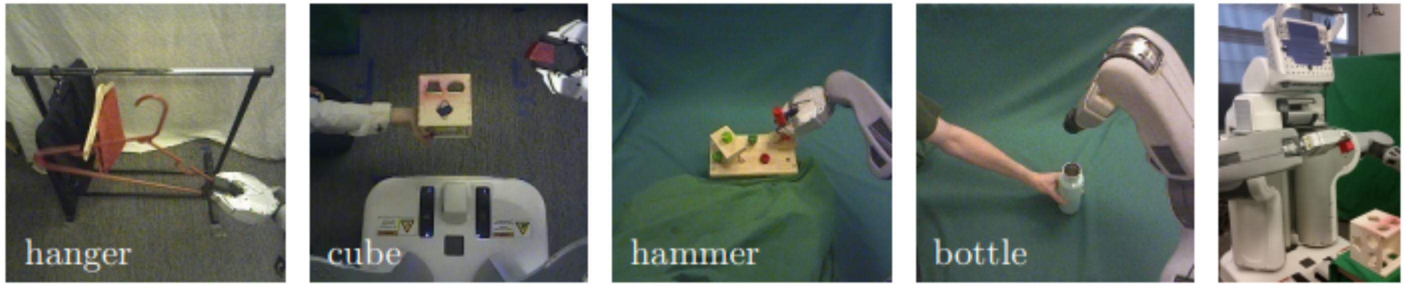
VISUAL FORESIGHT

Model-Based Deep Reinforcement Learning for Vision-Based Robotic Control

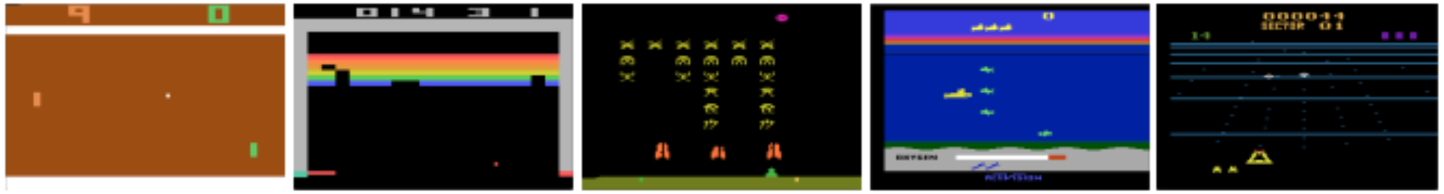
Chenhui Zhao

Introduction

End-to-end Learning of Deep Visuomotor Policies[1]



Playing Atari with Deep Reinforcement Learning[2]



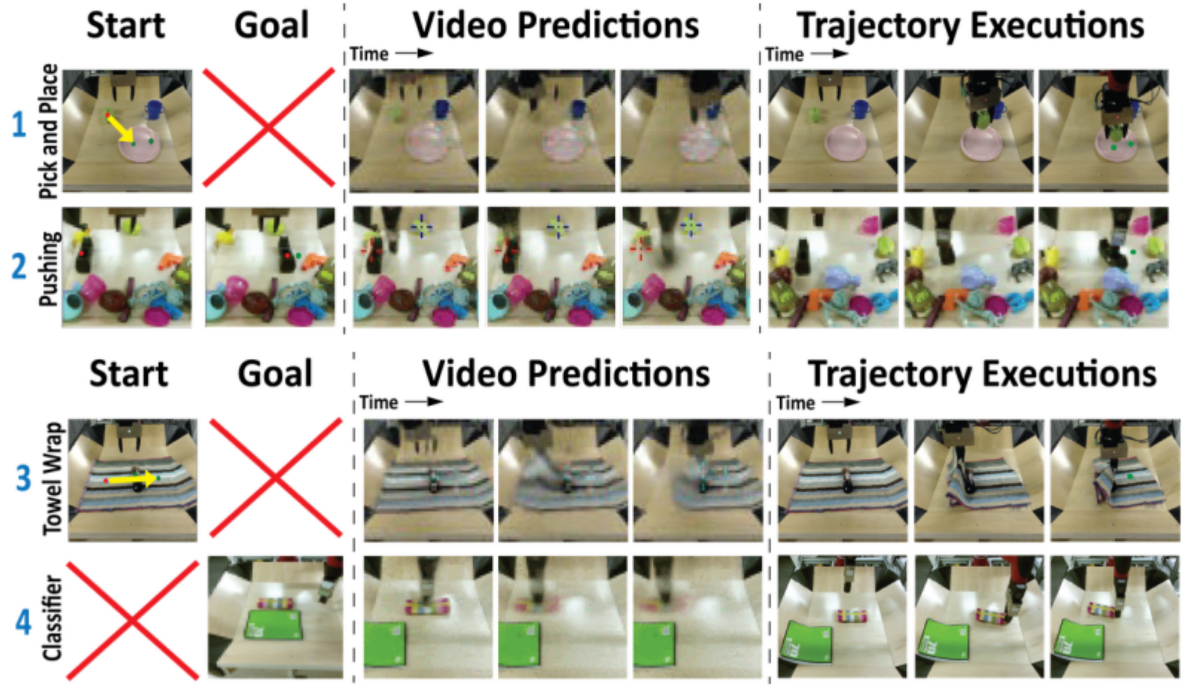
The ability to learn behaviors that generalize to new tasks is still limited. The key to generalization is diversity.

[1] S. Levine, C. Finn, T. Darrell, and P. Abbeel, "End-to-end learning of deep visuomotor policies," Journal of Machine Learning Research (JMLR), 2016.

[2] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," arXiv:1312.5602, 2013.

INTRODUCTION

The key to generalization is diversity and prediction is often considered a fundamental component of intelligence.



