

Single-Shot Learning of Stable Dynamical Systems for Long-Horizon Manipulation Tasks

Alexandre St-Aubin¹, Amin Abyaneh², and Hsiu-Chin Lin^{1,2}

Abstract—

Mastering complex sequential tasks continues to pose a significant challenge in robotics. While there has been progress in learning long-horizon manipulation tasks, most existing approaches lack rigorous mathematical guarantees for ensuring reliable and successful execution. In this paper, we extend previous work on learning long-horizon tasks and stable policies, focusing on improving task success rates while reducing the amount of training data needed. Our approach introduces a novel method that (1) segments long-horizon demonstrations into discrete steps defined by waypoints and subgoals, and (2) learns globally stable dynamical system policies to guide the robot to each subgoal, even in the face of sensory noise and random disturbances. We validate our approach through both simulation and real-world experiments, demonstrating effective transfer from simulation to physical robotic platforms.

I. INTRODUCTION

Learning to perform complex sequential tasks continues to be a fundamental challenge in robotics [1]–[4]. Many routine tasks, such as assembly and decluttering, require sequential decision-making and coordinated interaction between the robot and objects to perform a series of motion primitives.

Imitation learning works through these sequential tasks by learning to emulate a set of expert demonstrations [5]–[7]. However, most imitation learning methods are designed to learn a single task and tend to be unreliable when learning from complex, long-horizon expert demonstrations [1], [3], [8]. Notably, safety and stability guarantees are frequently disregarded in favor of achieving higher performance in stochastic environments [9].

Recent efforts to tackle long-horizon tasks focus on goal-conditioned imitation learning [10], [11]. Since long-horizon robotics tasks often involve multiple implicit subtasks or skills, a more viable direction refines the learning process by decomposing long-horizon and sequential tasks into a series of *subtasks* capable of reconstructing the original expert trajectories [4], [10], [12], [13].

Despite this progress, most of these methods cannot provide rigorous mathematical guarantees to generate reliable and successful outcomes. Therefore, deploying policies learned from long-horizon manipulation tasks in simulation to real robot systems raises a lot of safety concerns. Additionally, most methods require an impractically large number

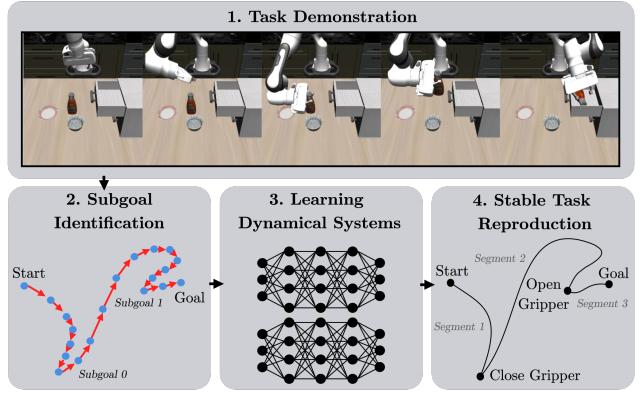


Fig. 1: Overview of our approach: Long-horizon demonstrations (1) are first segmented into subgoals (2). Low-level stable dynamical policies are then learned to robustly reach each subgoal, even in the presence of perturbations (3). Finally, a high-level policy orchestrates a cascade of these stable policies for each segment, replicating the long-horizon expert demonstrations (4).

of demonstrations, even for simple manipulation tasks such as pick-and-place.

The core challenge resides in the fact that long-horizon tasks require accomplishing a cascade of multiple *subgoals* before achieving the main objective, without the explicit knowledge about the subgoals. This brings about three significant challenges. First, as the horizon lengthens, the robot must handle increasing levels of uncertainty, an underlying cause of *compounding error*. Secondly, the robot must infer the subgoals from the demonstrations, introducing an additional layer of uncertainty to the problem. Third, if the robot fails to reach one subgoal, the whole task is jeopardized.

When learning from long-horizon demonstrations, it is crucial to ensure that each intermediary step is executed safely. Prior work on *stable* dynamical policies has been broadly utilized to imitate expert behavior while remaining resilient to perturbations [14]–[17]. The stability property ensures that all trajectories induced by dynamical policies converge to an equilibrium state. However, these methods are designed to learn from a single primitive. Even the most expressive dynamical policies struggle to learn long-horizon tasks [15], [17], as ensuring global stability becomes increasingly difficult over extended time horizons.

In this paper, we build upon prior work on learning long-horizon manipulation tasks and learning stable policies,

¹School of Computer Science, McGill University, Montreal, Canada

²Department of Electrical and Computer Engineering, McGill University, Montreal, Canada

alexandre.st-aubin2@mail.mcgill.ca

This work is sponsored by NSERC Discovery Grant, FRQNT Research Support for New Academics, FRQNT Doctoral Training Scholarships, and McGill Science Undergraduate Research Awards.

aiming to enhance task success rates while minimizing the required training data. We focus on the problem of the movement planning of the robot, without considering visual feedback and/or understanding human intention.

