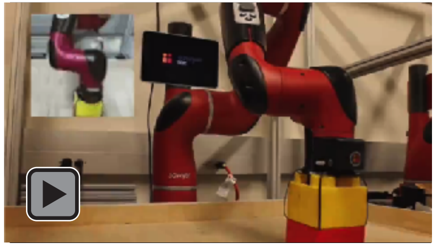
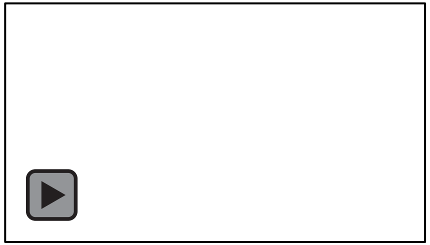
# SOLAR: Deep Structured Representations for Model-Based Reinforcement Learning

Marvin Zhang*, Sharad Vikram*, Laura Smith, Pieter Abbeel, Matthew J Johnson, Sergey Levine

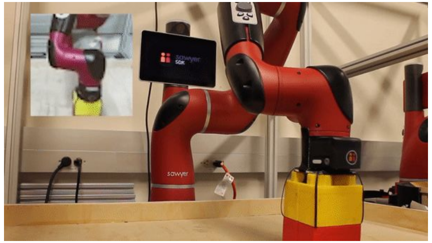
UC Berkeley, UC San Diego, Google

# Efficient reinforcement learning from images

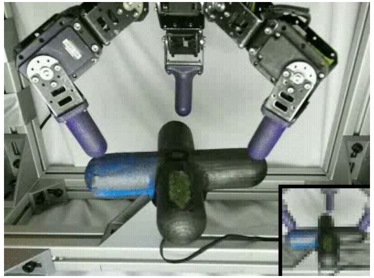




# Efficient reinforcement learning from images

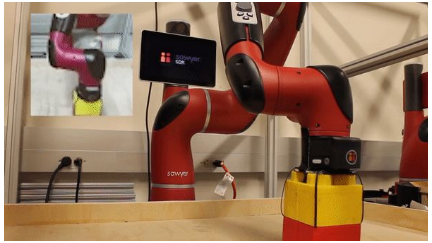


Model-free RL: 4 hours for image-based robotic task, 2 hours for block stacking from *states*



https://sites.google.com/view/sac-and-applications

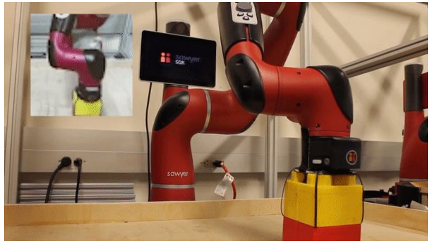
# Efficient reinforcement learning from images



Model-free RL: 20 hours for image-based robotic task, 2 hours for block stacking from *states*

Model-based RL from images: relies on *accurate forward prediction*, which is difficult

# Efficient reinforcement learning from images

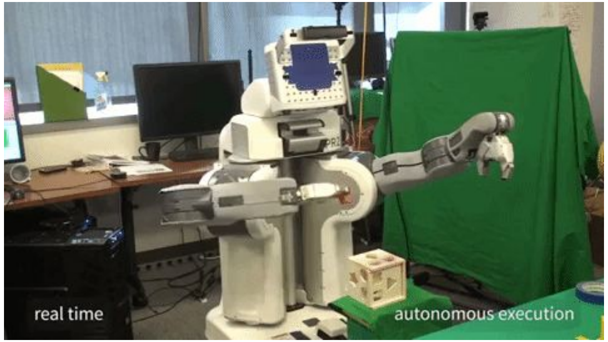


Model-free RL: 20 hours for image-based robotic task, 2 hours for block stacking from *states*

Model-based RL from images: relies on *accurate forward prediction*, which is difficult

Key idea: *structured representation learning* to enable accurate modeling with simple models; model-based method that does use forward prediction

