Code Available

Language Control Diffusion: Efficiently Scaling Through Space, Time, and Tasks

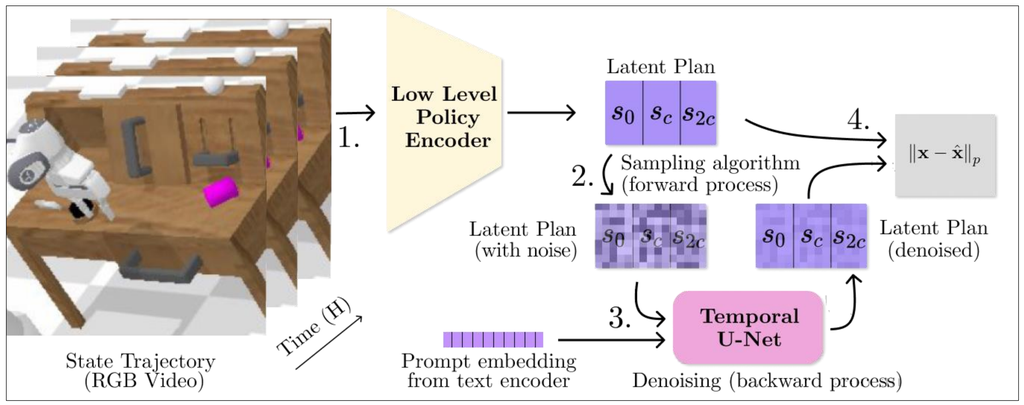
Edwin Zhang
Harvard, Founding

Yujie Liu, Shinda Huang, William Wang
University of California, Santa Barbara

Amy Zhang
UT Austin

1 Abstract

- Training generalist agents is difficult across several axes such as high-dimensional inputs (space), long horizons (time), and generalization to novel tasks
- We address all three axes by leveraging **L**anguage to **C**ontrol **D**iffusion models as a hierarchical planner conditioned on language (LCD)
- LCD outperforms other SOTA methodologies in multi-task success rates
- Improves inference speed over other comparable diffusion models by 3.3x~15x



2 Objective

We propose a hierarchical planner to train a generalist agent in the CALVIN benchmark to listen to language commands using language conditioned reinforcement learning and goal conditioned imitation learning.

3 Methodology

Hierarchical Diffusion Policies

$$\min_{\pi} \mathbb{E}_{s,\mathcal{R}\sim\mathcal{D}}\left[D_{\mathrm{KL}}\left(\pi_\beta(\cdot \mid s,\mathcal{R}), \pi(\cdot \mid s,\mathcal{R})\right)\right]. \quad (1)$$

$$\min_{\mathcal{P}} D_{\mathrm{KL}}\left(\mathcal{P}_\beta(\boldsymbol{\tau} \mid \mathcal{R}), \mathcal{P}(\boldsymbol{\tau} \mid \mathcal{R})\right) = \min_{\mathcal{P}} \mathbb{E}_{\boldsymbol{\tau},\mathcal{R}\sim\mathcal{D}}\left[\log \mathcal{P}_\beta(\boldsymbol{\tau} \mid \mathcal{R}) - \log \mathcal{P}(\boldsymbol{\tau} \mid \mathcal{R})\right]. \quad (2)$$

$$\min_{\theta} \mathbb{E}_{\boldsymbol{\tau}_0,\boldsymbol{\epsilon}}[\|\boldsymbol{\epsilon}_t - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\boldsymbol{\tau}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_t, t)\|^2]. \quad (3)$$

Starting from imitation learning (1), we define a state trajectory generator P in (2). This is reformulated to the diffusion training objective (3).

Practical Instantiation

Algorithm 1 Hierarchical Diffusion Policy Training

Input: baseline goal-conditioned policy $\pi_{\mathrm{lo}} := \phi(\mathcal{E}(s_t), g_t)$, diffusion variance schedule α_t, temporal stride c, language model ρ
Output: trained hierarchical policy $\pi(a_t|s_t) := \pi_{\mathrm{lo}}(a_t|s_t, \pi_{\mathrm{hi}}(g_t|s_t))$, where g_t is sampled every c time steps from π_{hi} as the first state in latent plan $\boldsymbol{\tau}^c$.

