DYffusion: A Dynamics-Informed Diffusion Model for Spatiotemporal Forecasting

Salva Rühling Cachay, Bo Zhao, Hailey Joren, Rose Yu

UC San Diego

NeurIPS 2023

arxiv.org/abs/2306.01984

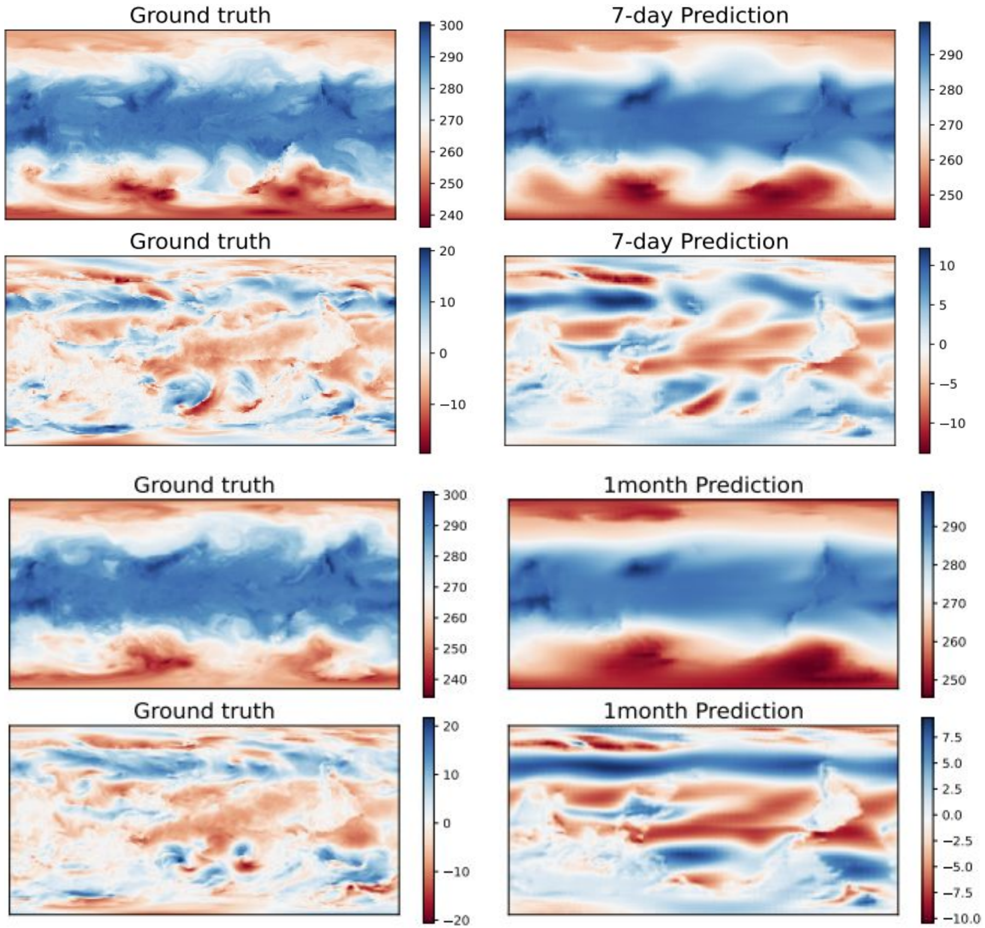
github.com/Rose-STL-Lab/dyffusion

NEURAL INFORMATION PROCESSING SYSTEMS

Motivation

Existing ML models for high-dimensional spatiotemporal forecasting tend to be:

- Deterministic → blurry, unrealistic long-range forecasts



ClimaX: A foundation model for weather and climate,
Nguyen et al., ICML 23, https://arxiv.org/abs/2301.10343

Motivation

Existing ML models for high-dimensional spatiotemporal forecasting tend to be:

