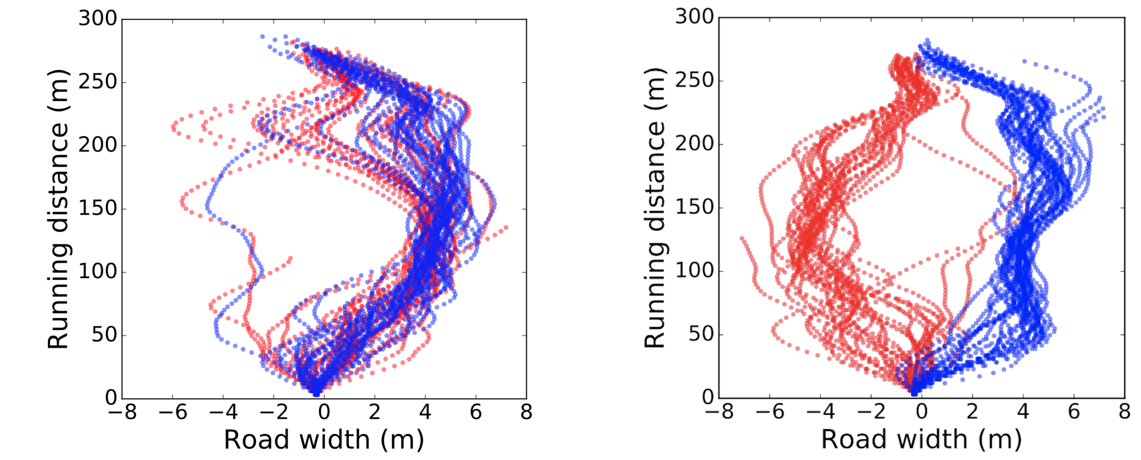
Experiments on Self-Driving

Interpretable Imitation Learning via Vision

- Using TORCS - a driving simulator
- Vision as only source of perceptual inputs

The learned policy

- successfully distinguishes expert behaviors.
- produces interpretable representations from high-dimensional visual behavioral data.
- imitates each mode accordingly.
- low-level actions controlled by specifying high-level latent codes.



(a) GAIL **(b)** InfoGAIL

Figure: Passing a car.

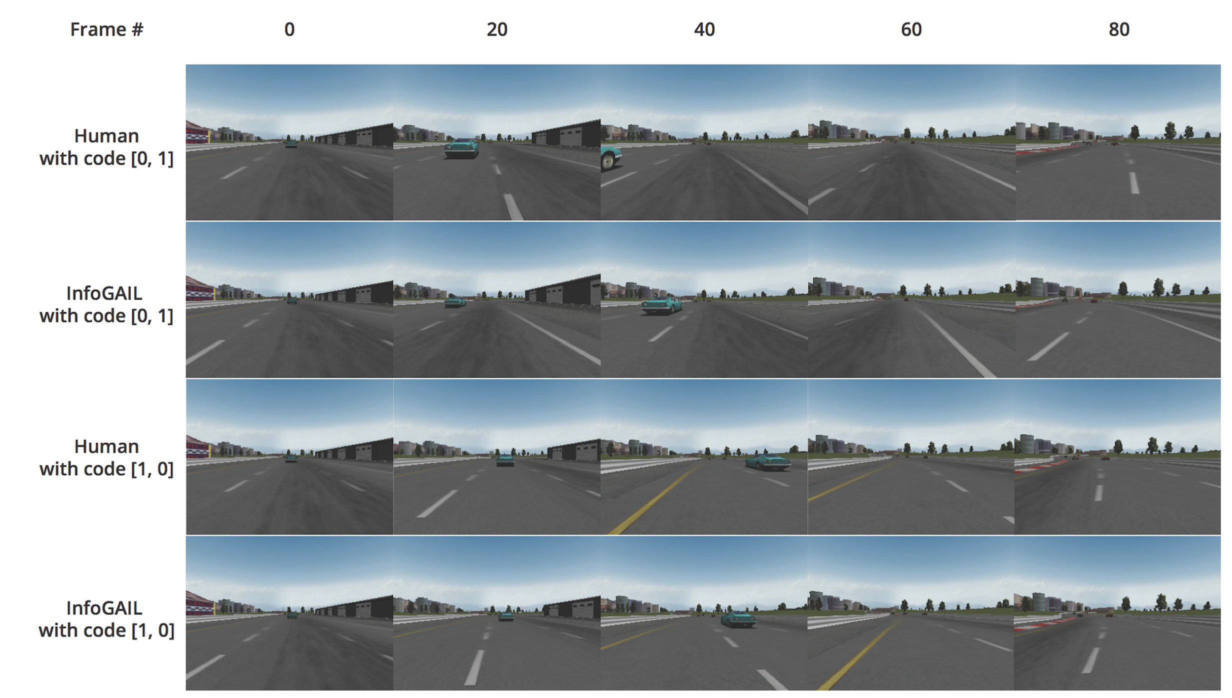


Figure: Visual inputs used for passing a car.

Table: Predictive accuracy

Method	Acc.
Chance	50%
K-means	55.4%
PCA	61.7%
InfoGAIL (Ours)	**81.9%**
SVM	85.8%
CNN	**90.8%**

Table: Ablation study

Method	Rollout dist.
Behavior Cloning	701.83
GAIL	914.45
InfoGAIL\RB	1031.13
InfoGAIL\RA	1123.89
InfoGAIL\WGAN	1177.72
InfoGAIL (Ours)	**1226.68**
Human	1203.51

Code : https://github.com/ermongroup/infogail

Acknowledgements

Toyota Research Institute (TRI) provided funds to assist the authors with their research. This research was also supported by Intel Corporation, FLI and NSF grants 1651565, 1522054, 1733686.