Training Robots to Evaluate Robots: Example-based Interactive Reward Functions

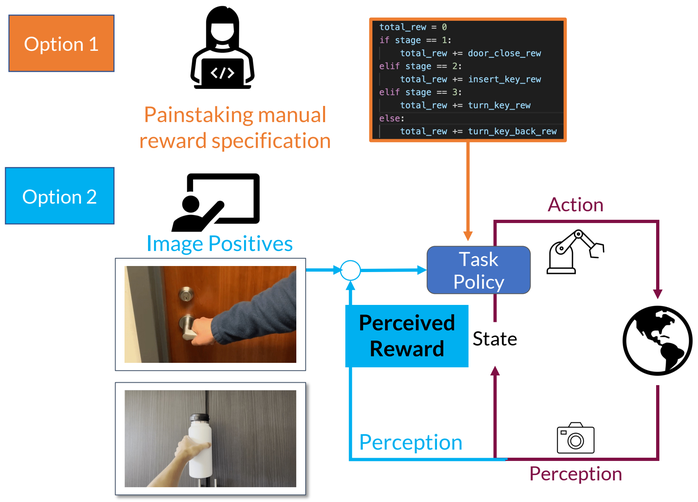
Kun Huang, Edward S. Hu, Dinesh Jayaraman

Penn Engineering | GRASP Laboratory
General Robotics, Automation, Sensing & Perception Lab

Motivation: Scalable Task Specification

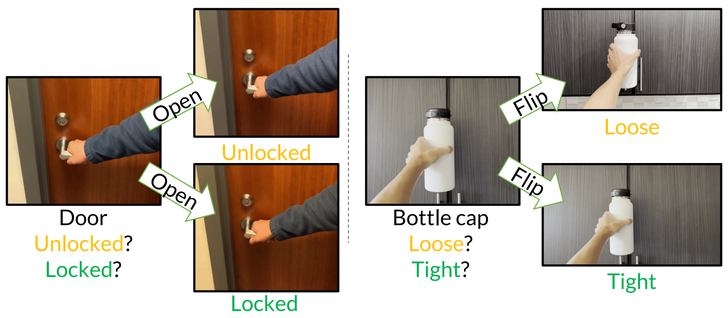
How can we scalably specify tasks like these to a reinforcement learning agent?

Specifying Tasks with Image Examples?



Example-based task specification scales well, but …

Perceiving Task Rewards Can Be Hard!

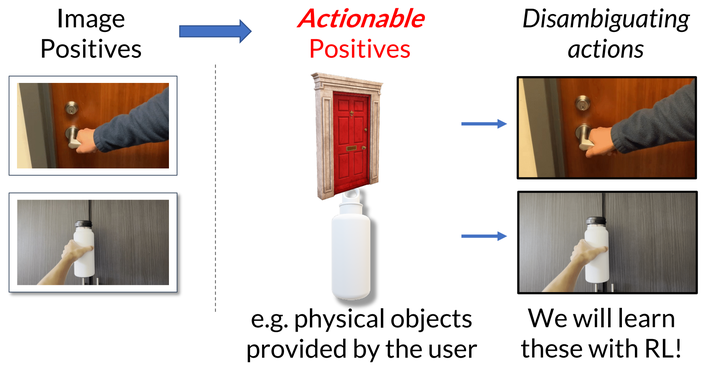


Physical interaction reveals task success/failure!

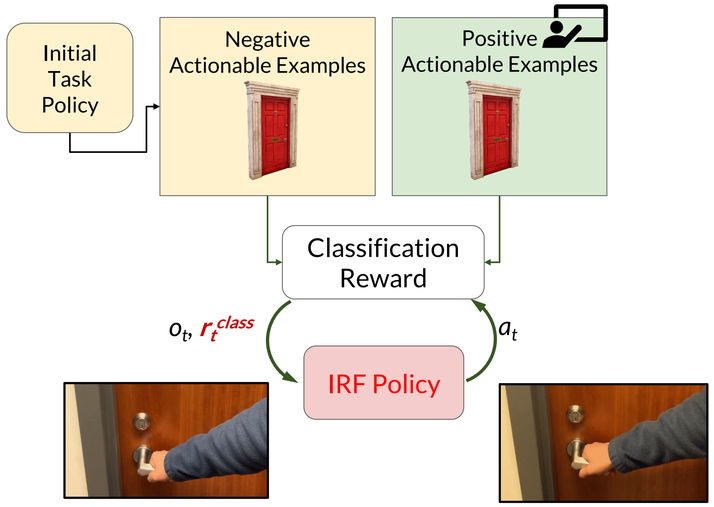
Key Idea: Train "interactive reward function" policies that interactively evaluate task policy outcomes!

But how to train these?

Solution: ~~Image Snapshots~~ Actionable Examples

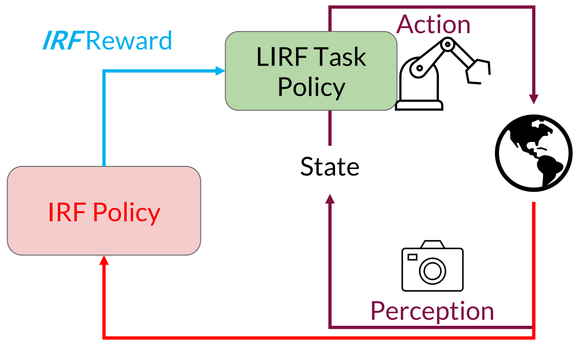


IRF Policy Training



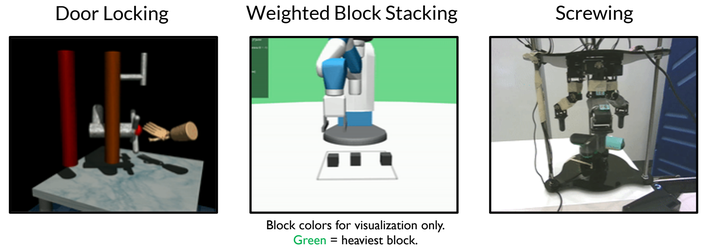
IRF policies are trained to interact with the objects to reveal task failure or success.

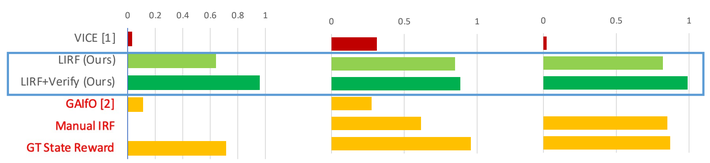
Learning from Interactive Reward Functions



IRF policies can provide task completion rewards to improve task policy learning!

Experiments





Privileged baselines – ground truth rewards, kinesthetic demonstrations, hand engineered reward policies!

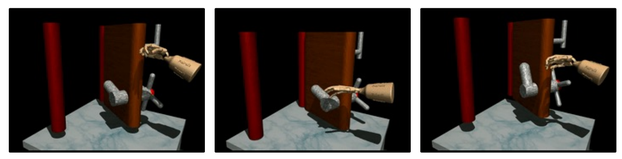
Only policies trained on IRF rewards solve partially observed tasks.

What do IRF policies do?

Evaluating successes

Evaluating failures

Using IRFs at test time for verification



Red ball appears: LIRF policy execution;
Green ball appears: IRF policy execution