- Deterministic → blurry, unrealistic long-range forecasts
- Autoregressive → Inference differs from training, errors accumulate, and rollouts become long-term unstable

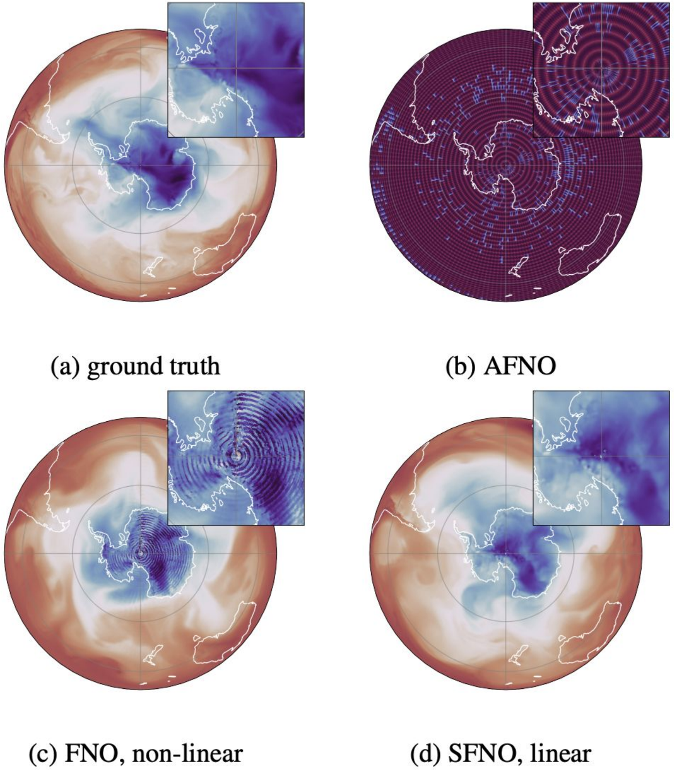
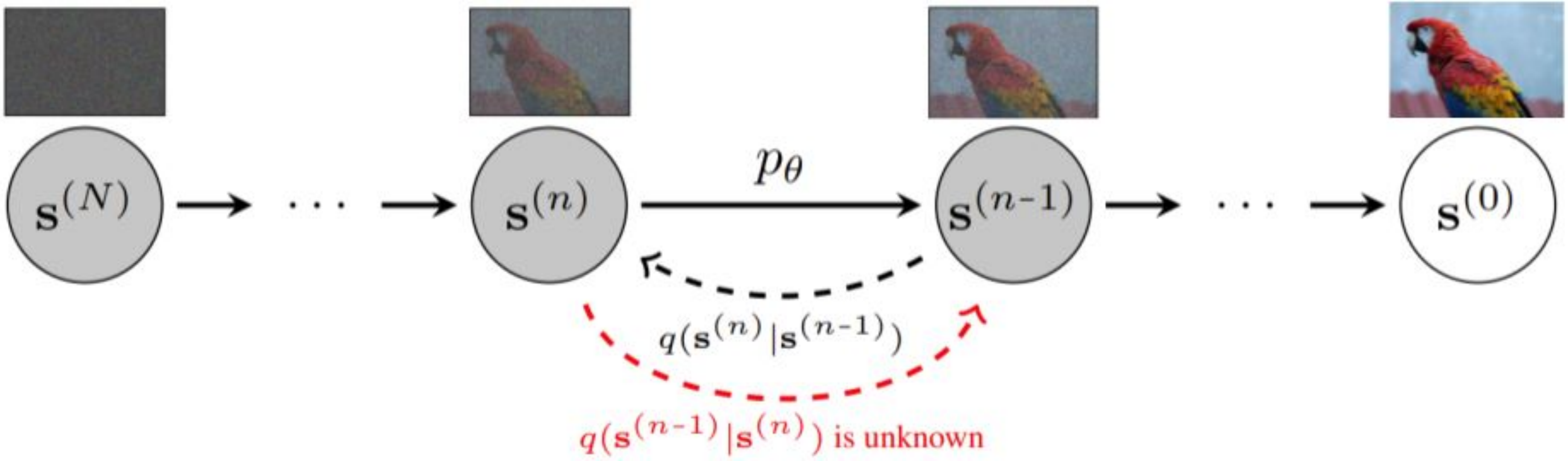