INTRODUCTION

Visual Model Predictive Control

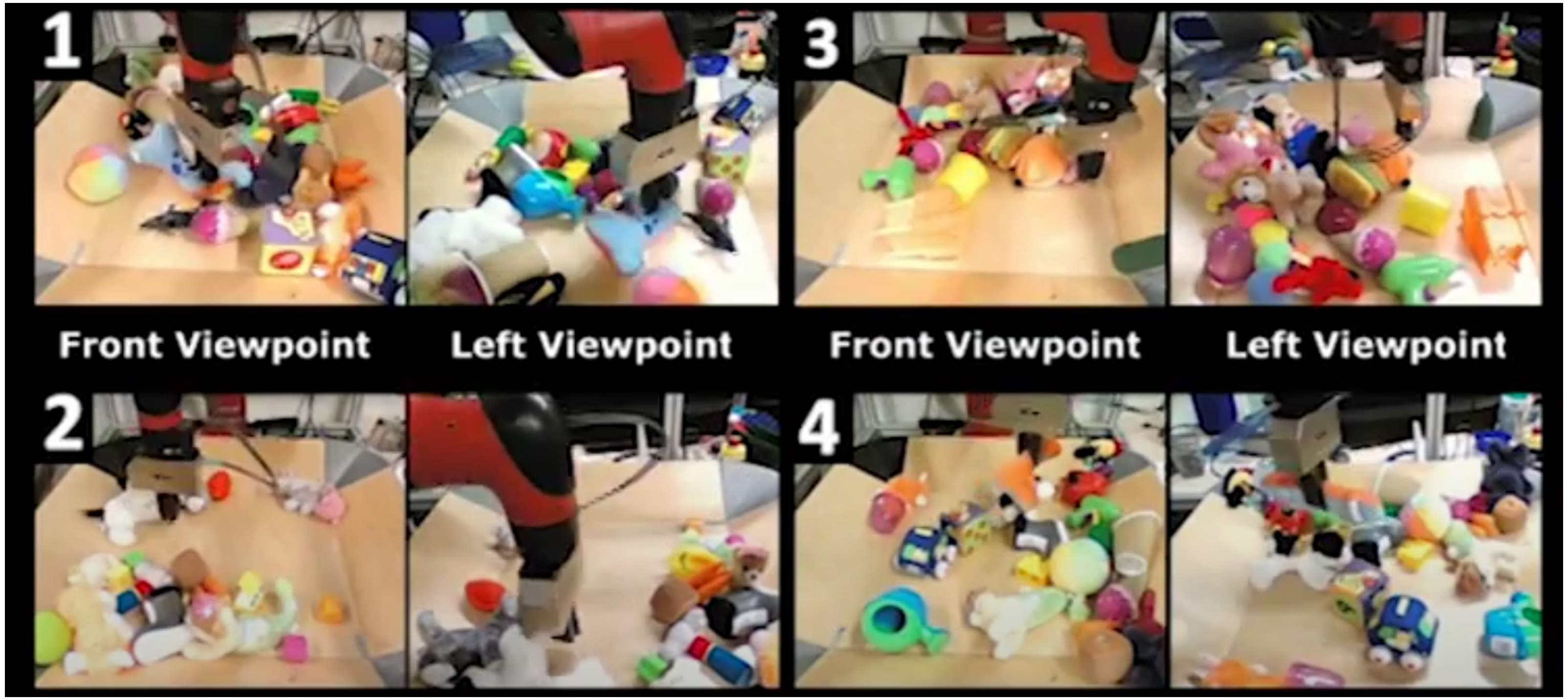
Describe deep neural network architectures that are effective for predicting pixel-level observations amid occlusions and with novel objects.

Present several practical methods for specifying and evaluating progress towards the goal:

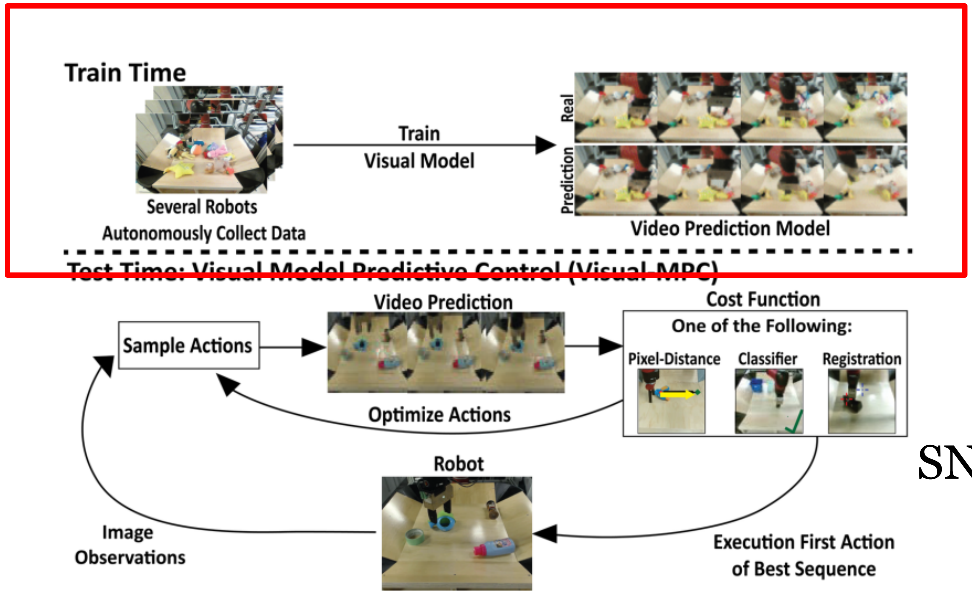
- Distances to goal pixel positions
- Registration to goal images
- Success classifiers

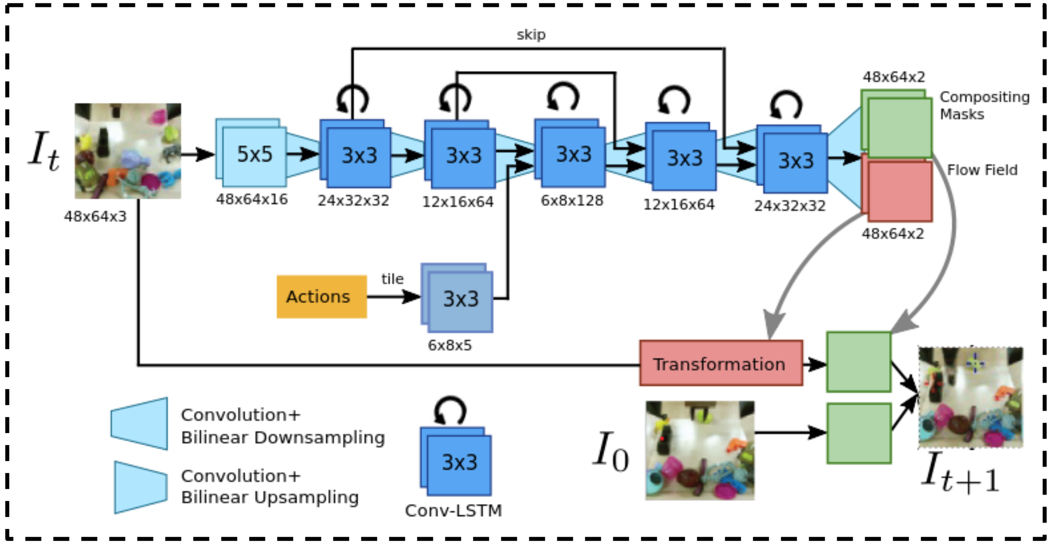
Usupervised Data Collection

Appling random actions sampled from a pre-specified distribution.

