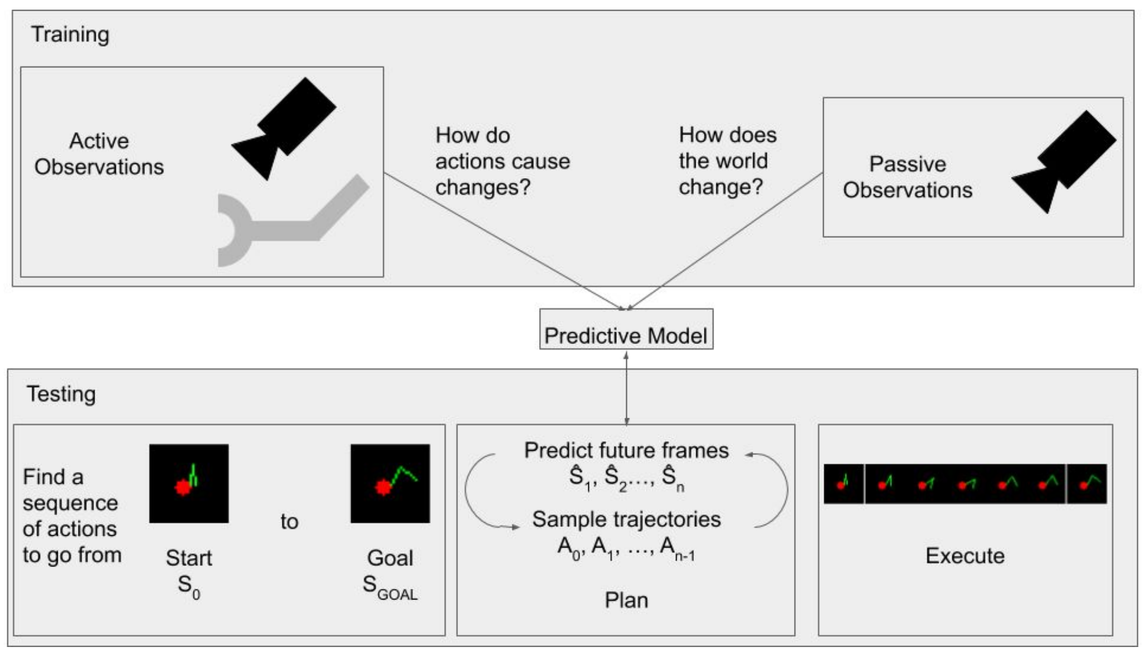
Visual Planning with Semi-Supervised Stochastic Action Representations

Penn Engineering
GRASP LABORATORY
General Robotics, Automation, Sensing & Perception Lab

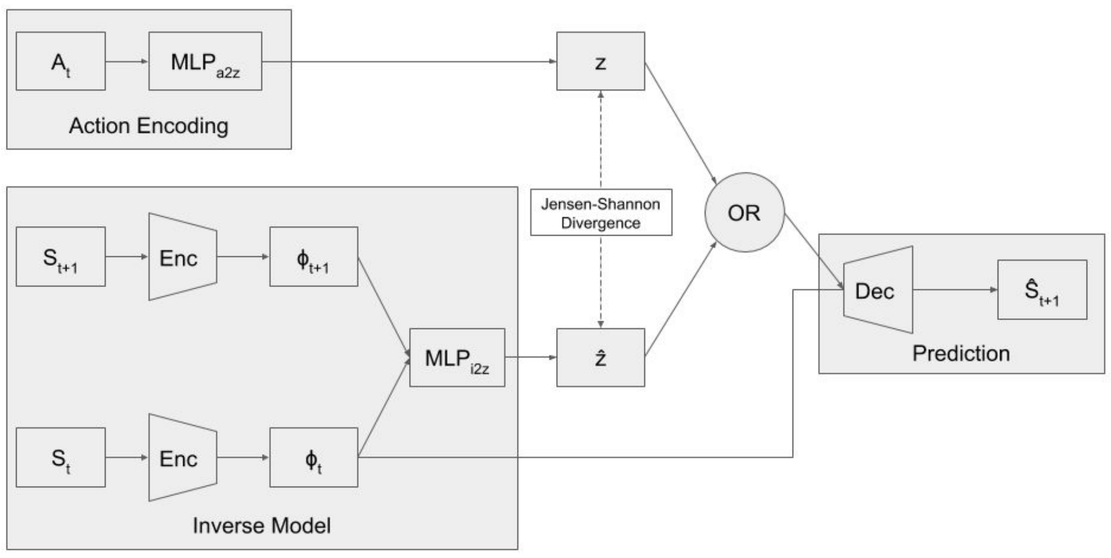
Karl Schmeckpeper, David Han,
Kostas Daniilidis, Oleh Rybkin