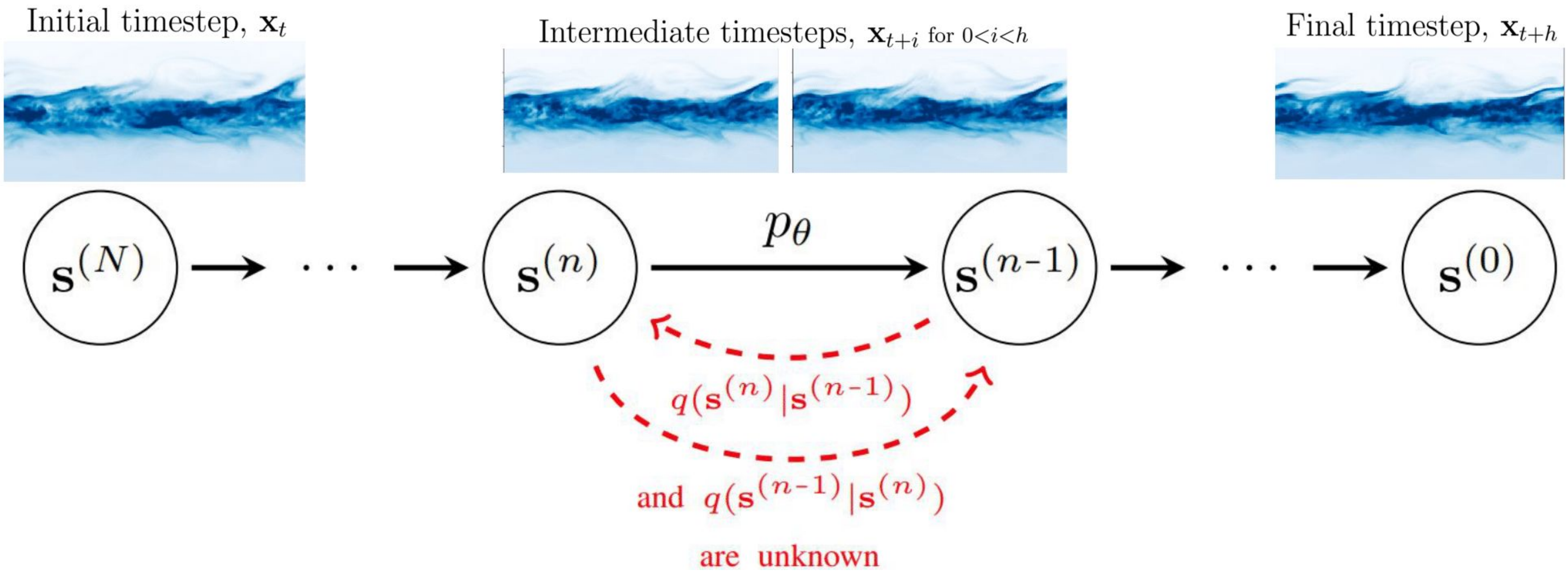
Figure 1. Qualitative comparison of temperature predictions (`t850`) over Antarctica at 4380h (730 autoregressive steps). The SFNO shows no visible artifacts even after six-month-long rollouts.