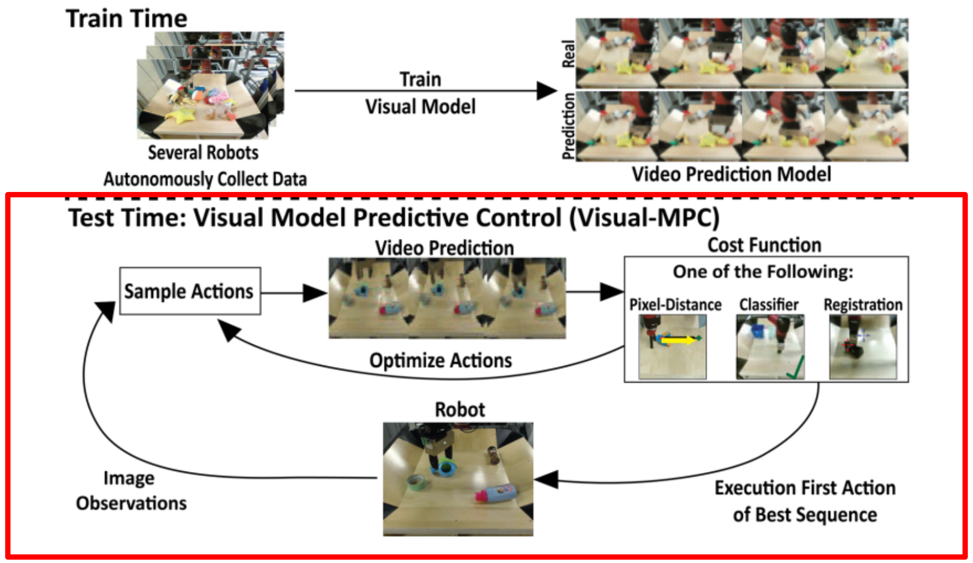


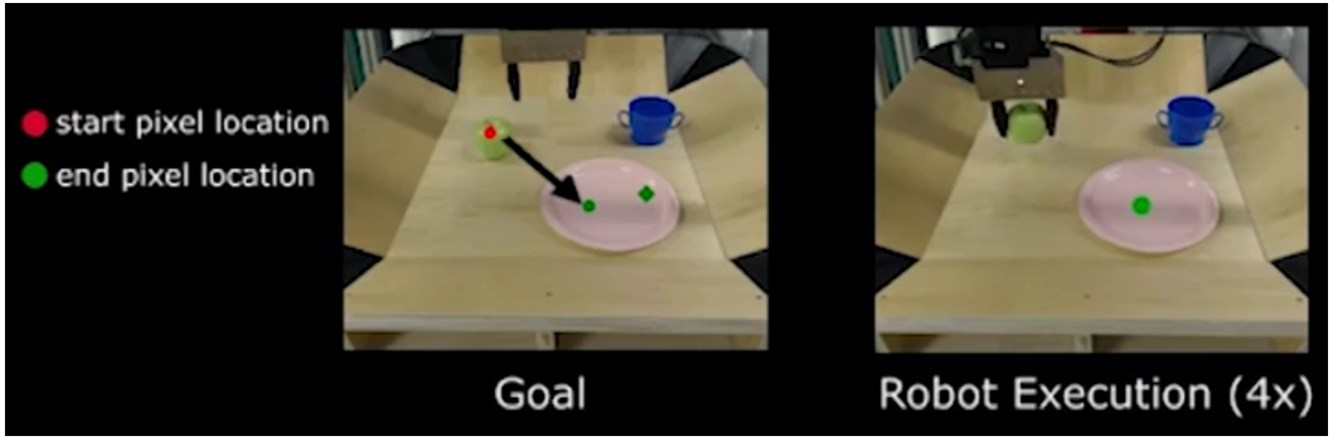
Video Prediction for Control



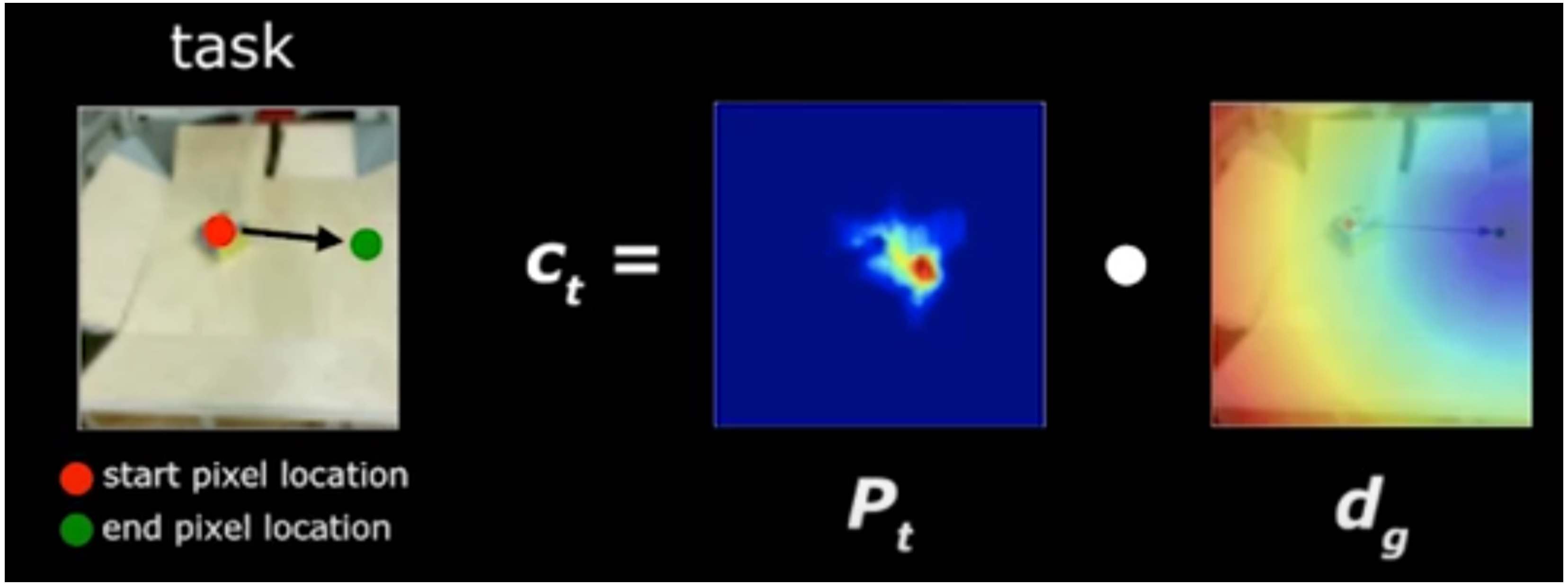


Test Time Control





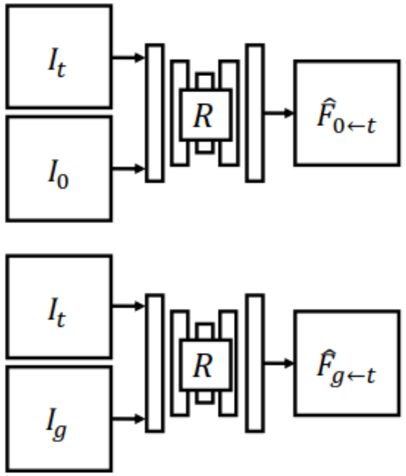
Planning Cost Functions: Pixel Distance Cost



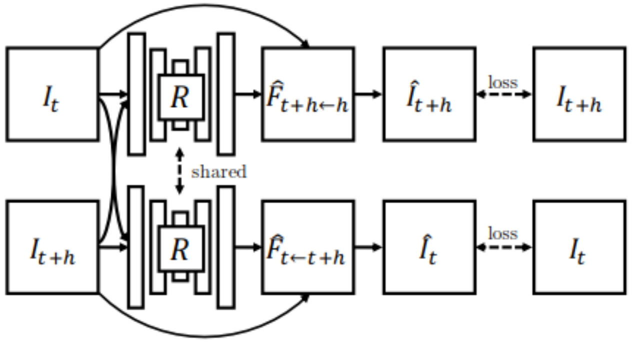
$$c = \sum_{t=1,\ldots,T} c_t = \sum_{t=1,\ldots,T} \mathbb{E}_{\hat{d}_t \sim P_t} \left[\|\hat{d}_t - d_g\|_2 \right]$$

Planning Cost Functions: Registration-Based Cost

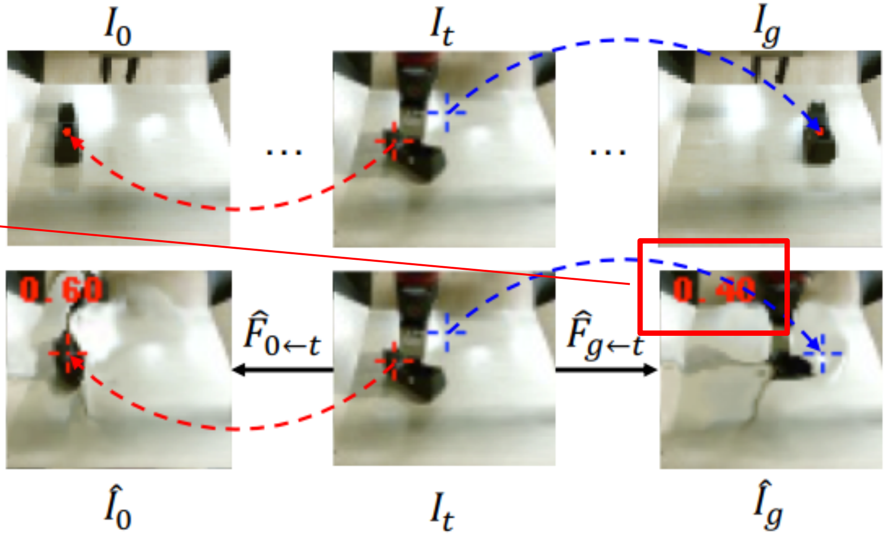
$$\lambda_i = \frac{||I_i(d_i) - \hat{I}_i(d_i)||_2^{-1}}{\sum_j^N ||I_j(d_j) - \hat{I}_j(d_j)||_2^{-1}}. \qquad c = \sum_i \boxed{\lambda_i} c_i$$

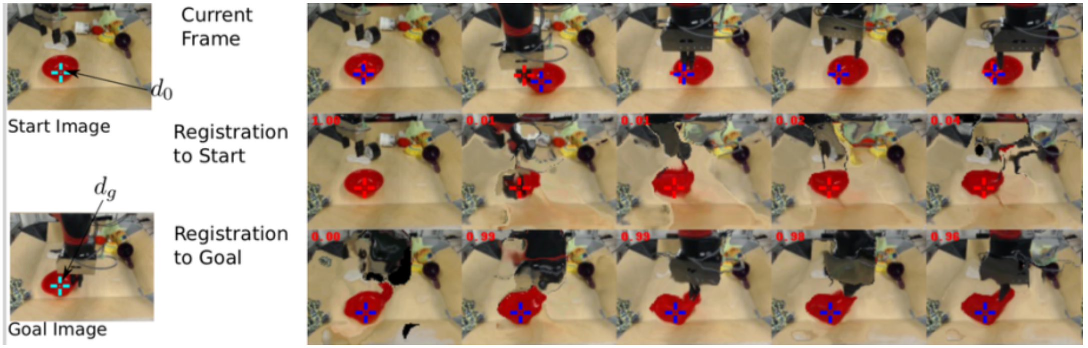


(a) Testing usage.