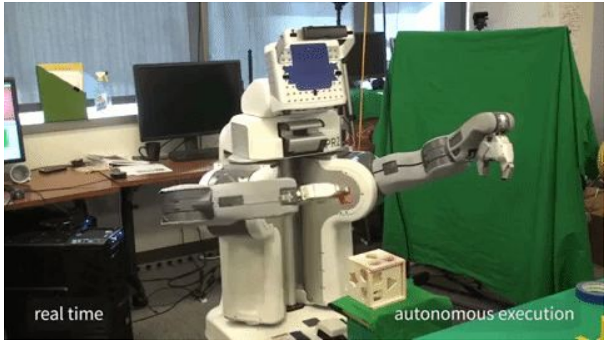
# Preliminary: LQR-FLM





Levine and Abbeel, “Learning Neural Network Policies with Guided Policy Search under Unknown Dynamics”. NIPS 2014.
Levine*, Finn*, Darrell, and Abbeel, “End-to-End Training of Deep Visuomotor Policies”. JMLR 2016.
Chebotar*, Hausman*, Zhang*, Sukhatme, Schaal, and Levine, “Combining Model-Based and Model-Free Updates for Trajectory-Centric Reinforcement Learning”. ICML 2017.

# Preliminary: LQR-FLM

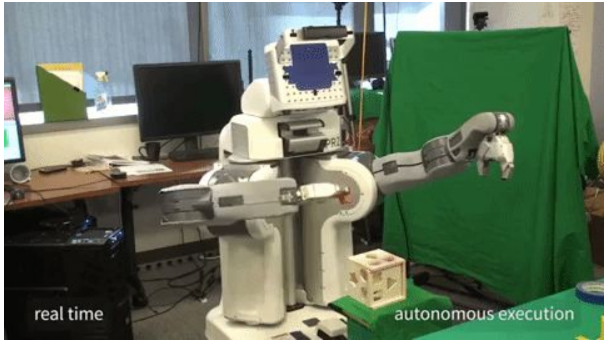




LQR-FLM fits *local models* for policy improvement, not forward prediction

Levine and Abbeel, “Learning Neural Network Policies with Guided Policy Search under Unknown Dynamics”. NIPS 2014.
Levine*, Finn*, Darrell, and Abbeel, “End-to-End Training of Deep Visuomotor Policies”. JMLR 2016.
Chebotar*, Hausman*, Zhang*, Sukhatme, Schaal, and Levine, “Combining Model-Based and Model-Free Updates for Trajectory-Centric Reinforcement Learning”. ICML 2017.

# Preliminary: LQR-FLM





LQR-FLM fits *local models* for policy improvement, not forward prediction

LQR-FLM has worked on complex robotic systems *from states*

Levine and Abbeel, "Learning Neural Network Policies with Guided Policy Search under Unknown Dynamics". NIPS 2014.
Levine*, Finn*, Darrell, and Abbeel, "End-to-End Training of Deep Visuomotor Policies". JMLR 2016.
Chebotar*, Hausman*, Zhang*, Sukhatme, Schaal, and Levine, "Combining Model-Based and Model-Free Updates for Trajectory-Centric Reinforcement Learning". ICML 2017.

# Preliminary: LQR-FLM

LQR-FLM fits linear dynamics and quadratic cost models for policy improvement:

$$\hat{p}(\mathbf{s}_{t+1} \mid \mathbf{s}_t, \mathbf{a}_t) = \mathcal{N}\left(\mathbf{F}\begin{bmatrix}\mathbf{s}_t \\ \mathbf{a}_t\end{bmatrix}, \Sigma\right)$$

$$\hat{C}(\mathbf{s}_t, \mathbf{a}_t) = \mathbf{s}_t^\top \mathbf{R}_t \mathbf{s}_t + \mathbf{a}_t^\top \mathbf{U}_t \mathbf{a}_t$$