We propose a novel approach that (1) splits a long-horizon demonstration into a set of segments, characterized by waypoints and a subgoal and (2) learns a set of globally stable dynamical system policies that guide the robot to each subgoal despite sensory noise and stochastic perturbations (see Fig. 1). The proposed method is validated both in simulation and on robotic hardware, with a direct transfer from simulation to real-world implementation. The main contributions include:

- We extend the previous work on learning dynamical systems [17] to problems of long-horizon tasks where the robot reaches each subgoal with rigorous theoretical guarantees.
- We show that our learned policies can be deployed directly onto a real-world system with a *single demonstration*.

Our code is available at github.com/Alestaubin/stable-imitation-policy-with-waypoints.

II. BACKGROUND

In this section, we provide a formal problem statement, and a brief background on learning stable dynamical policies and waypoint extraction methods later used in this paper.

A. Problem Definition

Given a demonstration of a long-horizon manipulation task $\mathcal{D} = \{(\mathbf{x}_1, \dot{\mathbf{x}}_1), (\mathbf{x}_2, \dot{\mathbf{x}}_2), \dots, (\mathbf{x}_N, \dot{\mathbf{x}}_N)\}$, where $(\mathbf{x}_n, \dot{\mathbf{x}}_n)$ denote the position and the velocity at the n^{th} step in the demonstration, and N is the task's horizon length. The variables $\mathbf{x} \in \mathbb{R}^d$, $\dot{\mathbf{x}} = \frac{d\mathbf{x}}{dt} \in \mathbb{R}^d$ unambiguously define the d -dimensional position and velocity of a robotic system. For example, \mathbf{x} could be the joint angles or end-effector pose.

We assume that the demonstration \mathcal{D} can be decomposed into a set of K sub-demos, $\{\mathcal{D}^1, \mathcal{D}^2, \dots, \mathcal{D}^K\}$, such that \mathcal{D}^k represents the k^{th} subtask with its corresponding subgoal \mathbf{g}^k . Our goal is to provide the robot with the velocity $\dot{\mathbf{x}}$ given the current position \mathbf{x} at each time step, to reproduce the expert demonstration *without* prior knowledge about the dimensionality of the task K and the subgoal \mathbf{g}^k .

B. Stable Neural Dynamical Systems (SNDS)

Assuming \mathcal{D} represents the expert trajectory, Stable Neural Dynamical Systems (SNDS) [17] efficiently learns a stable dynamical policy,

$$\dot{\mathbf{x}} = \pi_\theta(\mathbf{x}), \quad \pi_\theta : \mathbb{R}^d \rightarrow \mathbb{R}^d. \quad (1)$$

Modeled by an ordinary differential equation, $\pi_\theta(\mathbf{x})$ is optimized such that for any initial state, $\mathbf{x}_0 \in \mathcal{D}$, the forward Euler method ¹ generates a sequence,

$$\tau_{\mathbf{x}_0}^{\pi_\theta} = \{\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_M \mid \mathbf{x}_{m+1} = \mathbf{x}_m + \pi_\theta(\mathbf{x}_m)\Delta t\}, \quad (2)$$

¹This holds for any method used to solve an initial value problem, with the forward Euler method selected for simplicity.

which accurately replicates the corresponding trajectory in \mathcal{D} for sufficiently large M . On top of that, SNDS employs Lyapunov theory [18] to ensure global stability [17], [19], meaning that the sequence $\tau_{\mathbf{x}_0}^{\pi_\theta}$ converges to a predefined equilibrium, \mathbf{g} , for any arbitrary initial condition, $\mathbf{x}_0 \notin \mathcal{D}$.

A dynamical policy exhibits global stability if there exists a positive-definite function $v : \mathbb{R}^d \rightarrow \mathbb{R}$, known as a Lyapunov candidate, such that $\dot{v}(\mathbf{x}) < 0$ for all $\mathbf{x} \neq \mathbf{g}$ and $\dot{v}(\mathbf{x}) = 0$. SNDS learns both the policy, π_θ , and the Lyapunov candidate, v , by minimizing the following hybrid loss, $\mathcal{L}(\pi_\theta, \mathcal{D})$, on expert data:

$$\gamma \mathbb{E}_{\mathbf{x}, \dot{\mathbf{x}} \in \mathcal{D}} [(\pi_\theta(\mathbf{x}) - \dot{\mathbf{x}})^2] + (1 - \gamma) \mathbb{E}_{\mathbf{x}_i, \tau_{\mathbf{x}_i}^{\mathcal{D}} \in \mathcal{D}} [(\tau_{\mathbf{x}_i}^{\pi_\theta} - \tau_{\mathbf{x}_i}^{\mathcal{D}})^2], \quad (3)$$

in both position and velocity spaces. In Eq. 3, $\gamma \in [0, 1]$ controls the trade-off between the position and velocity components in the loss function. The terms $\tau_{\mathbf{x}_i}^{\mathcal{D}}$ and $\tau_{\mathbf{x}_i}^{\pi_\theta}$ represent partial trajectories from the demonstration data \mathcal{D} and those generated by π_θ , respectively.