Spherical Fourier Neural Operators: Learning stable dynamics on the sphere, Bonev et al., ICML 23, https://arxiv.org/abs/2306.03838

Motivation: **Diffusion models are mostly designed for *static* data**



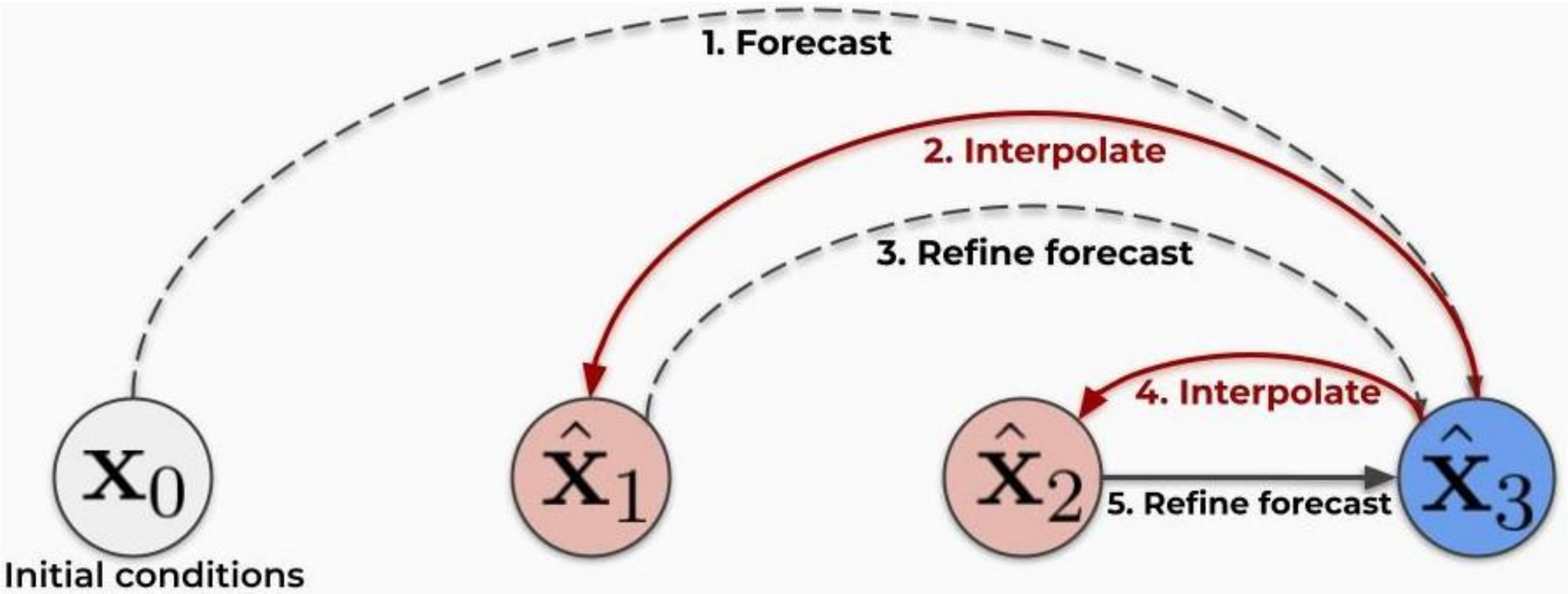
Key idea: Replace the forward & reverse processes with temporal interpolation & forecasting



Forecasting with DYffusion at inference time

DYffusion forecasts a sequence of h snapshots $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_h$ given the initial conditions $\mathbf{x}_0$ similarly to how standard diffusion models are used to sample from a distribution.

Forecasting with DYffusion at inference time



DYffusion forecasts a sequence of h snapshots $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_h$ given the initial conditions $\mathbf{x}_0$ similarly to how standard diffusion models are used to sample from a distribution.

