Auto-conditioned Recurrent Mixture Density Networks for Learning Generalizable Robot Skills

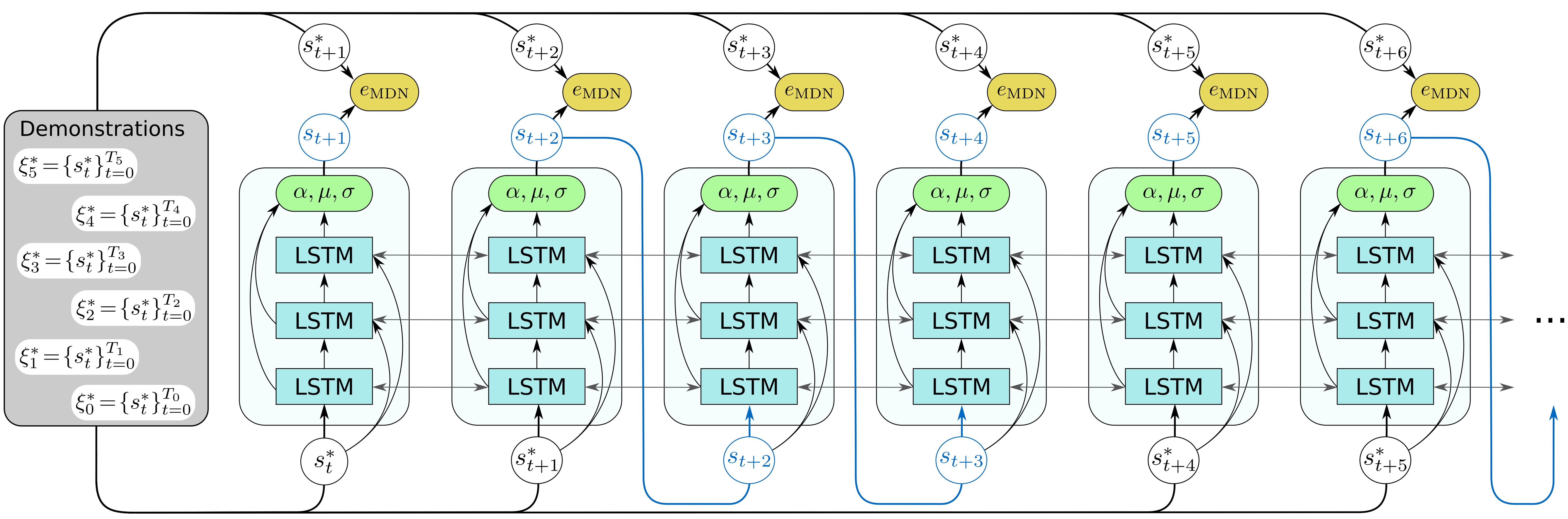
Hejia Zhang Eric Heiden Stefanos Nikolaidis Joseph J. Lim Gaurav S. Sukhatme

Introduction

In this work, we introduce a *state transition model* (STM) that generates joint-space trajectories by imitating motions from expert behavior. The learned STM can quickly generalize to unseen tasks and synthesize motions having longer time horizons than the expert demonstrations.

Approach

Our model combines a *Mixture Density Model* (MDN), to capture the uncertainty and multimodality of the state space, and *Long Short-Term Memory* (LSTM), to synthesize sequences of states. Training our model via *auto-conditioning* allows it to robustly learn trajectories over long time horizons.

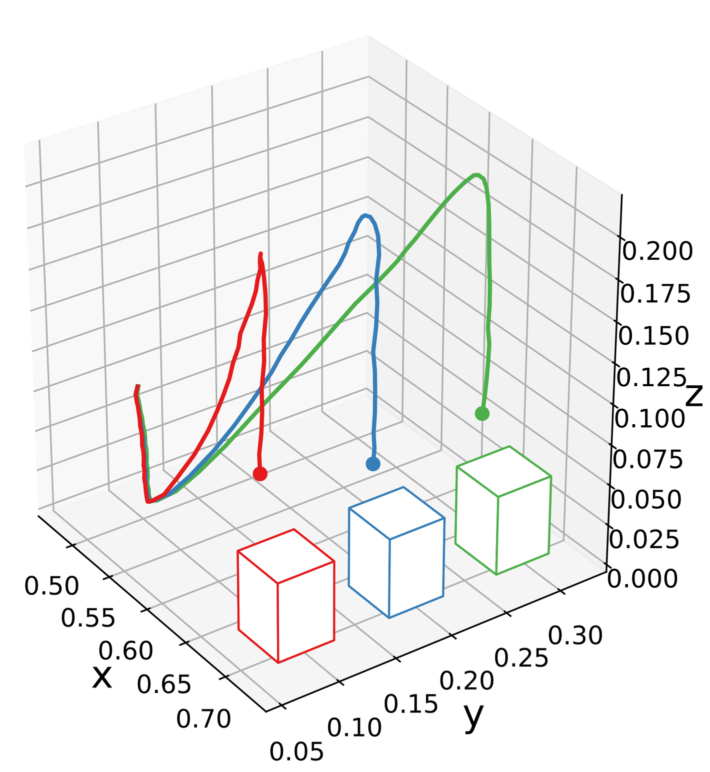


State space: $$(\underbrace{\Delta q_t^0, \Delta q_t^1, \ldots, \Delta q_t^6}_{\text{joint position change}}, \underbrace{\phi_t}_{\text{task input}}, \underbrace{\psi_t}_{\text{task description}})$$