C. Automatic Waypoint Extraction

Automatic Waypoint Extraction (AWE) aims to find waypoints that can reproduce a demonstration. Specifically, it aims to select a set of waypoints \mathbf{W} and reconstruct a trajectory by interpolating between every consecutive pairs of waypoints. AWE is formulated as an optimization problem

$$\begin{aligned} \min_{\mathbf{W}} & \|\mathbf{W}\| \\ \text{s.t. } & \mathcal{L}(f(\mathbf{W}), \mathcal{D}) \leq \eta \end{aligned} \quad (4)$$

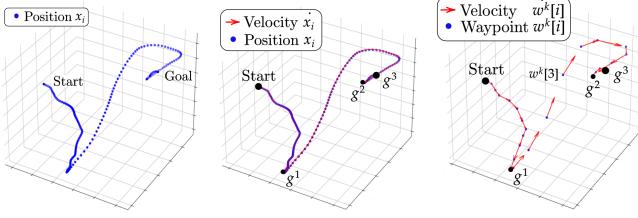
where $f(\mathbf{W})$ is a function that takes a set of waypoints and reconstructs a trajectory, \mathcal{L} is the loss function, and η is a threshold of maximum loss.

The parameter η allows the user to decide the precision of the reconstructed motion. A lower η indicates that the user prefers high-precision imitation, in which the method will automatically select more waypoints to fit the demonstration \mathcal{D} . In contrast, a higher η will yield a smaller set of waypoints from Equ. 4, and the reconstructed trajectory will be a rougher approximation of \mathcal{D} .

D. Discussion

The prior work on data-driven methods for learning stable policies (such as the one in Sec. II-B) was designed for solving a single task. In our work, we will adapt the same network architecture for each subtask of a long-horizon problem (see Sec. III-B).

AWE, discussed in Sec. II-C automatically selects waypoints given demonstrations. The hyper-parameter η in Equ. 4 allows the user to make a tradeoff between accuracy and smoothness. If a demonstration contains a discontinuous movement, AWE typically marks it as a waypoint. However, such motion can be the intention of the demonstration or the results of noise present in the data. In our work, we adapt AWE to select waypoints within a segment of a long horizon manipulation task (Sec. III-A). This allows smooth movement within a segment without sacrificing the precision for achieving the subtask.



(a) Expert trajectory (b) Selected subgoals (c) AWE waypoints

Fig. 2: An example of (a) a single expert demonstration in the robot's task space, (b) three subgoals selected, and (c) outcomes of the automatic waypoint selections in each segment.

III. METHODS

We aim to learn long-horizon manipulation tasks in a one-shot manner using a hierarchical approach. Our approach leverages [20], breaking down the policy into high-level decision-making and low-level motion planning. At the high level, we define the task as a series of subgoals (Sec III-A). Next, a unique and stable dynamical policy is learned for each segment based on the corresponding sub-demo (Sec III-B). Lastly, at execution time, we reconstruct the trajectory by selecting the appropriate policy to perform each part of the task (Sec. III-C).

A. Subgoal Identification and Waypoint Selection

Our first step is to identify key states in the trajectory where major stages of the overall task take place, thereby breaking down complex trajectories into more manageable segments for learning. We opt for a straightforward method, defining a subgoal as the activation of the gripper. Our insight is that in most household tasks, meaningful subgoals typically involve the hand or gripper either grasping or releasing an object. By defining these actions as subgoals, we can divide the demonstrations into sub-demos, where each segment can be easily described by a single dynamical policy.

Formally, we define untrimmed demonstrations as $\mathcal{D} = \{(\mathbf{x}_1, \dot{\mathbf{x}}_1), (\mathbf{x}_2, \dot{\mathbf{x}}_2), \dots, (\mathbf{x}_N, \dot{\mathbf{x}}_N)\}$, we need to find the key frames that divide the demonstration \mathcal{D} into K sub-demos, such that \mathcal{D}^k resembles a simple motion conditioned on a subgoal \mathbf{g}^k .

We perform a forward pass in the trajectory to find indices h^1, h^2, \dots, h^K such that \mathbf{x}_{h^k} denotes the k^{th} occasions where the gripper opens or closes. Based on the selected indices, we divided the trajectory \mathcal{D} into K segments, and define the subgoal for each segment as $\mathbf{g}^k = \mathbf{x}_{h^k}$.

An example can be seen in Fig. 2. The original demonstration (blue) contains a discontinuous motion (Fig. 2a), and our method labels the transition points with 3 subgoals $\mathbf{g}^1, \mathbf{g}^2, \mathbf{g}^3$ (Fig. 2b).

From each segment \mathcal{D}^k , we filter out the noise and reduce the data complexity by finding waypoints that approximate \mathcal{D}^k , thereby simplifying the task even further. We leverage AWE (see Sec. II-C) to automatically extract waypoints from data (using Equ. 4). For each segment, a set of waypoints

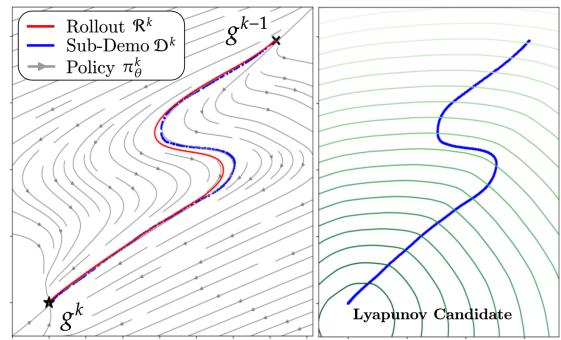


Fig. 3: Stable dynamical policy rollout by an optimized SNDS model (left), and its Lyapunov candidate (right). The learned Lyapunov candidate ensures the induced trajectories always move toward the lowest energy point, regardless of the initial state or perturbations.

is selected $\mathbf{W}^k = \{\mathbf{w}_0^k, \mathbf{w}_1^k, \dots, \mathbf{g}^k\}$ and the last waypoint is simply the subgoal of the segment.