# Preliminary: LQR-FLM

LQR-FLM fits linear dynamics and quadratic cost models for policy improvement:

$$\hat{p}(\mathbf{s}_{t+1} \mid \mathbf{s}_t, \mathbf{a}_t) = \mathcal{N}\left(\mathbf{F}\begin{bmatrix}\mathbf{s}_t \\ \mathbf{a}_t\end{bmatrix}, \Sigma\right)$$

$$\hat{C}(\mathbf{s}_t, \mathbf{a}_t) = \mathbf{s}_t^\top \mathbf{R}_t \mathbf{s}_t + \mathbf{a}_t^\top \mathbf{U}_t \mathbf{a}_t$$

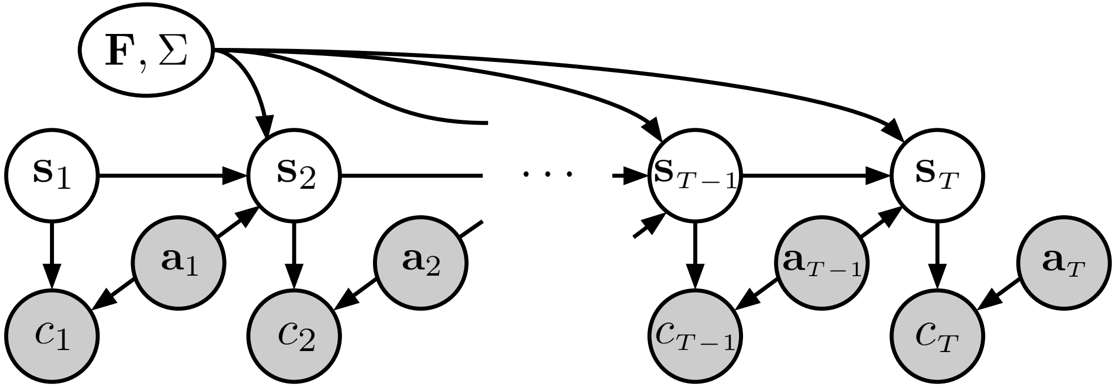
This works well, even for complex systems, if the state is relatively simple, but this doesn't work if the state is complex, such as images

# Our method: SOLAR

In this work, we enable LQR-FLM for images using *structured representation learning*

# Our method: SOLAR

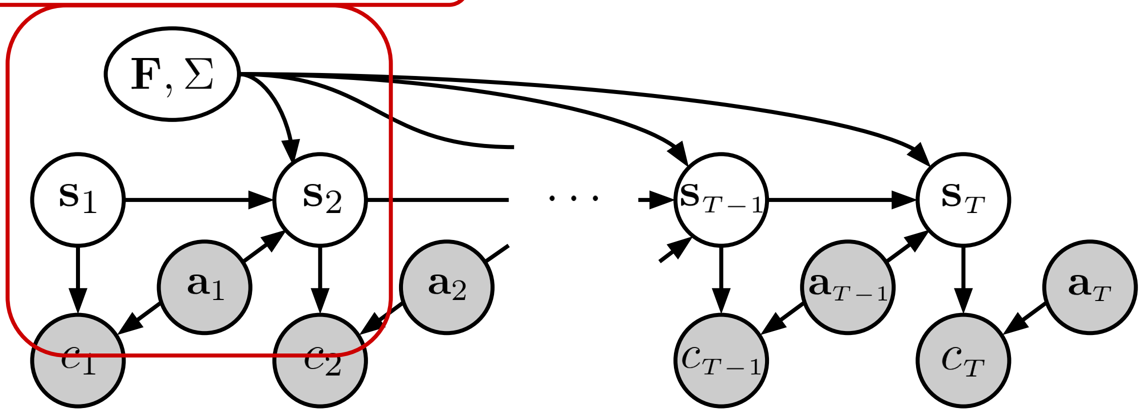
In this work, we enable LQR-FLM for images using *structured representation learning*



