Thanard Kurutach, Ignasi Clavera, Yan Duan, Aviv Tamar, Pieter Abbeel

*Berkeley AI Research, UC Berkeley, CA*

**Goal**

To deploy RL in real world, we need to be sample efficient.

**Previous Model-Based Approaches**

- **Linear:**
  - Bagwell and Schneider 2001; Abbeel et al. 2006; Levine and Abbeel 2014; Watter et al. 2015; Levine et al. 2009a; Kumar et al. 2016

- **GP:**
  - Reichardt et al. 2008; Ho et al. 2007; Osawa and Doya 2011

- **NN:**
  - Silver et al. 2016, 2017; Mine et al. 2015; Schulman et al. 2015, 2016

**Limitations**

- Relatively simple domains: Nguyen and Williams 1999; Deisenroth and Rasmussen 2010; Silver et al. 2016; Pinto and Gupta 2016; Levine et al. 2016b; Finn and Levine 2017; Nair et al. 2017

- Combine with MF methods: On et al. 2016; Oh et al. 2015; Heess et al. 2015; Nagabandi et al. 2017

- Domain specific: Finn and Levine 2017; Nair et al. 2017

- Batch RL: Depeweg et al. 2017; Matos et al. 2017

**Successes in Deep RL**

**Contribution**

- Achieve MF performance with **100x** less data.
- Reduce model bias using model ensemble.
- Avoid numerical instability and bad local optima using TRPO.

**Data Efficiency Comparison**

- **Swimmer**
- **Snake**
- **Half-Cheetah**
- **Hopper**
- **Ant**
- **Humanoid**

**Environments**

- **Swimmer**
- **Half-Cheetah**
- **Ant**
- **Humanoid**

**Size of Ensemble**

- **Snake**
- **Half-Cheetah**

**Learn More**

*Videos: https://sites.google.com/view/me-trpo*