Architecture of the proposed auto-conditioned recurrent mixture density network to model state transitions, unrolled over 6 time steps, with auto-conditioning length 2 and groundtruth length 2.

Experiments

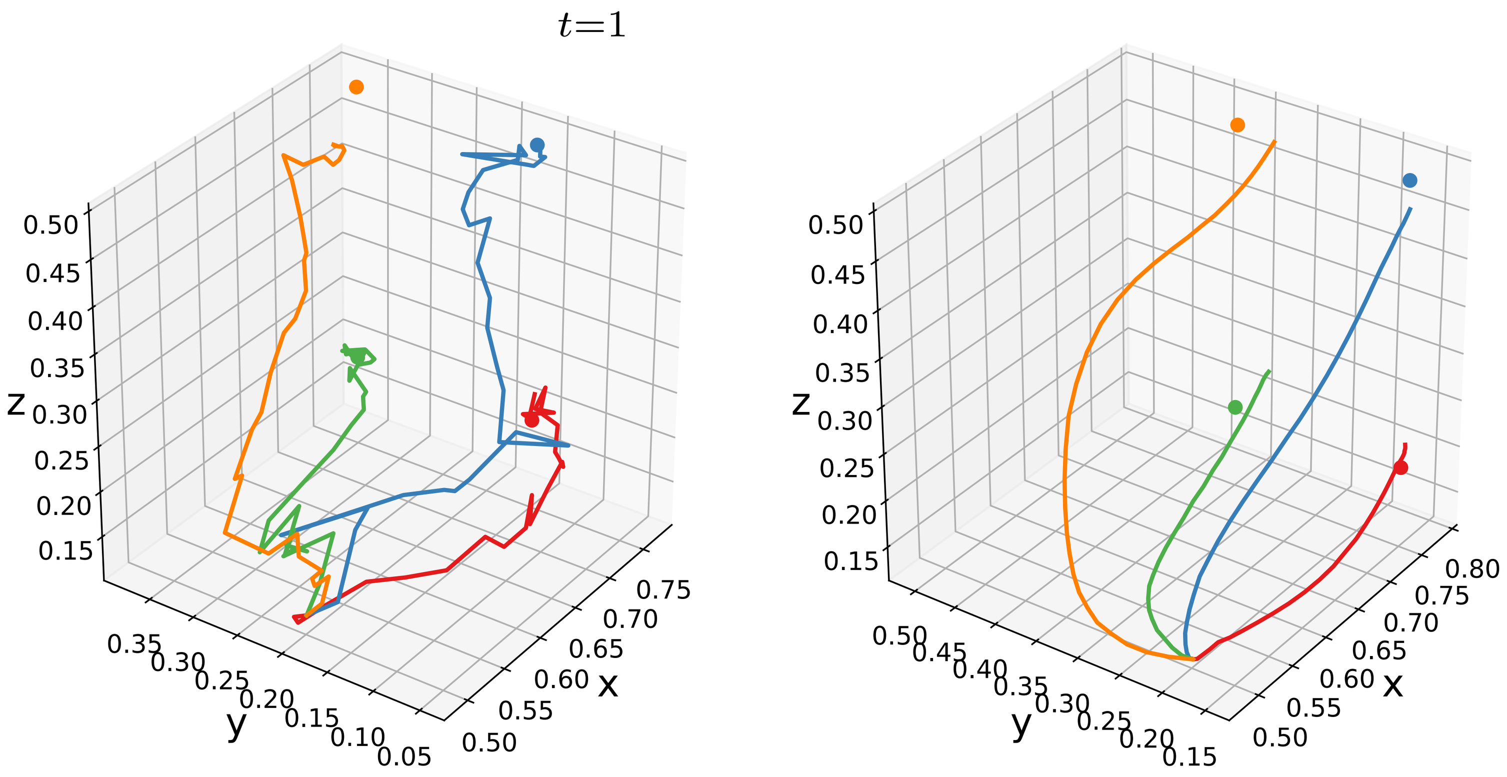
We evaluate our model and training procedure on various tasks on the Sawyer robot arm, such as servoing, pick-and-place, and scenarios where the robot has to draw circles with a given radius. By training our model and various demonstrations, it is able to generalize to many different initial and goal conditions for each task.