AWE selects waypoints sparingly when the trajectory is straight and more densely when the trajectory is complex, providing more data where necessary and less where it's not. By adjusting the trajectory reconstruction loss threshold η , the smoothness of the trajectory can be controlled.

Fig. 2c is an example of waypoint extraction. Note that only a small amount of data is retained since this is sufficient to reconstruct the original demonstration through interpolations between waypoints.

Remark 1.1: In contrast to the previous work, the waypoint extraction is performed solely within a segment. Our insight is that the most important requirement of manipulation tasks lies in achieving the subgoal, while precise imitation may not be essential. We can filter the noise or sacrifice the accuracy of imitation by reconstructing the demonstration with the sampled waypoints, but the subgoal cannot be approximated.

B. Learning Dynamical Systems

The second step is to learn a set of K models that can reproduce the motion from the segmented trajectories derived from Sec. III-A. For each k^{th} segment, we train an SNDS policy, $\tilde{\mathbf{x}} = \pi_\theta^k(\mathbf{x})$ with the following objective, with the loss \mathcal{L} defined in Eq. 3.

$$\theta_k^* \triangleq \arg \min_{\theta \in \mathbb{R}^{N_\theta}} \mathcal{L}(\pi_\theta^k, \mathcal{D}^k) \quad (5)$$

The training process is conducted exclusively on the data from the k^{th} segment, denoted as \mathcal{D}^k , with the subgoal \mathbf{g}^k set as the stable equilibrium. Since π_θ is formulated using standard automatic differentiation tools, the optimization problem can be efficiently solved to determine the optimal parameter, θ^* .

An example is illustrated in as illustrated in Fig. 3. The left figure shows the segmented demonstration \mathcal{D}^k (blue) and the policy rollout (red) from the dynamical system π_θ^k . The grey arrows represent the vector field produced by

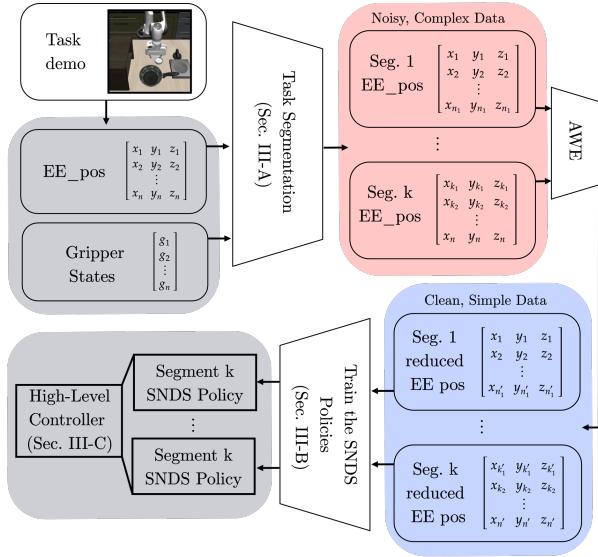


Fig. 4: Our framework on learning stable policies for long-horizon manipulation tasks

the dynamical system. In regions where demonstrations are absent, the motion is mostly determined by the Lyapunov candidate function (right) which enforces movement toward the subgoal.

Remark 2.1: As explained in Sec II-B, the representation of SNDS is specifically designed to enforce global stability. Consequently, each learned policy, π_θ^k , generates velocities $\tilde{\mathbf{x}}$ based on the current state to imitate expert data within the segment, ensuring that all trajectories converge to the subgoal \mathbf{g}^k . For this reason, even in the presence of external disturbances or noisy inputs, the dynamical system can still bring the robot to the subgoal.

C. Stable Task Reproduction

Having learned a unique dynamical system for each segment, what remains is to define a high-level controller \mathcal{C} to imitate the task by returning desired velocities at each state during execution.

The high-level controller \mathcal{C} takes as input the set of subgoals (from Sec. III-A) and learned dynamical systems (from Sec. III-B). At each time step, the high-level controller evaluates the current state \mathbf{x} and determines which subgoal should be the target and whether the current subgoal was achieved, based on a distance threshold ε^k . The parameter ε^k can be adjusted depending on the nature of the task.

Then, the high-level controller applies policy $\tilde{\mathbf{x}} = \pi_\theta(\mathbf{x}, \mathbf{g}^k)$ and executes the predicted velocity $\tilde{\mathbf{x}}$ during the execution of segment k of the trajectory. Specifically,

$$\tilde{\mathbf{x}} = \begin{cases} \pi_\theta^k(\mathbf{x}), & \|\mathbf{x} - \mathbf{g}^k\| > \varepsilon^k \\ \pi_\theta^{k+1}(\mathbf{x}), & \|\mathbf{x} - \mathbf{g}^k\| \leq \varepsilon^k \wedge k < K-1 \\ \pi_\theta^K(\mathbf{x}), & k = K-1 \end{cases} \quad (6)$$

Proposition 3.1: The high-level policy outlined in Eq. 6 is globally stable at the last subgoal, \mathbf{g}^K .