Methods: **Training**

- Standard forward process $\rightarrow$ a stochastic temporal interpolation net, $\mathcal{I}_\phi$
- Standard reverse process $\rightarrow$ a deterministic forecaster network, F_θ, that predicts h steps ahead
- Train networks in two stages with simple time-conditioned objectives
- In the second stage and during sampling, use a schedule that maps diffusion steps to interpolation timesteps. In the simplest case $[i_n]_{i=0}^{N-1} = \{0, 1, \ldots, h-1\}$

Algorithm 1 DYffusion, Two-stage Training

Input: networks $F_\theta, \mathcal{I}_\phi$, norm $||\cdot||$, horizon h, schedule $[i_n]_{i=0}^{N-1}$

Stage 1: Train interpolator network, $\mathcal{I}_\phi$

1. Sample $i \sim \texttt{Uniform}(\{1, \ldots, h-1\})$
2. Sample $\mathbf{x}_t, \mathbf{x}_{t+i}, \mathbf{x}_{t+h} \sim \mathcal{X}$
3. Optimize $\min_\phi ||\mathcal{I}_\phi(\mathbf{x}_t, \mathbf{x}_{t+h}, i) - \mathbf{x}_{t+i}||^2$

Stage 2: Train forecaster network (diffusion model backbone), F_θ

1. Freeze $\mathcal{I}_\phi$ and enable inference stochasticity (e.g. dropout)
2. Sample $n \sim \texttt{Uniform}(\{0, \ldots, N-1\})$ and $\mathbf{x}_t, \mathbf{x}_{t+h} \sim \mathcal{X}$
3. Optimize $\min_\theta ||F_\theta(\mathcal{I}_\phi(\mathbf{x}_t, \mathbf{x}_{t+h}, i_n), i_n) - \mathbf{x}_{t+h}||^2$

Methods: **Sampling**

- DYffusion models the dynamics $\mathbf{x}(s)$ as follows, given initial conditions $\mathbf{x}(t) = \mathbf{x}_t$:

$$\mathbf{x}(s) = \mathbf{x}(t) + \int_t^s \frac{d\mathcal{I}_\phi\left(\mathbf{x}_t, F_\theta(\mathbf{x}, s), s\right)}{ds} ds \qquad \text{for } s \in (t, t+h].$$

- At inference time, we evaluate the integral using cold sampling [1].

Proposition 1. *Cold Sampling is an approximation of the Euler method.*

Proposition 2. *In Cold Sampling, the discretization error per step is bounded by $O(\Delta s)$. Naive sampling does not have this property.*

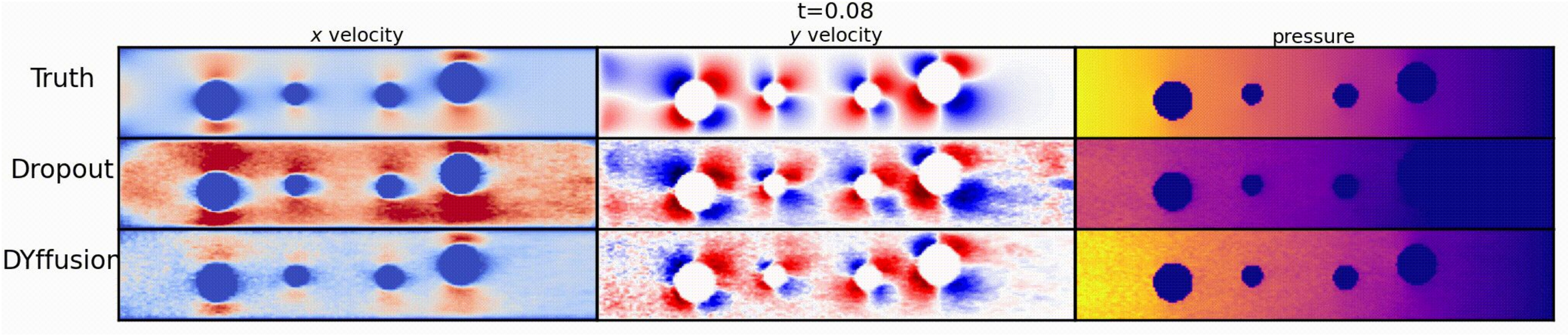
- Different discretizations are allowed: flexible sampling schedules at inference time

