InfoGAIL: Interpretable Imitation Learning from Visual Demonstrations
Yunzhu Li\textsuperscript{1}, Jiaming Song\textsuperscript{2} and Stefano Ermon\textsuperscript{2}
\textsuperscript{1}MIT Computer Science and Artificial Intelligence Laboratory
\textsuperscript{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.

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

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

Expert demonstrations provided by humans,
\[ \pi \]
tries to fool the discriminator.
\[ E \left[ \log D(s, a) + \mathbb{E}_z \left[ \log (1 - D(s, a)) \right] \right] \]

\[ -\lambda_1 L_i(\pi, Q) \]

\[ \min_{Q} \frac{1}{d} \mathbb{E}_i \left[ \log D(s, a) + \mathbb{E}_z \left[ \log (1 - D(s, a)) \right] \right] \]

InfoGAIL: Interpretable Imitation Learning from Visual Demonstrations

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

\[ \min_{Q} \mathbb{E}_i \left[ \log D(s, a) + \mathbb{E}_z \left[ \log (1 - D(s, a)) \right] \right] \]

\[ -\lambda_1 L_i(\pi, Q) \]

\[ L_i(\pi, Q) \] is a variational lower bound of MI.

\[ L_i(\pi, Q) = E_c p(c) a \sim \pi(c) \left[ \log Q(c|\tau) \right] + H(c) \]

\( \lambda \) 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 \( \tau \), \( \tau_E \) containing state-action pairs.
Output: Learned policy \( \pi_\theta \)

\[ \text{for } i = 0 \text{ to } m \text{ do} \]

Sample latent codes: \( c \sim p(c) \)
Sample trajectories: \( \tau_i \sim \pi(c) \)
Sample state-action pairs: \( \tau_E \)
Update \( \omega_1 \) to \( \omega_{i+1} \) by ascending with gradients.

\[ \Delta_\omega = \mathbb{E}_i \left[ \nabla_c \log D_i(s, a) E_c p(c) \nabla_w \log (1 - D_i(s, a)) \right] \]

Update \( \psi_1 \) to \( \psi_{i+1} \) by descending with gradients.

\[ \Delta_\psi = -\lambda_2 \mathbb{E}_i \left[ \nabla_w \log Q_i(c|s) \right] \]

Update \( \theta_i \) to \( \theta_{i+1} \) using TRPO with the objective:

\[ \hat{E}_i \left[ \log D_{i+1}(s, a) \right] - \lambda_1 L_i(\pi_\theta, Q_{\psi_{i+1}}) \]

end for

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.

Figure: Visualization of different training stages.

InfoGAIL Training
• Reward Augmentation
Incorporate prior knowledge by adding state-based incentives.
• Improved Objective
Using WGAN to alleviate the problems of (1) vanishing gradient and (2) mode collapse.

\[ \min_{\pi} \mathbb{E}_i \left[ D_i(s, a) - D_i(s, a) \right] - \lambda_1 \log i(s) \]

\[ -\lambda_1 L_i(\pi_\theta, Q_{\psi_{i+1}}) \]

Figure: Passing a car.

Figure: Visual inputs used for passing a car.

Table: Predictive accuracy

<table>
<thead>
<tr>
<th>Method</th>
<th>Acc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior Cloning</td>
<td>701.83</td>
</tr>
<tr>
<td>GAIL</td>
<td>914.45</td>
</tr>
<tr>
<td>InfoGAIL</td>
<td>1031.13</td>
</tr>
<tr>
<td>InfoGAIL</td>
<td>1123.89</td>
</tr>
<tr>
<td>InfoGAIL</td>
<td>1177.72</td>
</tr>
<tr>
<td>InfoGAIL</td>
<td>1226.68</td>
</tr>
<tr>
<td>InfoGAIL</td>
<td>1203.51</td>
</tr>
</tbody>
</table>

Table: Ablation study

<table>
<thead>
<tr>
<th>Method</th>
<th>Rollout dist.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavior Cloning</td>
<td>701.83</td>
</tr>
<tr>
<td>GAIL</td>
<td>914.45</td>
</tr>
<tr>
<td>InfoGAIL</td>
<td>1031.13</td>
</tr>
<tr>
<td>InfoGAIL</td>
<td>1123.89</td>
</tr>
<tr>
<td>InfoGAIL</td>
<td>1177.72</td>
</tr>
<tr>
<td>InfoGAIL</td>
<td>1226.68</td>
</tr>
<tr>
<td>InfoGAIL</td>
<td>1203.51</td>
</tr>
</tbody>
</table>

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.