Reinforcement and Imitation Learning
for diverse visuomotor skills.

Summary:

1. Use demonstration to speed up training of RL policies in simulation.

2. Leverage privileged and task-specific information to speed up training in simulation.

3. Some success in zero-shot sim2real transfer.

Combine RL and IL rewards:

1. $r(s_t, a_t) = \lambda r_{gail}(s_t, a_t) + (1-\lambda) r_{task}(s_t, a_t)$ $\lambda \in [0,1]$.

2. They argue that this new. fun. helps the agents to solve tasks that neither RL or IL can solve.

Leveraging physical states in simulation:

1. Because they train in simulation, the full state of the system is available.

2. Even tho. this info. is unavailable at test time on a real robot, they can use it during training:

demonstration as a curriculum:

1. They alter the initial state dist. with the demo. state.
2. They build a curriculum that contains cluster of states in different stages of a task.
3. During training, with prob. ϵ, they start the episode from a random initial state, and with prob $1-\epsilon$, they uniformly select a cluster and initialize the episode with a demonstration state from that cluster.

learning value fun. from states

1. Since the value fun. is not used at test time, we can leverage privileged info. to speed up its training.
2. The input to the value fun. is low-level physical states

object-centric discriminator

1. task-specific features are used as input to the discriminator of GAIL.
2. they found that obj-centric representation (e.g. positions of the objs) provide salient signals to the discriminator.
3. If they use the states of the robot arm, the discriminator focuses on irrelevant aspects of the behavior of the controller and are detrimental to training the policy.
4. The construction of the obj-centric rep. requires domain knowledge.

state prediction auxiliary tasks

1. they found this helps with accelerating training.

sim2real policy transfer

1. Randomize position of camera and also its orientation
2. Other parameters were set by hands.

Questions:

1. They mention in the related work section that they do not need demonstrator action. This seems to contradict their GAIL formulation ?