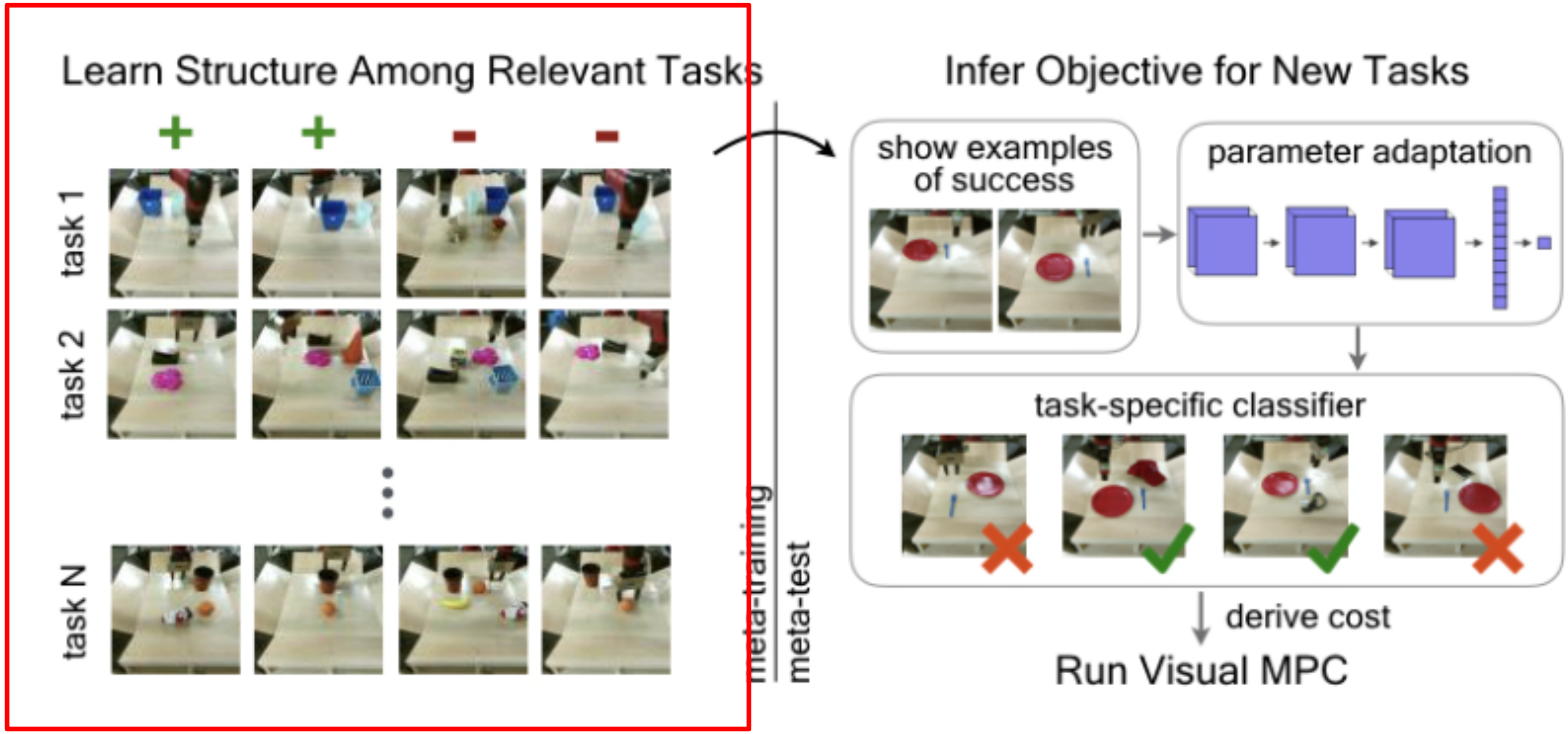


(b) Training usage.