Fig. 5: Examples of tasks in Robosuite (top) and an overview of the task demonstrations (bottom).

Proof. The proof is intuitive and follows from the global stability of each low-level dynamical policy. Formally, according to the Lyapunov global stability theorem,

$$\forall \mathbf{x}_0^k \in \mathcal{D}^k, \lim_{t \rightarrow \infty} \mathbf{x}_t = \mathbf{g}^k, \text{ if } \mathbf{x}_{t+1} = \mathbf{x}_t + \Delta t \pi_\theta^k(\mathbf{x}),$$

for sufficiently small Δt . Then, every $\mathbf{g}^k \in \mathcal{D}^k$ of $\pi_\theta^k(\mathbf{x})$ can be viewed as the initial condition of $\pi_\theta^{k+1}(\mathbf{x})$: \mathbf{g}^{k+1} . Hence, the core definition of global stability can be written seen as a cascade, yielding: $\forall \mathbf{x}_0^k \in \mathcal{D}, \lim_{t \rightarrow \infty} \mathbf{x}_t = \mathbf{g}^K$, when $\mathbf{x}_{t+1} = \mathbf{x}_t + \Delta t \pi_\theta(\mathbf{x})$. \square

A summary of our proposed method is illustrated in Fig. 4. This architecture ensures resilience against noise and external perturbations. This allows the system to quickly return to the original path and avoid collisions in cluttered environments.

IV. EXPERIMENTS

Our experiments aim to demonstrate the following questions:

- 1) Can we improve the overall success rate by enforcing the success of each subtask?
- 2) Can we learn a long-horizon manipulation task from a single demonstration?

A. Evaluation Criteria

To evaluate the performance, we consider both the success rate on the success rate of completing (1) the sub-task within each segment and (2) the whole task.

B. Baselines

We compare our work against the following baselines.

- 1) **BC**: standard behavioral cloning
- 2) **SNDS**: original Stable Neural Dynamical Systems [17]
- 3) **BC+Ours**: To evaluate our proposed work on trajectory segmentation with waypoint selections (Sec. III-A), we also compare the performance of using standard behavior cloning on each segment data.

Lastly, we use **Ours** to denote the model trained with our method described in both Sec. III-A and Sec. III-B

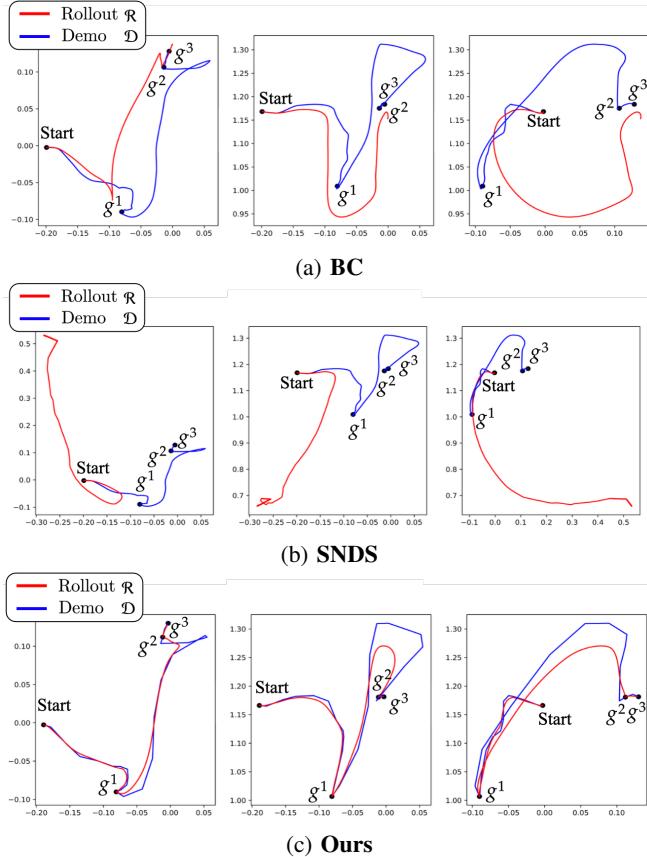


Fig. 6: 2D projections (from left to right: xy, xz, yz) of policy rollouts (red) learned from demo (blue) in **Ketchup** task in a *deterministic* environment.

C. Experimental setups

We evaluate our work in Robosuite [21] with tasks defined in public benchmark LIBERO [22] and robomimic [23] (see Fig. 5). We selected the following pick-and-place tasks from the benchmark:

- 1) Ketchup: grab a ketchup bottle and place it in a drawer.
- 2) Square: pick up a square nut and fit it on a square rod.
- 3) Wine: grab a wine bottle and set it on a rack.
- 4) Bowl: lift a small bowl and place it on a cabinet.
- 5) Cheese: pick up a bar of cheese and put it in a bowl.

