HIERARCHICAL NEURAL DYNAMIC POLICIES

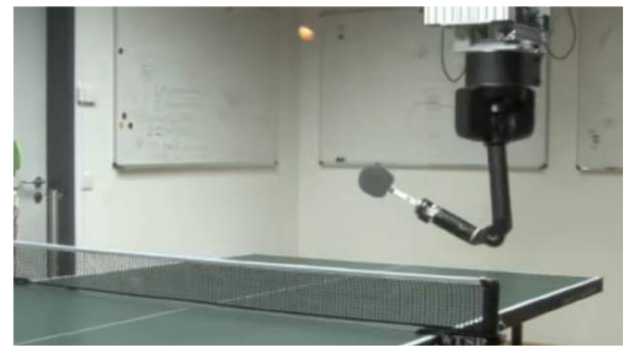
Shikhar Bahl Abhinav Gupta Deepak Pathak

CMU

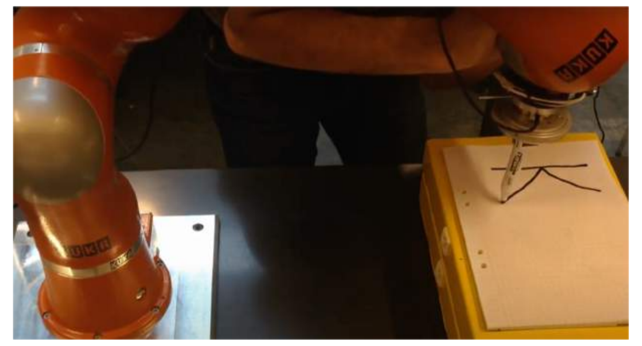
For videos and paper!
https://shikharbahl.github.io/hierarchical-ndps/

How can we build *generalizable* policies *for real world dynamic tasks*?

Dynamical systems in robotics literature have been used to perform dynamic tasks (e.g. **DMPs** [Schaal., 2002])



[Muelling et. al, 2013]



[Steinmetz, 2014]