1: Collect dataset $\mathcal{D}_{\mathrm{onpolicy}}$ by rolling out trajectories $\boldsymbol{\tau} \sim \pi_{\mathrm{lo}}, \rho$
2: Instantiate π_{hi} as diffusion model $\epsilon_\theta(\boldsymbol{\tau}_{\mathrm{noisy}}, t, \rho(L))$
3: **repeat**
4: Sample mini-batch $(\boldsymbol{\tau}, L) = B$ from $\mathcal{D}_{\mathrm{onpolicy}}$.
5: Subsample latent plan $\boldsymbol{\tau}^c = (\mathcal{E}(s_0), \mathcal{E}(s_c), \mathcal{E}(s_{2c}), ..., \mathcal{E}(s_T))$.
6: Sample diffusion step $t \sim \mathrm{Uniform}(\{1, ..., T\})$, noise $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
7: Update high-level policy π_{hi} with gradient $-\nabla_\theta \|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t}\boldsymbol{\tau}^c + \sqrt{1-\bar{\alpha}_t}\epsilon, t, \rho(L))\|^2$
8: **until** converged

High-Level Policy: Before this point in Algorithm 1, we assume that a low-level policy (LLP) has been trained. We first train our LLP through hindsight relabeling.
Low-Level Policy: We adopt the HULC architecture for our low-level policy, and the state encoder as the visual encoder within the HULC policy.
Model Architecture: We adopt the T5-XXL model as our textual encoder, which contains 11B parameters and outputs 4096 dimensional embeddings.

4 Results

Experimental Setup: Dataset, Metric, and Baselines

Horizon	GCBC	MCIL	HULC	SPIL	Diffuser-1D	Diffuser-2D	Ours
One	64.7 ± 4.0	76.4 ± 1.5	82.6 ± 2.6	84.6 ± 0.6	47.3 ± 2.5	37.4 ± 3.2	$\mathbf{88.7 \pm 1.5}$
Two	28.4 ± 6.2	48.8 ± 4.1	64.6 ± 2.7	65.1 ± 1.3	18.8 ± 1.8	9.3 ± 1.3	$\mathbf{69.9 \pm 2.8}$
Three	12.2 ± 4.1	30.1 ± 4.5	47.9 ± 3.2	50.8 ± 0.4	5.9 ± 0.4	1.3 ± 0.2	$\mathbf{54.5 \pm 5.0}$
Four	4.9 ± 2.0	18.1 ± 3.0	36.4 ± 2.4	38.0 ± 0.6	2.0 ± 0.5	0.2 ± 0.0	$\mathbf{42.7 \pm 5.2}$
Five	1.3 ± 0.9	9.3 ± 3.5	26.5 ± 1.9	28.6 ± 0.3	0.5 ± 0.0	0.07 ± 0.09	$\mathbf{32.2 \pm 5.2}$
Avg horizon len	1.11 ± 0.3	1.82 ± 0.2	2.57 ± 0.12	2.67 ± 0.01	0.74 ± 0.03	0.48 ± 0.09	$\mathbf{2.88 \pm 0.19}$

LCD is good for lang. cond. RL

	GCBC	HT	Ours
Seen	45.7 ± 2.5	38.0 ± 2.2	$\mathbf{53.7 \pm 1.7}$
Unseen	46.0 ± 2.9	36.7 ± 1.7	$\mathbf{54.0 \pm 5.3}$
Noisy	42.7 ± 1.7	33.3 ± 1.2	$\mathbf{48.0 \pm 4.5}$

LCD can generalize to new tasks

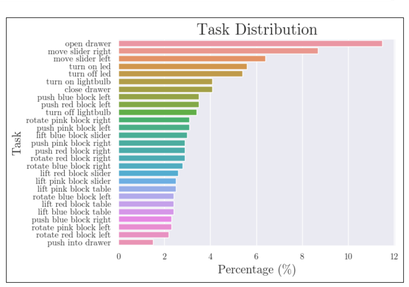
Task	Diffuser-1D	Ours
Lift Pink Block Table	31.67 ± 10.27	$\mathbf{55.00 \pm 16.33}$
Lift Red Block Slider	13.35 ± 10.25	$\mathbf{88.33 \pm 8.50}$
Push Red Block Left	1.64 ± 2.35	$\mathbf{35.00 \pm 7.07}$
Push Into Drawer	3.34 ± 4.71	$\mathbf{90.00 \pm 10.80}$
Rotate Blue Block Right	5.00 ± 4.10	$\mathbf{36.67 \pm 14.34}$
Avg SR	12.67 ± 3.56	$\mathbf{61.00 \pm 7.79}$