Each demonstration contains approximately 100 to 250 data points. We use our method described in Sec. III-A to divide the demonstration and select waypoints with an error threshold of $\eta = 0.01$ for each segment. This yields 10 to 25 waypoints from the demonstration. See Fig. 5 for more information on the tasks we evaluate on.

We define the state of the system at the end-effector position and the gripper state. The control commands are generated through standard inverse kinematics. For each task, all policies are trained from **one demonstration** over 10,000 epochs. We evaluate both the overall task completion and the success of each individual subgoal. That is, we assess subgoals independently—if subgoal 1 fails (due to a timeout), the environment is reset to the state of a successful

TABLE I: Success rate (%) for behavior cloning benchmark and our work in *noisy* and *perturbed & noisy* environments across various tasks.

Task	Subgoal	Noisy		Perturbed & Noisy	
		BC+Ours	Ours	BC+Ours	Ours
Ketchup	1	96.7 ± 5.8	100.0 ± 0.0	73.3 ± 5.8	93.3 ± 5.8
	2	100.0 ± 0.0	100.0 ± 0.0	96.7 ± 5.8	96.7 ± 5.8
	3	100.0 ± 0.0	100.0 ± 0.0	96.7 ± 5.8	100.0 ± 0.0
	total	96.7 ± 5.8	100.0 ± 0.0	60.0 ± 10.0	80.0 ± 0.0
Square	1	100.0 ± 0.0	100.0 ± 0.0	70.0 ± 0.0	70.0 ± 0.0
	2	100.0 ± 0.0	100.0 ± 0.0	66.7 ± 5.8	70.0 ± 0.0
	3	100.0 ± 0.0	100.0 ± 0.0	90.0 ± 17.3	100.0 ± 0.0
	total	56.7 ± 15.3	56.7 ± 11.5	53.3 ± 11.5	70.0 ± 0.0
Wine	1	100.0 ± 0.0	100.0 ± 0.0	90.0 ± 0.0	90.0 ± 0.0
	2	100.0 ± 0.0	100.0 ± 0.0	30.0 ± 30.0	70.0 ± 0.0
	3	100.0 ± 0.0	100.0 ± 0.0	30.0 ± 30.0	80.0 ± 17.3
	total	63.3 ± 15.3	63.3 ± 15.3	20.0 ± 17.3	50.0 ± 10.0
Bowl	1	90.0 ± 17.3	100.0 ± 0.0	56.7 ± 5.8	93.3 ± 11.5
	2	100.0 ± 0.0	100.0 ± 0.0	90.0 ± 10.0	100.0 ± 0.0
	3	100.0 ± 0.0	100.0 ± 0.0	93.3 ± 11.5	97.6 ± 5.8
	total	83.3 ± 20.8	97.6 ± 5.8	46.7 ± 15.3	70.0 ± 0.0
Cheese	1	80.0 ± 26.5	100.0 ± 0.0	33.3 ± 32.1	60.0 ± 26.5
	2	100.0 ± 0.0	100.0 ± 0.0	53.3 ± 20.8	60.0 ± 0.0
	3	100.0 ± 0.0	100.0 ± 0.0	90.0 ± 17.3	100.0 ± 0.0
	total	80.0 ± 26.5	93.3 ± 11.5	16.7 ± 11.5	50.0 ± 10.0

subgoal 1, and the simulation continues with subgoal 2. Evidently, if the model completes the task after a reset to a later subgoal, it is not considered a successful attempt. All experiments are done with a maximum horizon of 1000 actions per subgoal.

During execution, the learned dynamical system outputs the linear velocity of the end-effector based on the current end-effector position. The angular movement is calculated separately using simple spherical linear interpolation (SLERP), as most of the manipulation tasks do not involve complex movement in the orientations. As long as the gripper reaches the subgoal with the correct orientation, the speed and timing of orientation adjustments do not affect task success. At each time-step, \mathcal{C} evaluates whether the current sub-goal was achieved, based on a distance threshold $\varepsilon^k = 0.008\text{m}$, which is precise enough for pick-and-place tasks.

D. Results in Deterministic Environments

We begin by evaluating all baseline models and our approach in deterministic environments. Each model is trained using 10 different random seeds, and the simulation is ran once to assess success or failure. Our results demonstrate **perfect success rates** (100%) across training seeds for both **Ours** and **BC+Ours**, while **BC** and **SNDS**, which lack segmentation, exhibit **near-zero success rates**—even when success is evaluated at the subgoal level. Figure 6 compares the rollouts of each model with the demonstration, clearly highlighting the poor performance of non-segmented models in contrast to their segmented counterparts. Since **BC+Ours** has a similar performance with **Ours** in this experiment, only one plot is shown here.