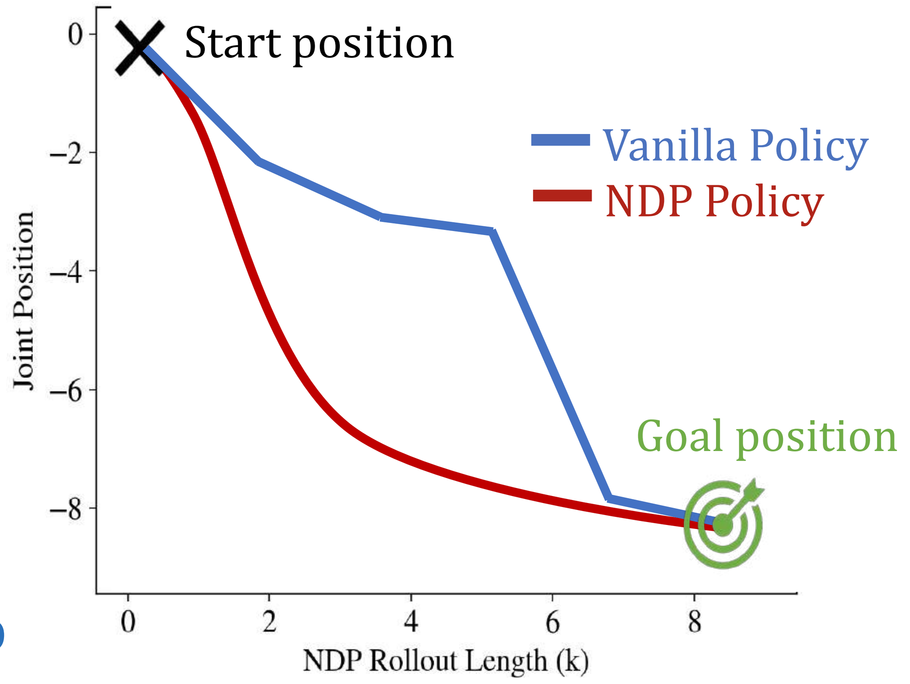
DMP Structure

$$\ddot{y} = \alpha(\beta(g-y) - \dot{y}) + f(x)$$

$$f(x,g) = \frac{\sum \psi_i w_i}{\sum \psi_i} x(g - y_0)$$

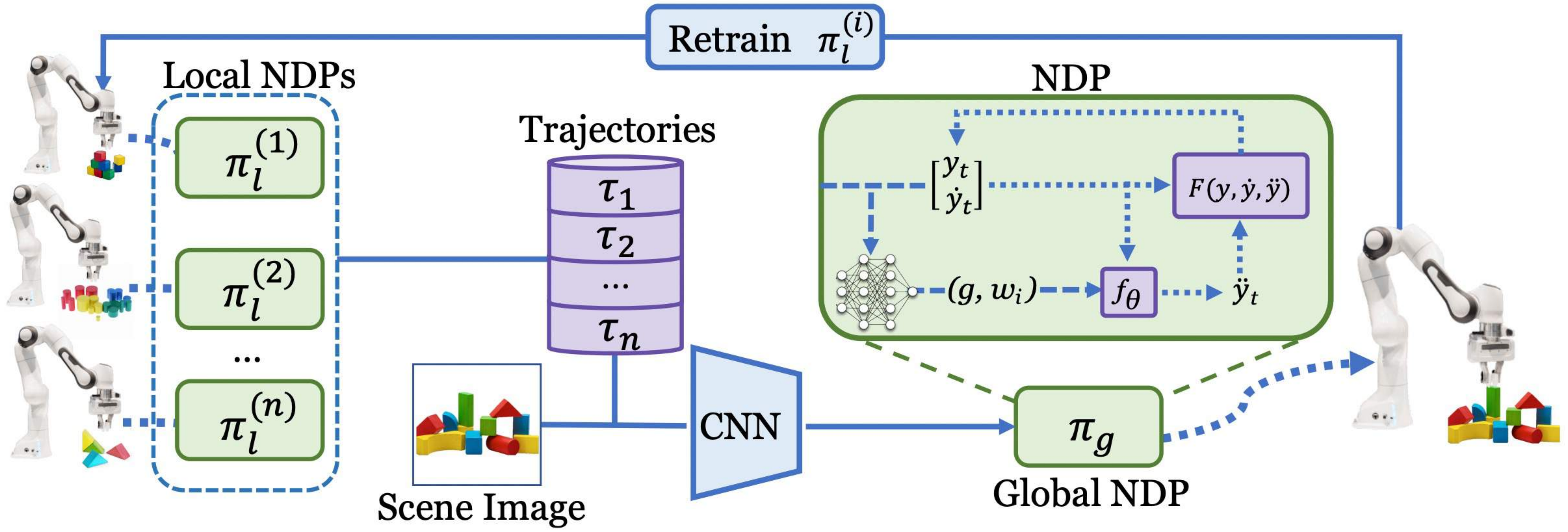
NDPs

- ✔ NDPs [Bahl et al., 2020] embed DMPs in **policy networks**
- ✔ NDPs reason the space of **trajectory shapes and goals**
- ✔ NDPs operate in a space of **smooth & plausible** trajectories
- ✘ Single dynamical systems tend to **overfit** to single trajectories