Penn Engineering
GRASP LABORATORY
General Robotics, Automation, Sensing & Perception Lab

Motivation

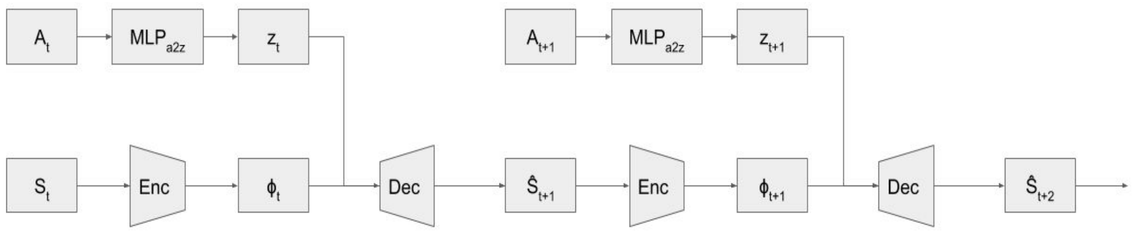


Semi-Supervised Action Representations



Our model generates stochastic action enbeddings by either encoding the true action and by using a visual encoder. The two representations are pulled together with a Jensen-Shannon divergence loss. This allows to train the model via the visual encoder even when only a portion of the observations have corresponding actions.

Planning



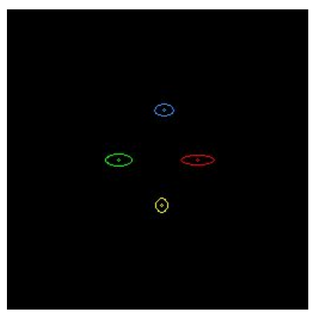
We use our learned forward model with the cross entropy method to plan sequences of actions that reach the goal image.

Algorithm 1 The Cross Entropy Method

function CrossEntropyMethod(X, μ_0, σ_0^2)
  for i = 1 to num_iters **do**
    $trajectories[1:N] \leftarrow$ sample(X, (μ_{i-1}, σ_{i-1}^2))
    $dists[1:N] \leftarrow$ minDistance($trajectories[1:N]$)
    $trajectories \leftarrow$ sortByDists($trajectories$, $dists$)
    $\mu_i \leftarrow$ mean($trajectories[1:N_{elite}]$)
    $\sigma_i^2 \leftarrow$ variance($trajectories[1:N_{elite}]$)
  end for
  return sample(X, (μ_{num_iters}, $\sigma_{num\_iters}^2$))
end function

Experiments

Our model learns to map the action representations z and $\hat{z}$ to the same locations in the latent space, visualized in a two dimensional action latent space



Action encoder

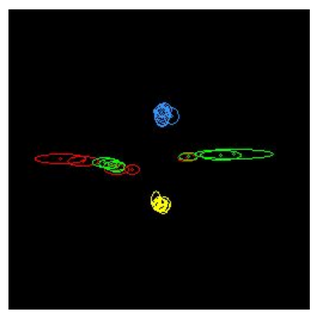
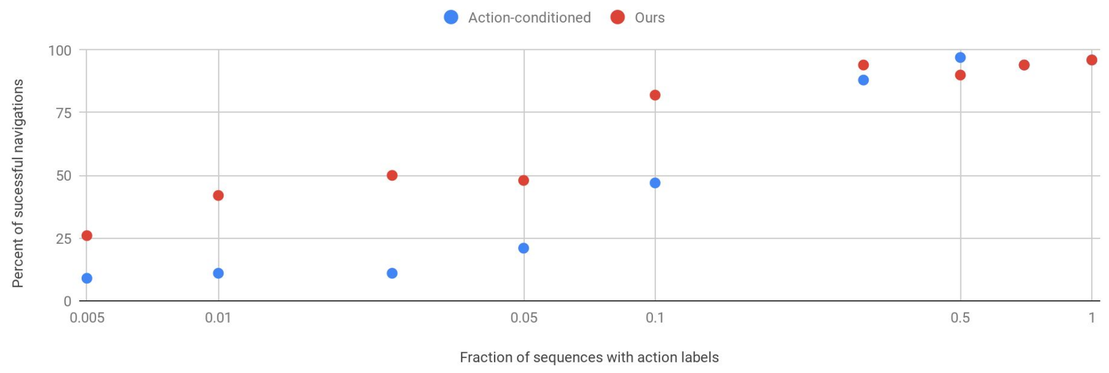
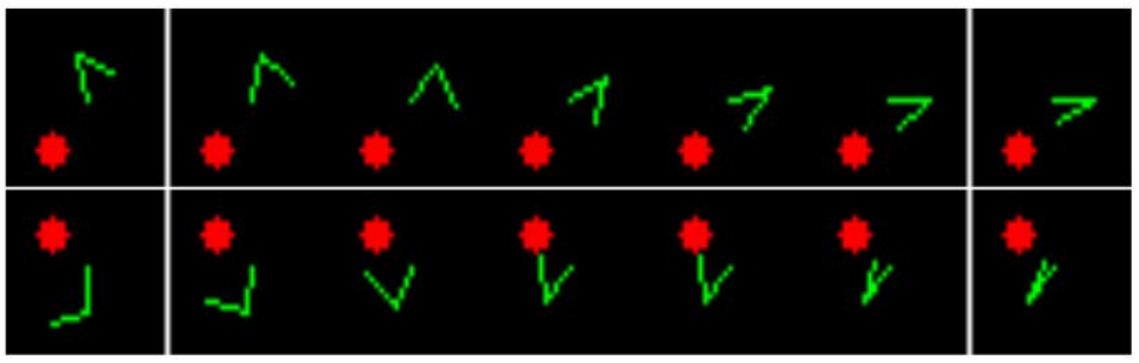