# Our method: SOLAR

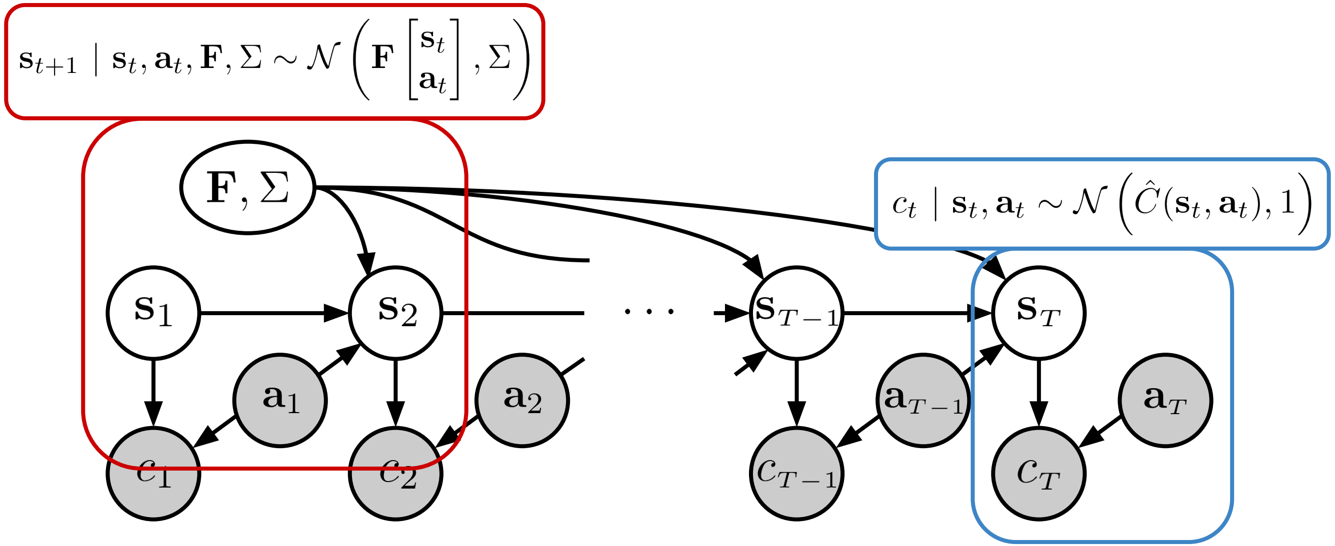
In this work, we enable LQR-FLM for images using *structured representation learning*

$$\mathbf{s}_{t+1} \mid \mathbf{s}_t, \mathbf{a}_t, \mathbf{F}, \Sigma \sim \mathcal{N}\left(\mathbf{F}\begin{bmatrix}\mathbf{s}_t \\ \mathbf{a}_t\end{bmatrix}, \Sigma\right)$$



