Example-Driven Model-Based Reinforcement Learning for Solving Long-Horizon Visuomotor Tasks
Bohan Wu, Suraj Nair, Li Fei-Fei†, Chelsea Finn† († Equal Advising) Stanford University

Long-Horizon Vision-Based Manipulation
Demands learning a repertoire of visuomotor skills that are:

- Robust
- High Success Rates
- Persistent
- Closed-loop and Reactive

Key Insight
Use human-provided example images as supervision to learn a repertoire of skills, groundings, and success detectors

- Provides supervision signal for reward learning in RL
- Enables skill grounding in long-horizon task planning
- Encourages closed-loop visuomotor control

Performance Evaluation

- Successful trials (out of 20) and success rates (%)

<table>
<thead>
<tr>
<th></th>
<th>EMBR (Ours)</th>
<th>EMBR w/o (f_{\text{ae}})</th>
<th>EMBR w/o (f_T)</th>
<th>EMBR w/o (f_T) (Qt-Opt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg. Skill Success</td>
<td>96.5%</td>
<td>89.1%</td>
<td>85.9%</td>
<td></td>
</tr>
<tr>
<td>Avg. Long-Horizon Task Success</td>
<td>85.0%</td>
<td>70.0%</td>
<td>58.3%</td>
<td></td>
</tr>
</tbody>
</table>

Kalashnikov et al. QT-Opt: Scalable Deep Reinforcement Learning for Vision-Based Robotic Manipulation, CoRL’18

Real-Robot Experiments: Long-Horizon Tasks with Novel Objects from Raw Image Observations

Example-Driven Model-Based Reinforcement Learning (EMBR)
A framework for learning grounded visuomotor skills sequenced by symbolic planners to complete long-horizon tasks.

- Step 1: Learn a repertoire of skills with example-driven model-based reinforcement learning (EMBR)
- Step 2: Specify model over skills, using image classifiers for representing pre- and post-conditions
- Step 3: Run symbolic planner with the repertoire of visuomotor skills learned in Step 1

Learning a Repertoire of Visuomotor Skills with EMBR

- Learns a low-dimensional latent space with a VAE and a latent dynamics model
- Obtains rewards by learning image classifiers
- Learns Q-functions for model-based control

Key Takeaways
1. Q-function is important for low-level skill performance
2. Robustness of replanning w/ model is critical for long-horizon performance

Future Work
1. Reduce the amount of human supervision
2. Expand task scope and handle partial observation
3. Generalization to new environment setup