Image to action encoder

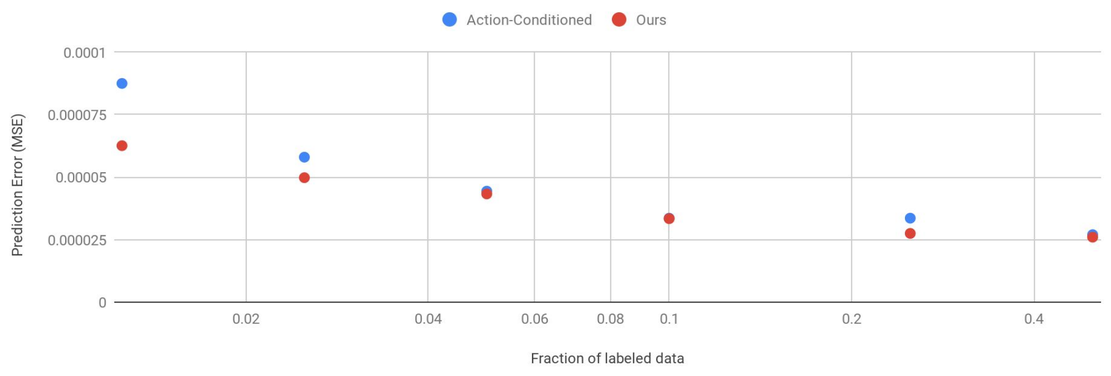
Planning success rate in the reacher environment



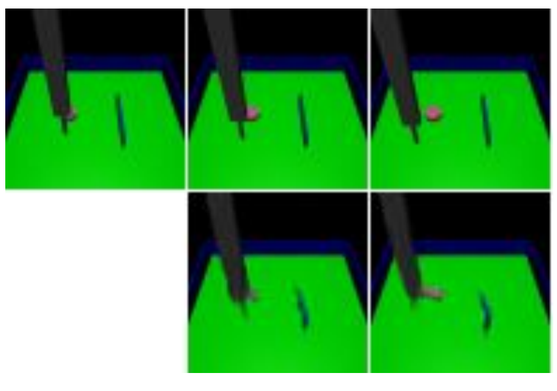
Our model learns a representation sufficient to plan executable trajectories for a 2-dof reacher.
As the amount of sequences with action labels decreases, the performance of our semi-supervised model degrades slower than a fully supervised model



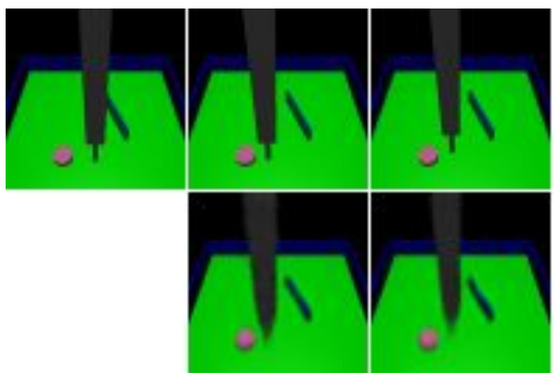
Prediction error in Mujoco



Our model was able to learn to predict future frames of video in a more complicated Mujoco simulation.
Even with only ~500 sequences of data with action labels, our system was able to generate realistic predictions of future frames.



Actioned-conditioned



Ours