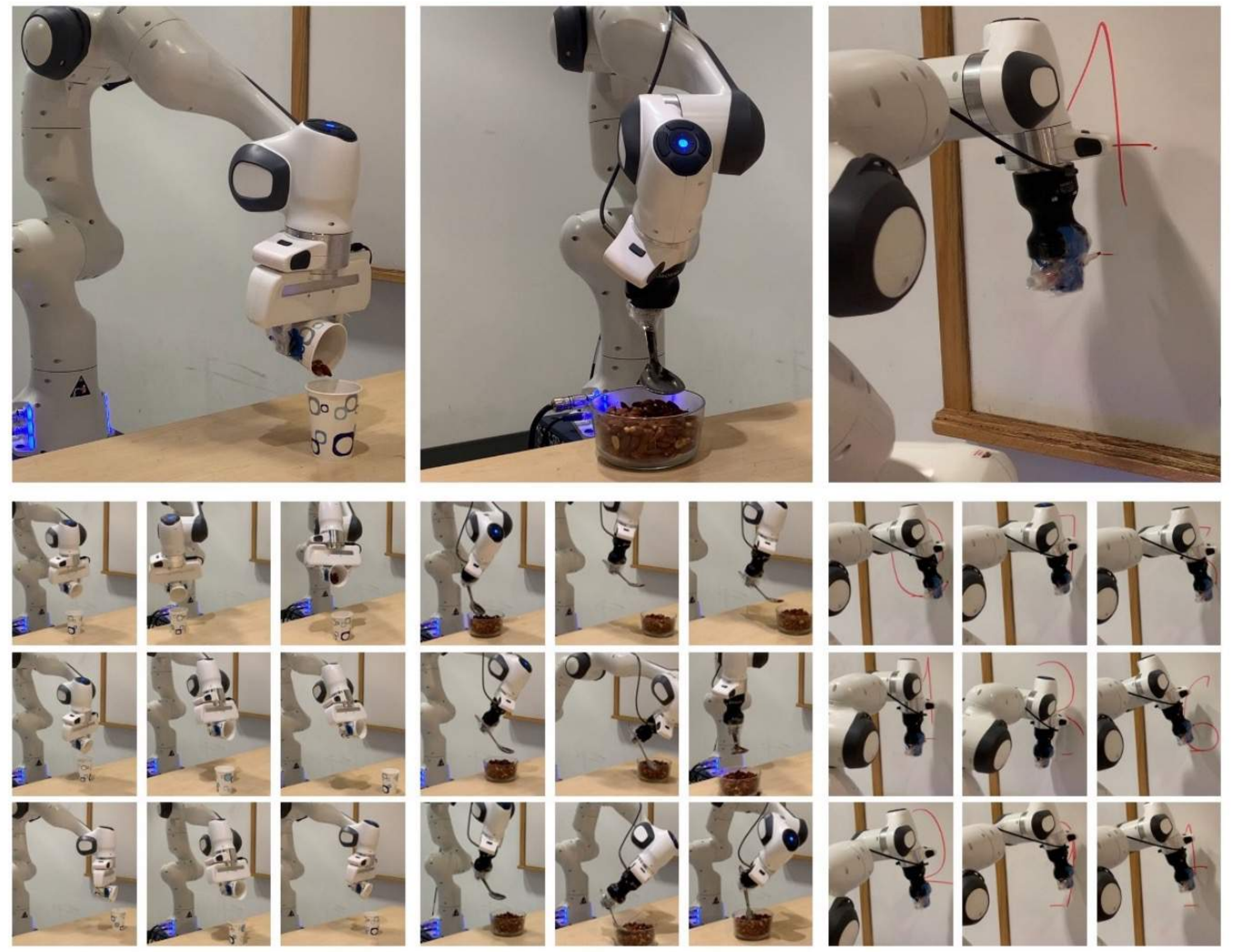


How can we leverage dynamical systems to handle diversity in the task and handle unstructured data?

Local-to-Global Structure: Hierarchical Neural Dynamical Policies (H-NDPs)



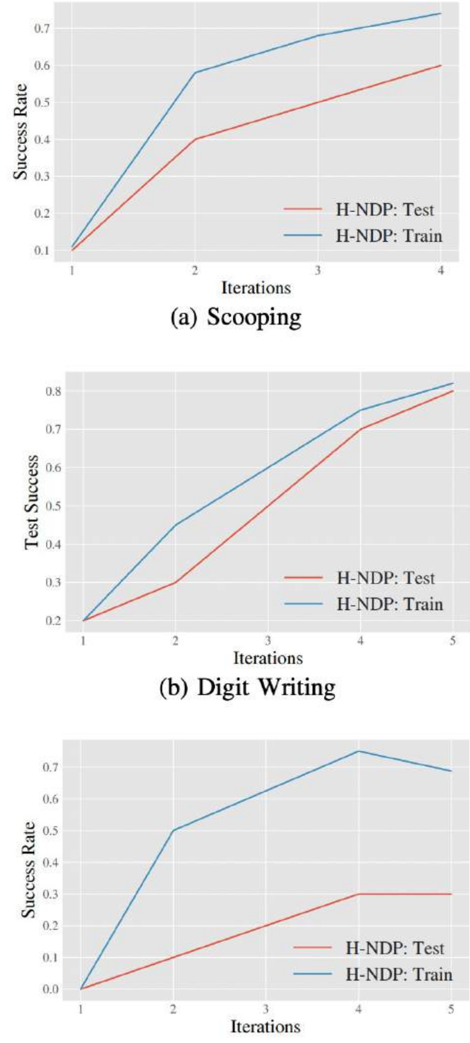
- Train **local NDPs** on individual task regions from state space
- Distill into a **global NDP** that operates from raw images only



H-NDPs can perform ***real-world*** dynamic tasks from ***raw images*** only and ***generalize*** to novel settings.