Algorithm 2 Adapted Cold Sampling [2] for DYffusion

1: **Input:** Initial conditions $\hat{\mathbf{x}}_t := \mathbf{x}_t$, schedule $[i_n]_{i=0}^{N-1}$, output timesteps J (by default $J = \{1, \ldots, h-1\}$)
2: **for** $n = 0, 1, \ldots, N-1$ **do**
3: $\quad \hat{\mathbf{x}}_{t+h} \leftarrow F_\theta(\hat{\mathbf{x}}_{t+i_n}, i_n)$
4: $\quad \hat{\mathbf{x}}_{t+i_{n+1}} = \mathcal{I}_\phi\left(\mathbf{x}_t, \hat{\mathbf{x}}_{t+h}, i_{n+1}\right) - \mathcal{I}_\phi\left(\mathbf{x}_t, \hat{\mathbf{x}}_{t+h}, i_n\right) + \hat{\mathbf{x}}_{t+i_n}$
5: **end for**
6: $\hat{\mathbf{x}}_{t+j} \leftarrow \mathcal{I}_\phi\left(\mathbf{x}_t, \hat{\mathbf{x}}_{t+h}, j\right), \ \forall j \in J$ # Optional refinement
7: **Return:** $\{\hat{\mathbf{x}}_{t+j} \mid j \in J\} \cup \{\hat{\mathbf{x}}_{t+h}\}$

A. Bansal, E. Borgnia, H.-M. Chu, J. S. Li, H. Kazemi, F. Huang, M. Goldblum, J. Geiping, and T. Goldstein. Cold diffusion: Inverting arbitrary image transforms without noise. *Advances in Neural Information Processing Systems*, 2023.

Results: **Competitive probabilistic rollouts**



Main benchmark results. Evaluation with 50-member ensembles for sea surface temperature forecasting of 1 to 7 days ahead, and Navier-Stokes flow full trajectory forecasting of 64 timesteps. Numbers are averaged out over the evaluation horizon. **Bold** indicates best, blue second best. Lower is better for CRPS and MSE; Closer to 1 is better for SSR.

Method	**SST**				**Navier-Stokes**		
	CRPS	MSE	SSR	Time [s]	CRPS	MSE	SSR
Perturbation	0.281 ± 0.004	0.180 ± 0.011	0.411 ± 0.046	0.4241	0.090 ± 0.001	0.028 ± 0.000	0.448 ± 0.002
Dropout	0.267 ± 0.003	0.164 ± 0.004	0.406 ± 0.042	0.4241	0.078 ± 0.001	0.027 ± 0.001	0.715 ± 0.005
DDPM	0.246 ± 0.005	0.177 ± 0.005	0.674 ± 0.011	0.3054	0.180 ± 0.004	0.105 ± 0.010	0.573 ± 0.001
MCVD	**0.216**	**0.161**	0.926	79.167	0.154 ± 0.043	0.070 ± 0.033	0.524 ± 0.064
DYffusion	0.224 ± 0.001	0.173 ± 0.001	**1.033 ± 0.005**	4.6722	**0.067 ± 0.003**	**0.022 ± 0.002**	**0.877 ± 0.006**

Results: **Temporal super-resolution (8x)**

Summary

- First study on diffusion models for spatiotemporal forecasting
- Novel adaptation of diffusion models to ensemble-based probabilistic forecasting
- Effective training approach for multi-step and long-range forecasting with low memory needs
- Competitive performance on probabilistic evaluations for forecasting complex dynamics in sea surface temperatures, Navier-Stokes flows, and spring mesh systems

+ *more details, results and extensive ablations in our paper!*

Thanks for listening!
Feel free to reach out :)

sruhlingcachay@ucsd.edu

@salvaRC7

Blog: https://salvarc.github.io/blog/2023/dyffusion

https://github.com/Rose-STL-Lab/dyffusion

Paper: arxiv.org/abs/2306.01984