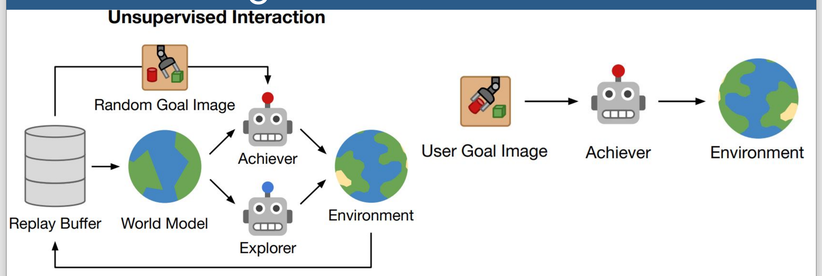
Discovering and Achieving Goals via World Models

Russell Mendonca[1] *, Oleh Rybkin[2] *, Kostas Daniilidis[2], Danijar Hafner[3], Deepak Pathak[1]

* equal contribution [1] Carnegie Mellon University [2] University of Pennsylvania [3] University of Toronto

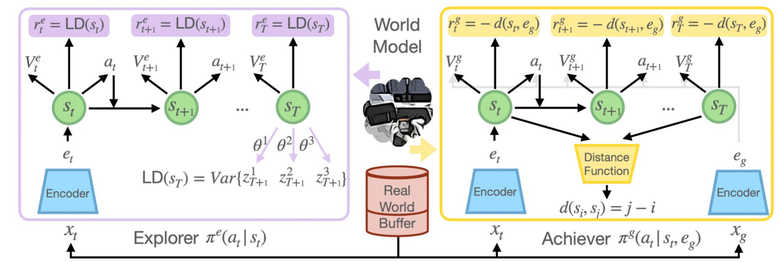
1 Introducing LEXA



1. Learns a world model without any supervision, trains two policies (explorer and achiever) in imagination.
2. The explorer finds new images via model disagreement and the achiever learns to reliably reach them.
3. Once trained, the achiever reaches user-specified goals zero-shot without further training at test time.

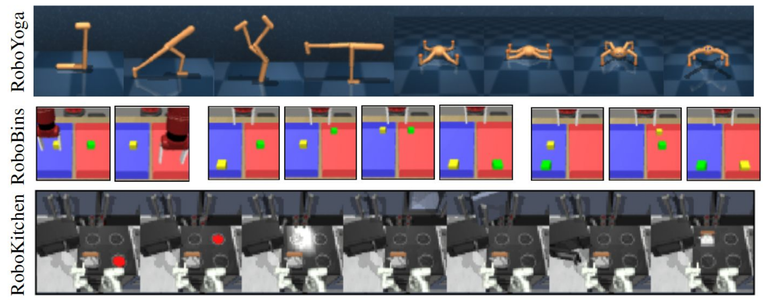
2 Unsupervised Discovery

describe world model, explorer, achiever



1. Explorer is trained on imagined latent state rollouts of the world model to maximize the disagreement objective
2. Achiever policy is conditioned on goals and trained on imagined rollouts to minimize a distance function. Goals are sampled randomly from replay buffer images.
3. Trains a temporal distance function, using the imagined rollouts of the achiever and predicst the number of time steps between states.