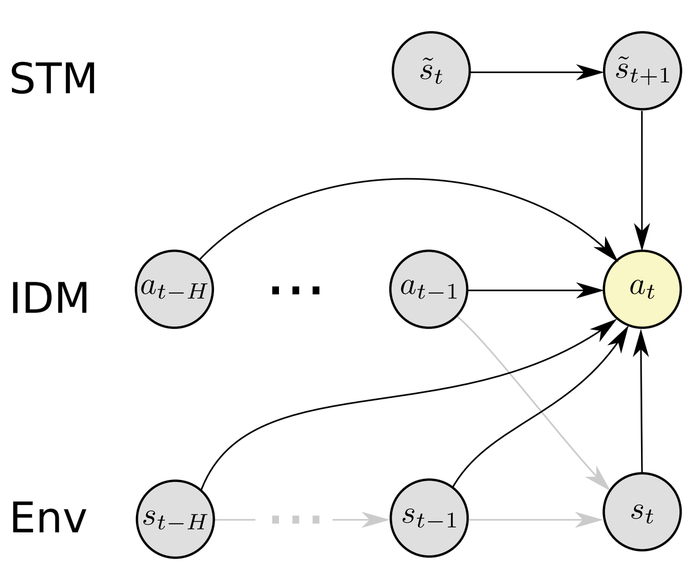
Trajectory optimization

To generate smooth trajectories, we optimize the generated motions using the following minimization criterion:

$$V(\{\mathbf{q}_t\}_{t=1}^{T}) = \sum_{t=1}^{T-1} \|\mathbf{q}_t - \tilde{\mathbf{q}}_t\|_2^2 + \gamma \|\mathbf{q}_{t+1} - \mathbf{q}_t\|_2^2$$



Combination with Inverse Dynamics Model



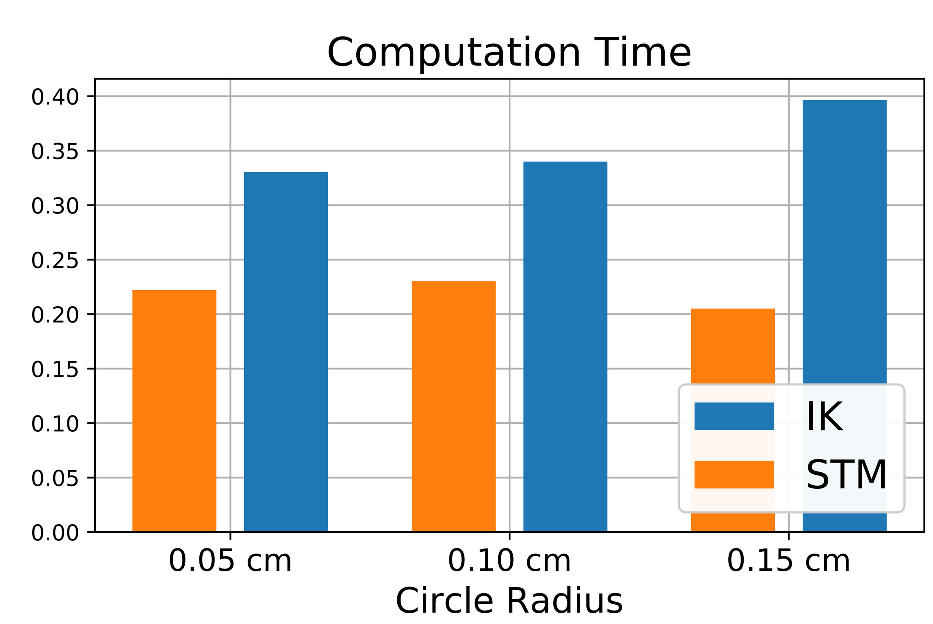
While the STM is able to generate states, i.e. joint positions, a joint position controller is required to execute the trajectories. In this experiment, we learned such *inverse dynamics model* (IDM) through self-supervised learning.

Results

Ablation studies over different model architecture choices experimentally validate our auto-conditioned LSTM-MDN approach.

	Reacher	Pick-and-place	Stacking
LSTM	80%	0%	0%
a.c. LSTM	90%	50%	25%
LSTM-MDN	90%	60%	30%
a.c. LSTM-MDN (Ours)	100%	100%	80%

Thanks to the neural network architecture, our model allows for fast inference times, outpacing traditional Inverse Kinematics solvers on a circle-drawing task.



Adaptation to changing goals

The auto-conditioned LSTM-MDN is able to generate trajectories much longer than the demonstrated sequences. It can adapt online to changing goals, which we demonstrate on reaching and pick-and-place tasks (Fig. below) where the the target is moved midway through the trajectory generation. Our model continues to synthesize motions that fulfill the new task descriptions.