Supplemental Materials

John D. Co-Reyes [*1] **YuXuan Liu** [*1] **Abhishek Gupta** [*1] **Benjamin Eysenbach** [2] **Pieter Abbeel** [1] **Sergey Levine** [1]

1. Experimental Details

For all experiments, we parameterize $\pi_{\theta_{PD}}$ and π_e as a three-layer fully connected neural networks with 400, 300, 200 hidden units and ReLU activations. The policies output either categorical or Gaussian distributions. The encoder is a two-layer bidirectional-LSTM with 300 hidden units, and we mean-pool over LSTM outputs over time before applying a linear transform to produce parameters of a Gaussian distribution. We use an 8-dimensional diagonal Gaussian distribution for z. The state decoder is a single-layer LSTM with 256 hidden units that conditions on the initial state and latent z, to output a Gaussian distribution over trajectories. We use trajectories of length $T = 19$, and plan over $K = 2048$ random latent sequences. We use horizons $H = 380$, $H_{MPC} = 5$, $H_e = 5$ for the 2D navigation task, $H = 950$, $H_{MPC} = 20$, $H_e = 10$ for the wheeled locomotion task, and $H = 950$, $H_{MPC} = 10$, $H_e = 10$ for the object manipulation task. These values were chosen empirically with a hyperparameer sweep.

2. Baseline Details

TRPO / VIME We used the rllab TRPO implementation, OpenAI VIME implementation with a batch size of 100 * task horizon and step size of 0.01.

MPC We use a learning rate of 0.001 and batch size of 512. The MPC policy simulates 2048 paths each time it is asked for an action. We verified correctness on half-cheetah.

Option Critic We use a version of Option Critic that uses PPO instead of DQN. We swept over number of options, reward multiplier, and entropy bonuses. We verified correctness on cartpole, hopper, and cheetah.

Feudal / A3C The Feudal and A3C implementations are based on chainerRL. We swept over the parameters β, t_{max}, and gradient clipping.

[*]Equal contribution [1]University of California, Berkeley [2]Google Brain. Correspondence to: John D Co-Reyes <jcoreyes@berkeley.edu>, YuXuan Liu <yuxuanliu@berkeley.edu>.

Proceedings of the 35^{th} International Conference on Machine Learning, Stockholm, Sweden, PMLR 80, 2018.