Learning from Demonstrations

We perform a **large scale, systematic** evaluation in the **real world**



(a) Scooping

(b) Digit Writing

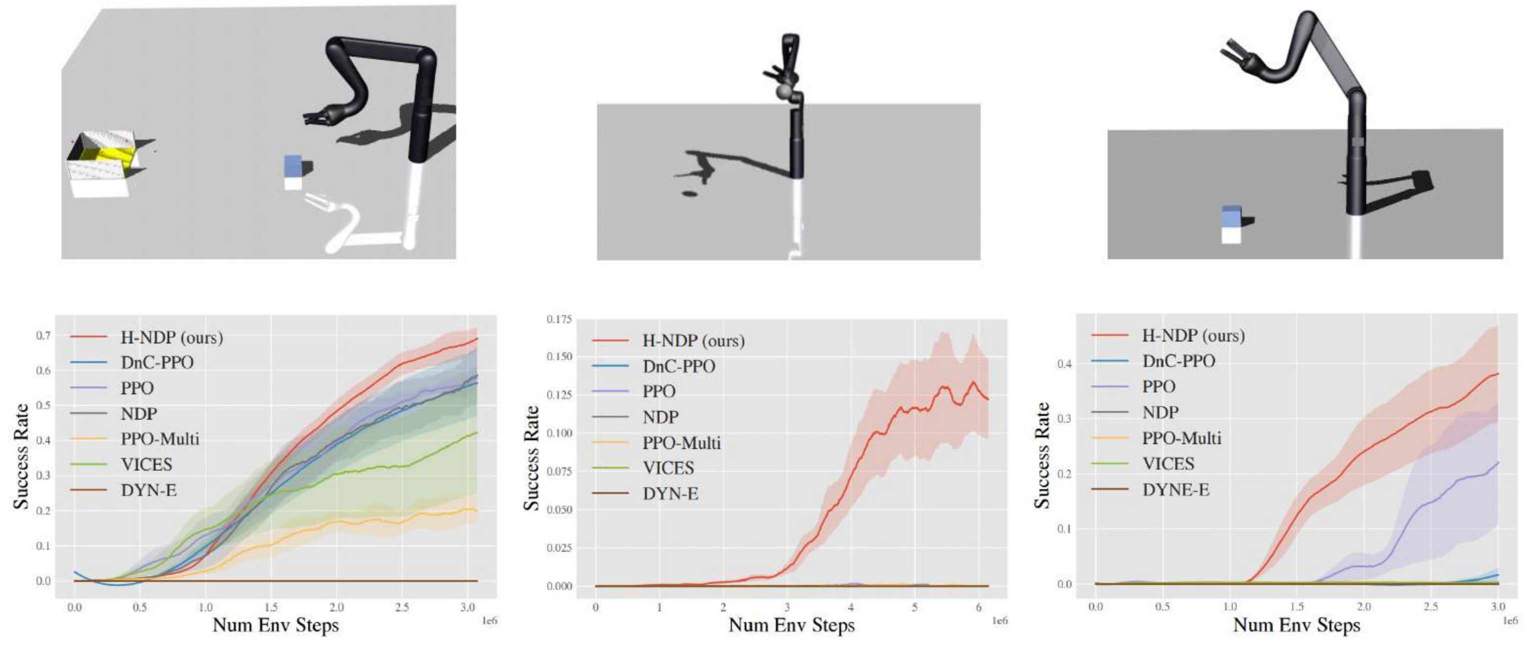
(c) Pouring

	#Demos	#Iter	Writing	Scooping	Pouring
No local-to-global structure:					
NDP	1x	1	0.2	0.2	0.0
Vanilla NN	1x	1	0.1	0.0	0.0
No local-to-global structure with 5x Demos:					
NDP	5x	1	0.5	0.3	0.0
Vanilla NN	5x	1	0.1	0.0	0.0
Local-to-global but no iterative refinement:					
GPS	5x	1	0.1	0.0	0.0
H-NDPs (ours)	5x	1	0.4	0.3	0.0
Both local-to-global and iterative refinement:					
GPS	1x	5	0.3	0.0	0.2
H-NDPs (ours)	1x	5	**0.8**	**0.6**	**0.3**

H-NDPs show **strong performance** against **state-of-the-art baselines**

(b) Input (c) Ours (d) Local NDP (e) GPS

Reinforcement Learning



(a) Throwing (b) Catching (c) Picking

H-NDPs outperform baselines for dynamic tasks with high diversity