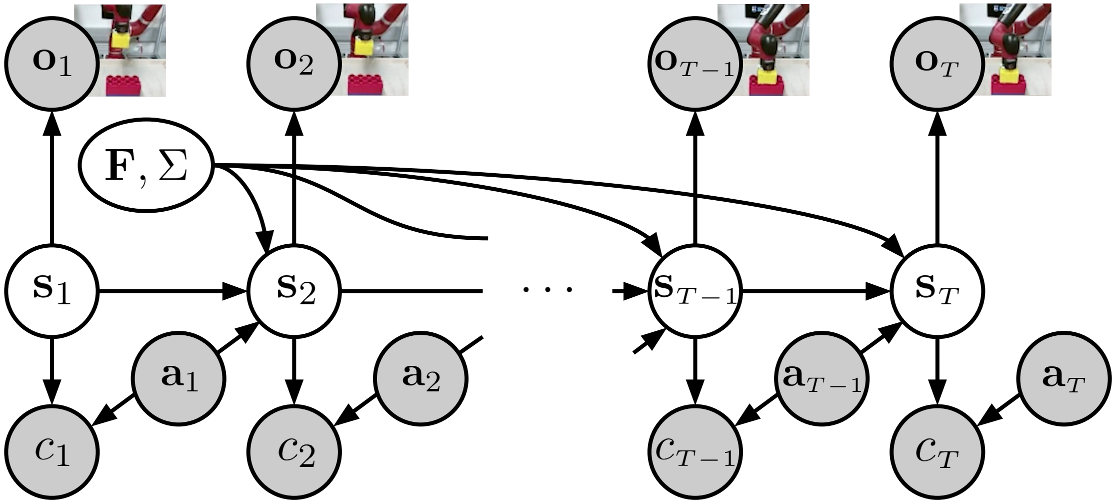
# Our method: SOLAR

In this work, we enable LQR-FLM for images using *structured representation learning*



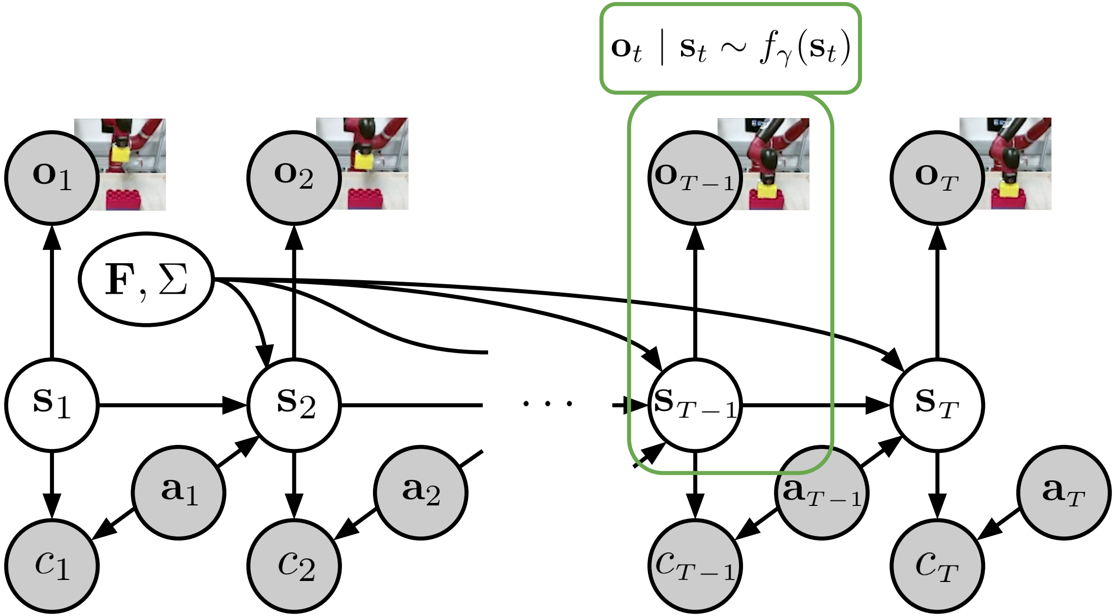
# Our method: SOLAR

In this work, we enable LQR-FLM for images using *structured representation learning*



# Our method: SOLAR

In this work, we enable LQR-FLM for images using *structured representation learning*



# Real robot results

Our method is more efficient than both prior model-free and model-based methods

# Real robot results

Our method is more efficient than both prior model-free and model-based methods

Block stacking: we can *transfer a representation and model* to multiple initial arm positions

# Real robot results

Our method is more efficient than both prior model-free and model-based methods

Mug pushing: We can solve this task from *sparse reward* using human key presses

# Thank you

Poster #34

Paper: https://arxiv.org/abs/1808.09105

Website: https://sites.google.com/view/icml19solar

Blog post: https://bair.berkeley.edu/blog/2019/05/20/solar

Code: https://github.com/sharadmv/parasol