This behaviour is expected since dividing the task into

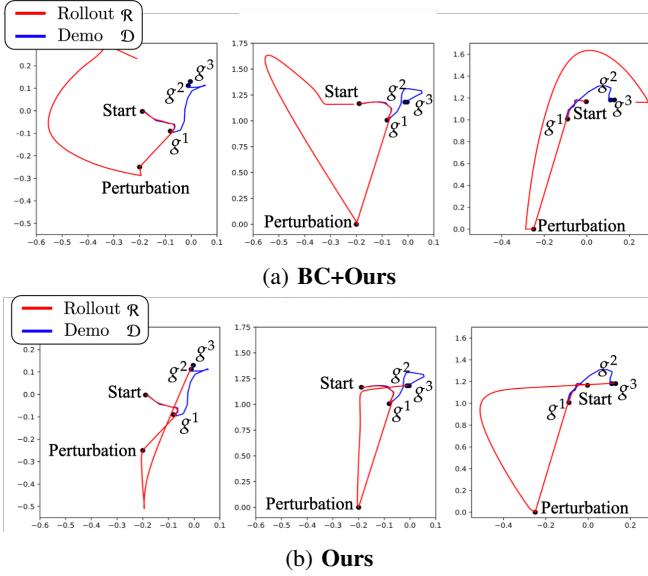


Fig. 7: 2D projections (from left to right: xy, xz, yz) of policy rollouts (red) learned from demo (blue) in **Ketchup** task with perturbations injected.

segments and then sampling the waypoints drastically simplifies the complexity of the motion and learning becomes easier. Also, those methods learn to be conditioned on achieving a specific sub-goal. Methods that take the whole demonstration as an input are unlikely to learn well with a single demonstration.

E. Rollouts with Noise and Perturbations

To further evaluate how a learned policy performs in a realistic setting, we add noise and perturbations during rollouts. Since **BC** and **SNDS** have not achieved any tasks in a deterministic setting, we only compare **BC + Ours** and **Ours** in this section.

We add Gaussian noise with a standard deviation of 0.01 to the end-effector position feedback. This is comparable to the noise level of a real robotic system. In a real-world application, the robot might experience unexpected disturbances (e.g., from inexperienced users). For this, we also investigate the effect of perturbations while executing the policy. We train different policies using 3 random seeds and evaluate each model 10 times. We repeat the same experiments and the average success rates across all 5 tasks are reported in Table I.

For both scenarios, we can see that **Ours** consistently performs better than **BC + Ours**. This is due to our choice of dynamical systems that enforce the movement toward each sub-goal. Therefore, it is robust under noise and disturbances.

To visualize the performance of the rollouts, we plot the demonstration and the rollout in Fig. 7. We can see that the rollout generated from **BC + Ours** (Fig. 7a) diverges from the trajectory and is unable to recover. Note that, the rollouts are generated using models trained from data using our segmentation and waypoint methods described in Sec. III-A, and each segment is a relatively simple task. However,

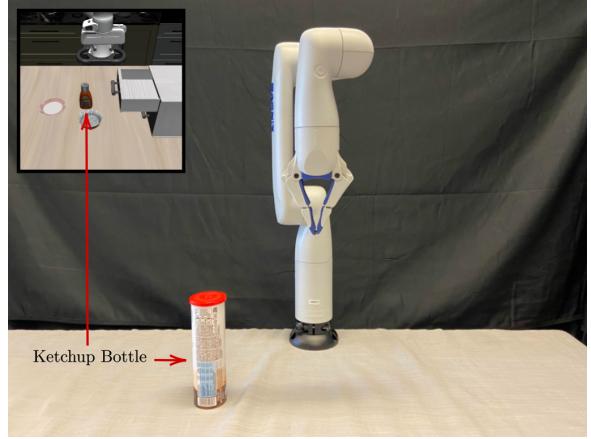


Fig. 8: Visualization of tasks on robotic hardware.

this phenomenon is expected since the perturbation pushes the robot to a region that is not covered by the demonstration. Similar to most standard data-driven algorithms, the model does not respond well to out-of-distribution scenarios.

In contrast, as seen from Fig. 7b, **Ours** is robust to perturbations and re-plans as needed. Even with perturbations, the constraints imposed by the Lyapunov condition can guide the robot back to the next subgoal. While most work in literature requires a large dataset to train a policy, our work only requires a single demonstration.

F. Zero-shot sim-to-real

The dynamical systems were trained with *one demonstration* in simulation, without further data augmentation. We deploy the dynamical systems learned from the simulation on robotic hardware (see Fig. 8). Note that, we did not observe any *sim-to-real gap*. This is because the Lyapunov condition ensures that the motion is constrained toward sub-goals, diminishing the likelihood of divergence.

V. CONCLUSION

In this paper, we build upon prior work on learning long-horizon manipulation tasks and stable learning with dynamical systems, with the goal of improving task success rates of execution while minimizing the required training data. We introduce a novel approach that (1) decomposes long-horizon demonstrations into segments defined by waypoints and subgoals, and (2) learns globally stable dynamical system policies that drive the robot toward each subgoal, even in the presence of sensory noise and stochastic disturbances. The proposed method is validated both in simulation and on robotic hardware, demonstrating seamless transfer from simulation to real-world implementation.

In this work, we demonstrate a proof-of-concept using a sequence of stable dynamical systems for long-horizon manipulation tasks. Our future work will investigate various automatic segmentation methods with visual feedback [10], [24] and make our work more generalizable to unseen scenarios. We will also apply this learning regime to a more complex state representation.