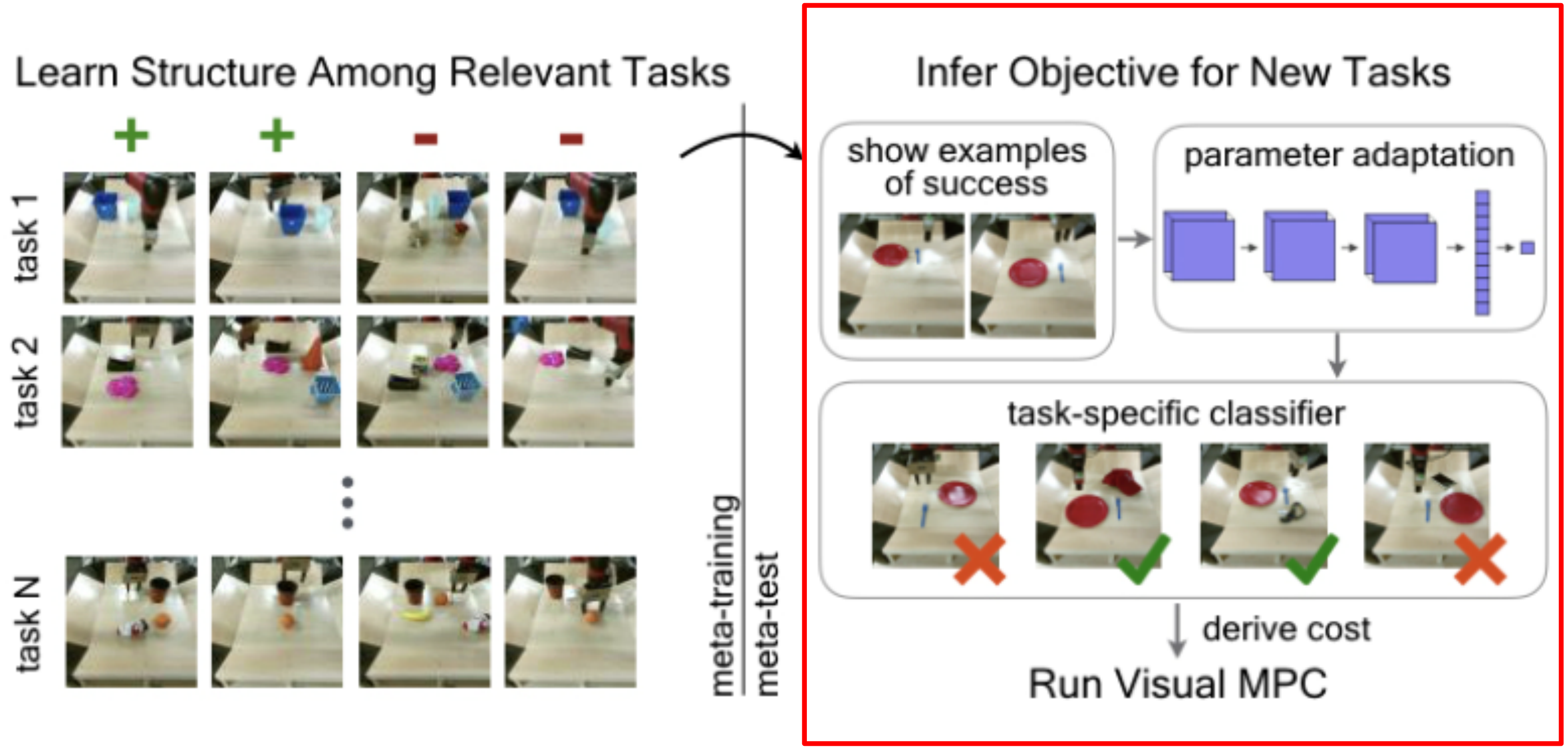




Planning Cost Functions: Classifier-Based Cost



Planning Cost Functions: Classifier-Based Cost

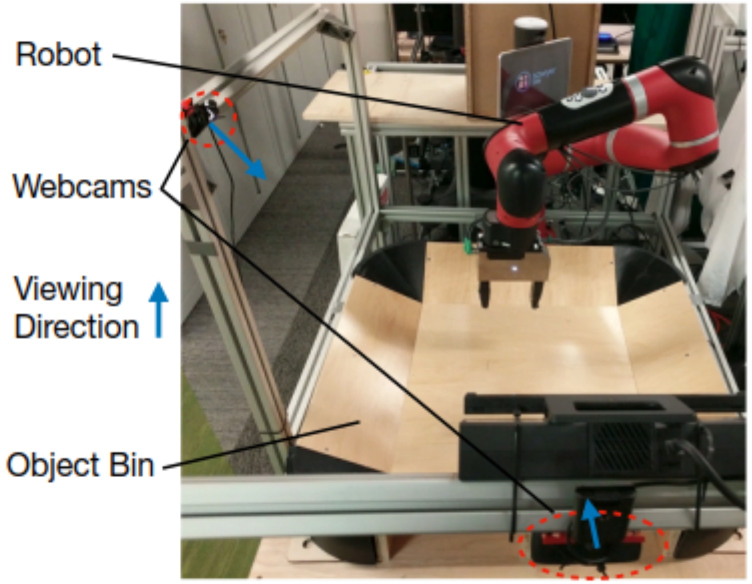


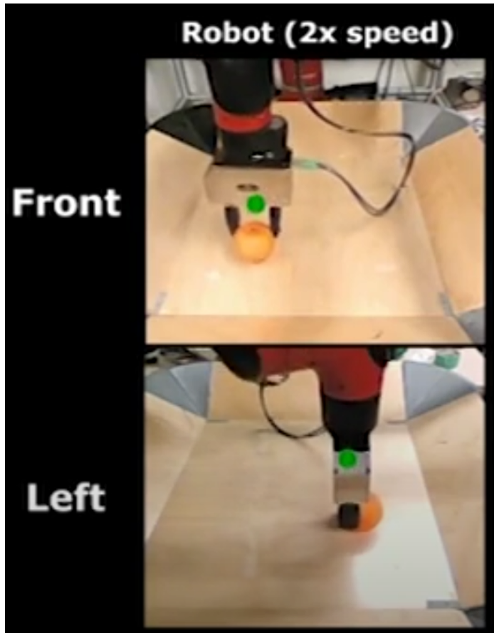
Optimizer & Custom Action Sampling Distributions

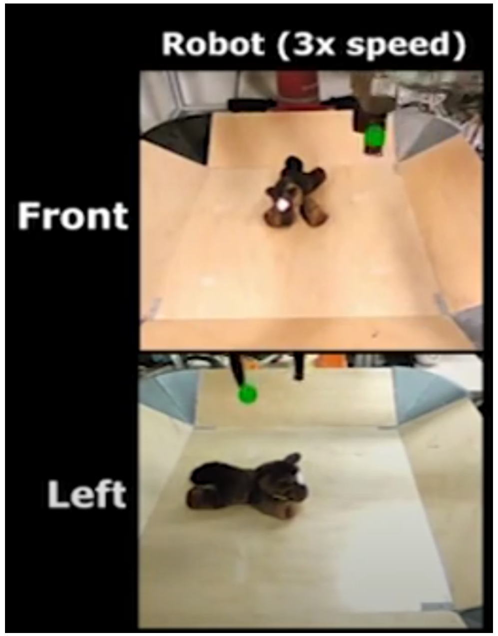
The role of the optimizer is to find actions sequences that minimize the sum of the per time step pixel distance costs. The key is using cross-entropy method to allow them to ensure actions stay within the distribution of actions the model encountered during training.

To allow picking up and placing of objects as well as folding of cloth to occur more frequently, the author incorporate a simple reflex during data collection, which is inspired by the palmar reflex observed in infants. The gripper automatically closes when the height of the wrist above the table is lower than a small threshold.