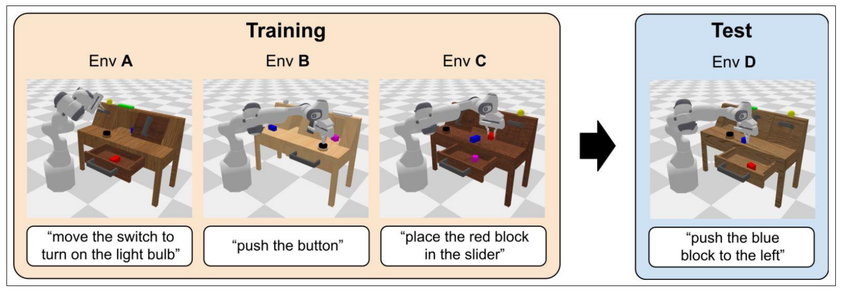
Faster inference time through hierarchy

	HULC	SPIL	Diffuser-1D	Diffuser-2D	Ours (HLP only)	Ours (full)
Training (hrs) (↓ Lower is Better)	82	86	20.8	49.2	**13.3**	95.3
Inference time (sec) (↓)	**0.005**	**0.005**	1.11	5.02	0.333	0.336
Avg ∇ updates/sec (↑)	.5	.5	4	2.1	**6.25**	**6.25**
Model size (↓)	47.1M	54.8M	74.7M	125.5M	**20.1M**	67.8M
Latent dims (↓)	N/A	N/A	256	1024	**32**	**32**

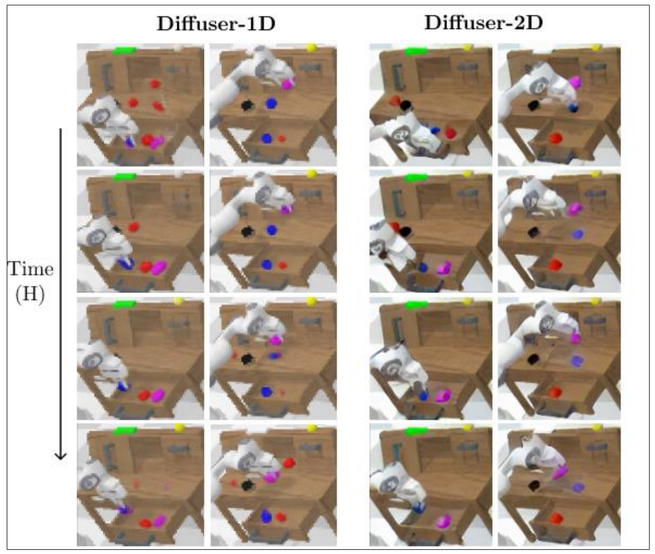
Diffusion vs Other Methods

Task	MLP	Transformer	Ours
One	86.6 ± 2.3	85.4 ± 0.5	$\mathbf{88.7 \pm 1.5}$
Two	64.1 ± 0.1	60.9 ± 0.5	$\mathbf{69.9 \pm 2.8}$
Three	48.1 ± 8.3	42.6 ± 1.9	$\mathbf{54.5 \pm 5.0}$
Four	35.7 ± 8.8	29.8 ± 2.5	$\mathbf{42.7 \pm 5.2}$
Five	25.5 ± 7.6	20.0 ± 1.8	$\mathbf{32.2 \pm 5.2}$
Avg SR	2.60 ± 0.33	2.36 ± 0.08	$\mathbf{2.88 \pm 0.19}$

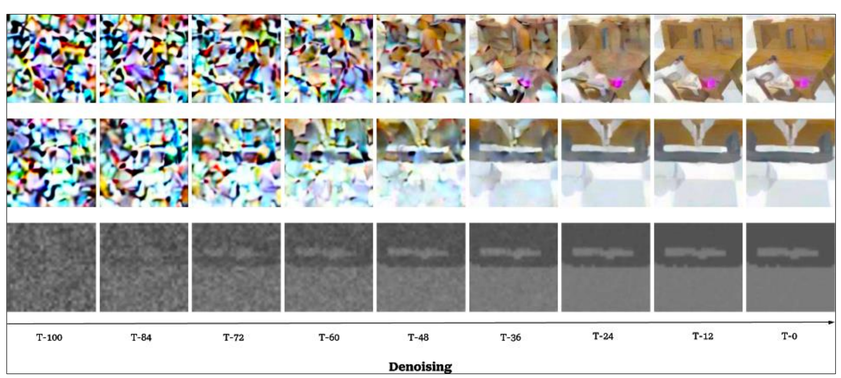




CALVIN benchmark example. Trained on long trajectories of a series of actions, the model will be tested on a new, never seen before trajectory.



Denoised Latent Representations. Directly using latent diffusion models fails. Hallucination occurs on Diffusor-1D, and loss of fine details occurs on Diffuser-2D.



An overview of our Denoising process. This an example of the denoising process of one of our ablations, the Diffuser-2D model.