REFERENCES

- [1] S. Nair and C. Finn, "Hierarchical foresight: Self-supervised learning of long-horizon tasks via visual subgoal generation," in *International Conference on Learning Representations*, 2020.
- [2] Y. Lee, E. S. Hu, and J. J. Lim, "Ikea furniture assembly environment for long-horizon complex manipulation tasks," in *IEEE International conference on robotics and automation (ICRA)*, 2021, pp. 6343–6349.
- [3] C. Lynch, M. Khansari, T. Xiao, V. Kumar, J. Tompson, S. Levine, and P. Sermanet, "Learning latent plans from play," in *Conference on robot learning*. PMLR, 2020, pp. 1113–1132.
- [4] L. Kou, F. Ni, Y. ZHENG, J. Liu, Y. Yuan, Z. Dong, and J. HAO, "KISA: A unified keyframe identifier and skill annotator for long-horizon robotics demonstrations," in *International Conference on Machine Learning*, 2024.
- [5] D. Morrison, P. Corke, and J. Leitner, "Closing the loop for robotic grasping: A real-time, generative grasp synthesis approach," *Robotics: Science and Systems XIV*, pp. 1–10, 2018.
- [6] Y. Ding, C. Florensa, P. Abbeel, and M. Philipp, "Goal-conditioned imitation learning," *Advances in neural information processing systems*, vol. 32, 2019.
- [7] E. Jang, A. Irpan, M. Khansari, D. Kappler, F. Ebert, C. Lynch, S. Levine, and C. Finn, "Bc-z: Zero-shot task generalization with robotic imitation learning," in *Conference on Robot Learning*. PMLR, 2022, pp. 991–1002.
- [8] B. Eysenbach, R. Salakhutdinov, and S. Levine, "C-learning: Learning to achieve goals via recursive classification," in *International Conference on Learning Representations*, 2021.
- [9] C. Dawson, S. Gao, and C. Fan, "Safe control with learned certificates: A survey of neural lyapunov, barrier, and contraction methods," *arXiv preprint arXiv:2202.11762*, 2022.
- [10] Z. Zhang, Y. Li, O. Bastani, A. Gupta, D. Jayaraman, Y. J. Ma, and L. Weihs, "Universal visual decomposer: Long-horizon manipulation made easy," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2024, pp. 6973–6980.
- [11] L. Mezghani, S. Sukhbaatar, P. Bojanowski, A. Lazaric, and K. Alahari, "Learning goal-conditioned policies offline with self-supervised reward shaping," PMLR, pp. 1401–1410, 2023.
- [12] K. Pertsch, O. Rybkin, F. Ebert, S. Zhou, D. Jayaraman, C. Finn, and S. Levine, "Long-horizon visual planning with goal-conditioned hierarchical predictors," *Advances in Neural Information Processing Systems*, vol. 33, pp. 17321–17333, 2020.
- [13] L. X. Shi, A. Sharma, T. Z. Zhao, and C. Finn, "Waypoint-based imitation learning for robotic manipulation," in *Conference on Robot Learning*. PMLR, 2023, pp. 2195–2209.
- [14] S. M. Khansari-Zadeh and A. Billard, "Learning stable nonlinear dynamical systems with gaussian mixture models," *IEEE Transactions on Robotics*, vol. 27, no. 5, pp. 943–957, 2011.
- [15] M. A. Rana, A. Li, D. Fox, B. Boots, F. Ramos, and N. Ratliff, "Euclideanizing flows: Diffeomorphic reduction for learning stable dynamical systems," in *Learning for Dynamics and Control*. PMLR, 2020, pp. 630–639.
- [16] A. Coulombe and H.-C. Lin, "Generating stable and collision-free policies through lyapunov function learning," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 3037–3043.
- [17] A. Abyaneh, M. S. Guzmán, and H.-C. Lin, "Globally stable neural imitation policies," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2024.
- [18] R. L. Devaney, *An introduction to chaotic dynamical systems*. CRC press, 2021.
- [19] J. Z. Kolter and G. Manek, "Learning stable deep dynamics models," *Advances in neural information processing systems*, vol. 32, 2019.
- [20] C. R. Garrett, R. Chitnis, R. Holladay, B. Kim, T. Silver, L. P. Kaelbling, and T. Lozano-Pérez, "Integrated task and motion planning," *Annual review of control, robotics, and autonomous systems*, vol. 4, no. 1, pp. 265–293, 2021.
- [21] Y. Zhu, J. Wong, A. Mandlekar, R. Martín-Martín, A. Joshi, S. Nasiriany, and Y. Zhu, "robosuite: A modular simulation framework and benchmark for robot learning," *arXiv preprint arXiv:2009.12293*, 2020.
- [22] B. Liu, Y. Zhu, C. Gao, Y. Feng, Q. Liu, Y. Zhu, and P. Stone, "Libero: Benchmarking knowledge transfer for lifelong robot learning," 2024.
- [23] A. Mandlekar, D. Xu, J. Wong, S. Nasiriany, C. Wang, R. Kulkarni, L. Fei-Fei, S. Savarese, Y. Zhu, and R. Martín-Martín, "What matters in learning from offline human demonstrations for robot manipulation," 2021. [Online]. Available: <https://arxiv.org/abs/2108.03298>
- [24] L. Kou, F. Ni, Y. Zheng, J. Liu, Y. Yuan, Z. Dong, and H. Jianye, "Kisa: A unified keyframe identifier and skill annotator for long-horizon robotics demonstrations," in *International Conference on Machine Learning*, 2024.