Multi-View Visual MPC



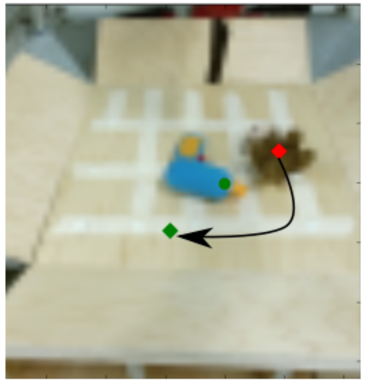


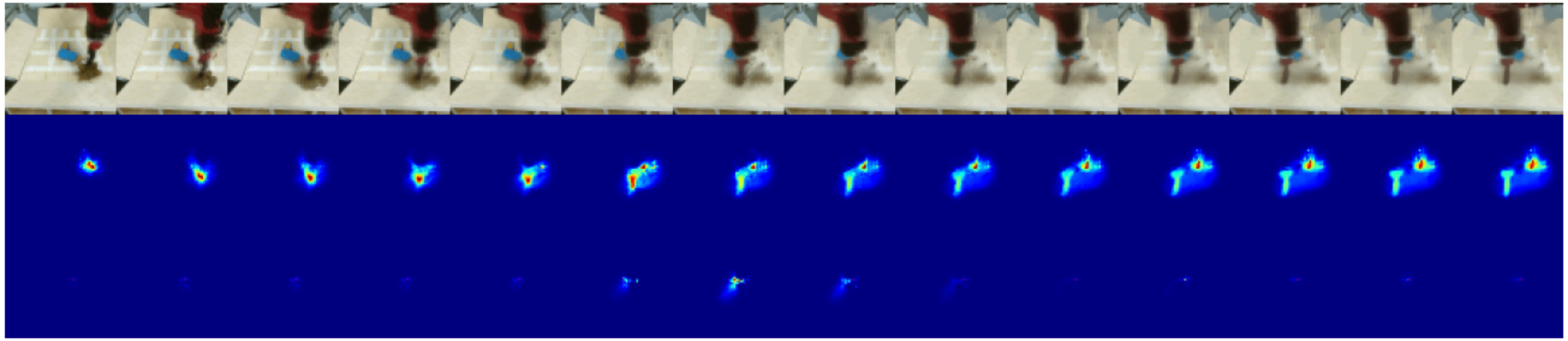


Experimental Evaluation

Video Prediction Architectures

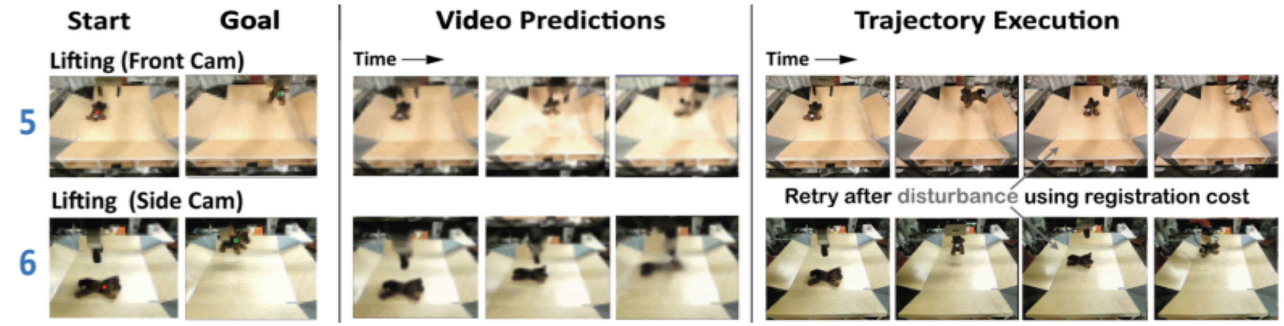
	moved imp. ± std err. of mean	stationary imp. ± std err. of mean
DNA [6]	0.83 ±0.25	-1.1 ± 0.2
SNA	**10.6 ± 0.82**	**-1.5 ± 0.2**





Experimental Evaluation

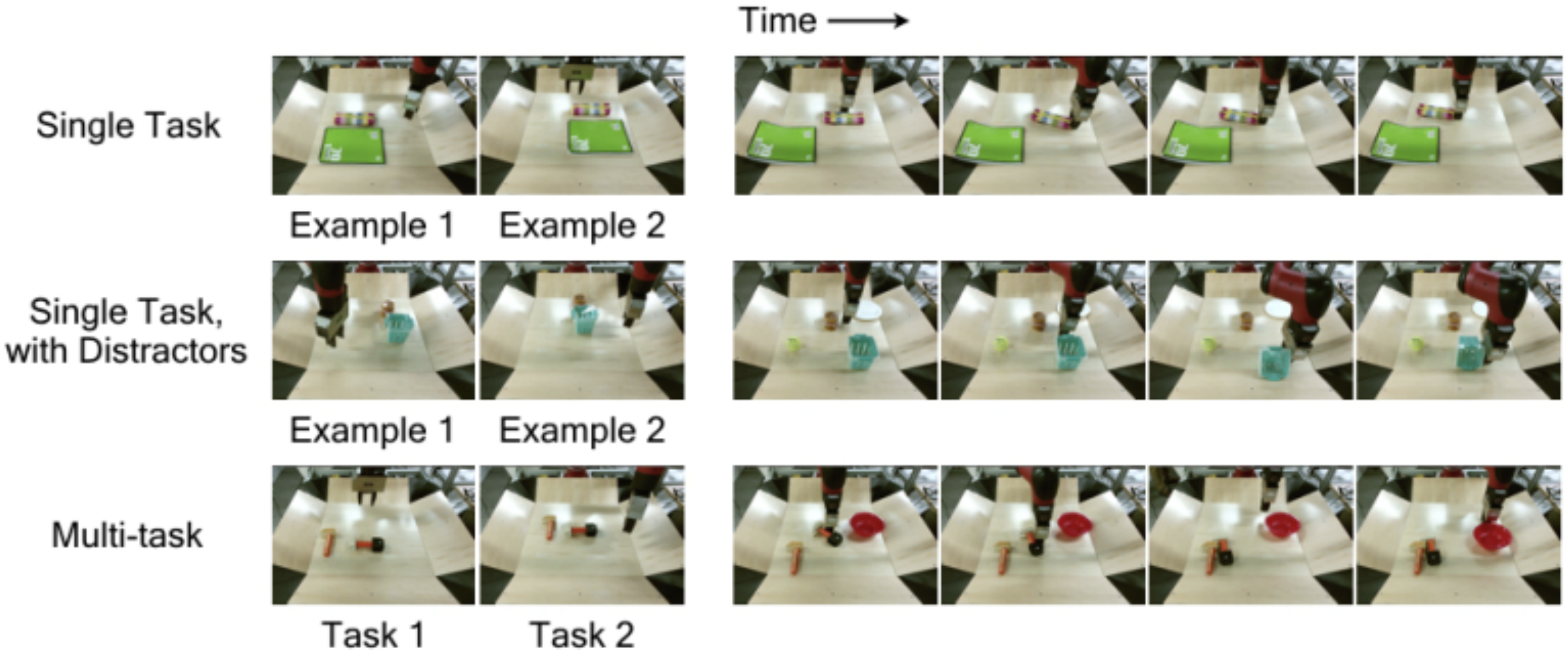
Evaluating Registration-Based Cost Functions



	Short	Long
Visual MPC + predictor propagation	83%	20%
Visual MPC + OpenCV tracking	83%	45%
Visual MPC + registration network	83%	**66%**