3 Goal Image Benchmark

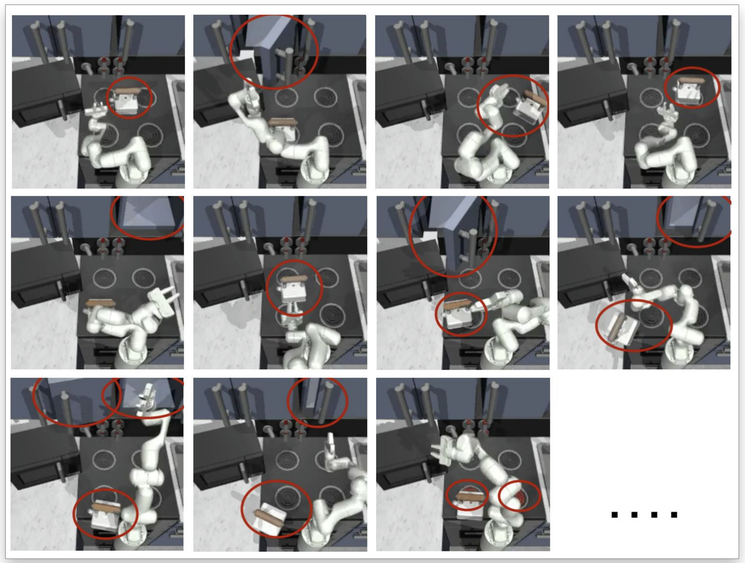


RoboYoga : Goals correspond to different body poses for Walker and Quadruped, such as lying down, standing up, and balancing.

RoboBins : Sawyer robotic arm with two bins, and two differently colored blocks. Tasks include reaching, manipulating one block, manipulating both blocks.

RoboKitchen: Goals specify single or multiple tasks that the franka can do in the kitchen.

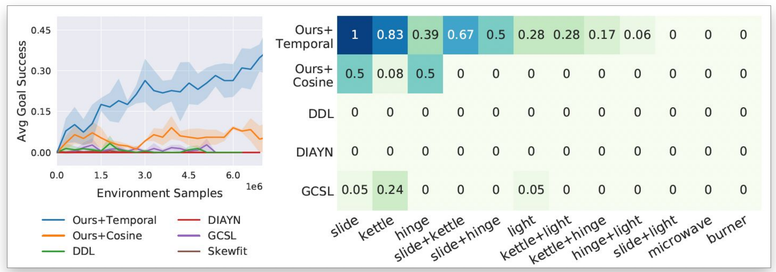
4 Diverse Exploration



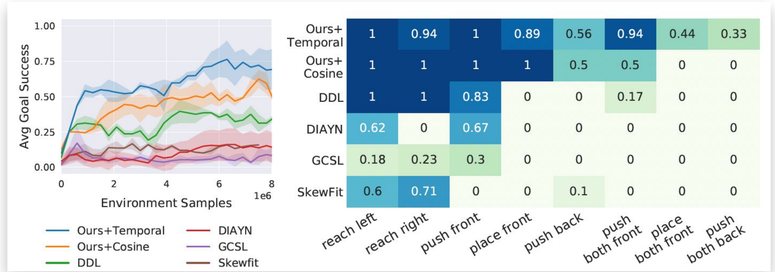
Images discovered by the explorer policy on the kitchen environment

5 Evaluation on Manipulation Envs

Kitchen evaluation

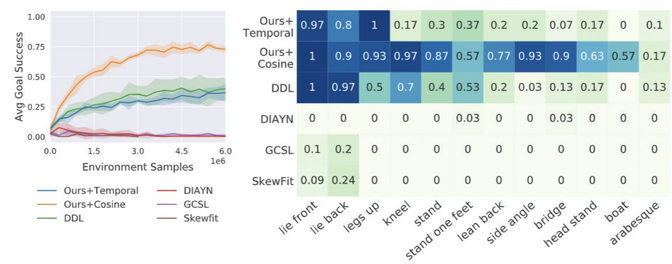


RoboBin evaluation

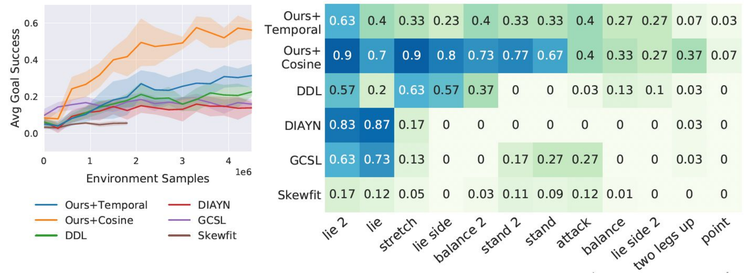


6 Evaluation on Pose Matching Envs

Walker evaluation



Quadruped evaluation



Website with videos: https://orybkin.github.io/lexa/