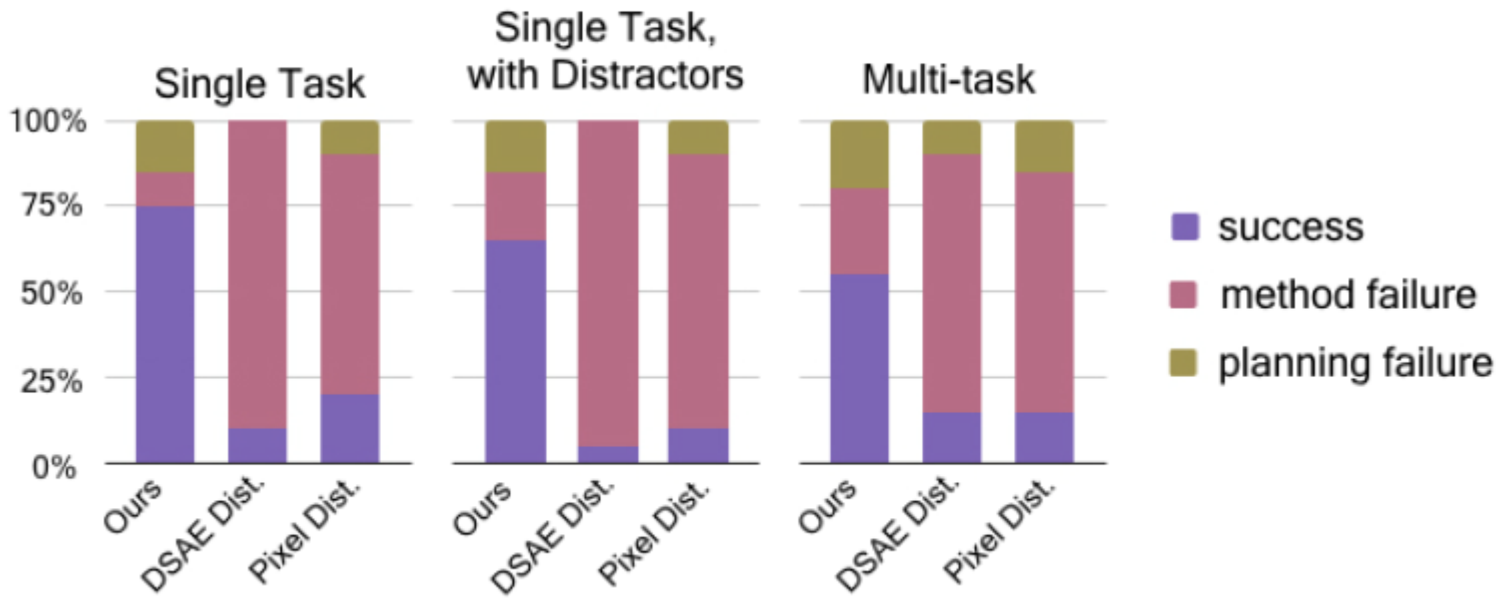
EXPERIMENTAL EVALUATION

Evaluating Classifier-Based Cost Functions



Experimental Evaluation

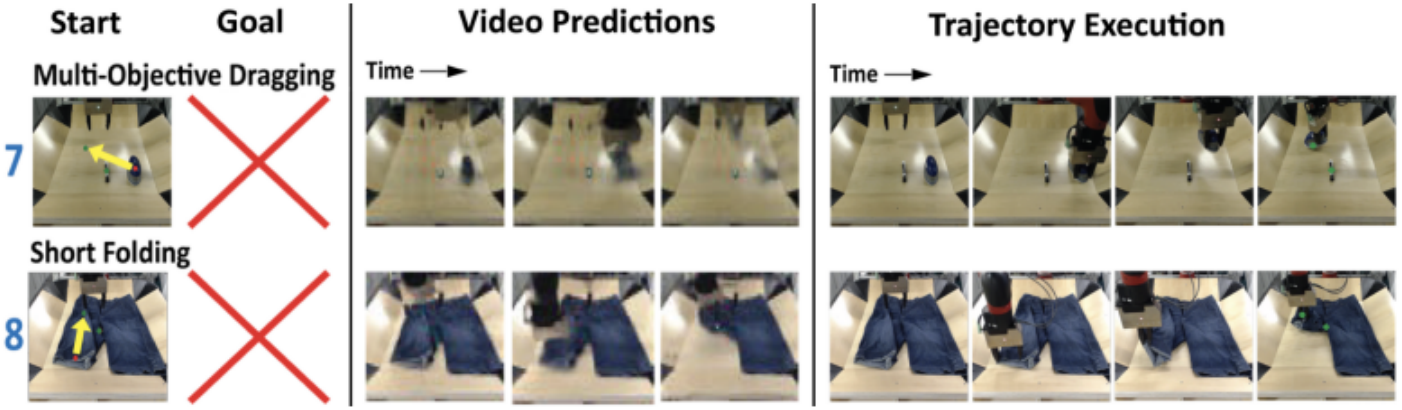
Evaluating Classifier-Based Cost Functions



Experimental Evaluation

Evaluating Multi-Task Performance

	% of Trials with Final Pixel Distance < 15
Visual MPC	**75%**
Calibrated Camera Baseline	18.75 %



Discussion

Most of the generalization performance is likely a result of large-scale self-supervised learning, which allows to acquire a rich, dynamics model of the environment.

The main limitations of the presented framework are that all target objects need to be visible throughout execution, it is currently not possible to handle partially observed domains.

Discussion

@82_f2

The intuitive idea behind their learning setup (predicting future sensor readings from earlier sensor readings) seems to be rooted in a couple of papers we read before about prediction enabling manipulation of objects and internal models that can predict the consequences of motor actions. I think it's very interesting to require a system to learn to do a similar thing in order to complete tasks. It seemed like a step towards robots understanding the effect that their actions will have on the world, and whether that result is desirable, which might be essential for more general intelligence.

Discussion

@82_f3

I think that the Visual-Model Predictive Control was an interesting way of implementing prediction into a reinforcement model. It draws a lot of similarities to how I imagine my mind would work, where internally I am making predictions about what should be happening when I am doing a task, and if something goes differently, my CNS would take that error and factor it into the next decision. The main difference I see is the efficiency at which these predictive errors are optimized. It's common for us as humans to make a wrong decision once, but uncommon for us to continuously make the same errors over and over again. I wonder what the vast difference in corrective efficiency could be related to. Could it be another one of our characteristics such as curiosity that allow us to find more optimal actions? I think this is the right idea, but there is work that could be done regarding the speed of learning.

Discussion

@82_f4

This paper combines a number of different novel contributions that it looks were developed in previous papers together in a unified model. One aspect of this research I found particularly interesting was how they used incorporated skip connections which enabled the model to predict pixel locations even if the targeted object was occluded for a number of subsequent frames. Apparently, without this addition, the model was not able to handle these occlusion events at all.