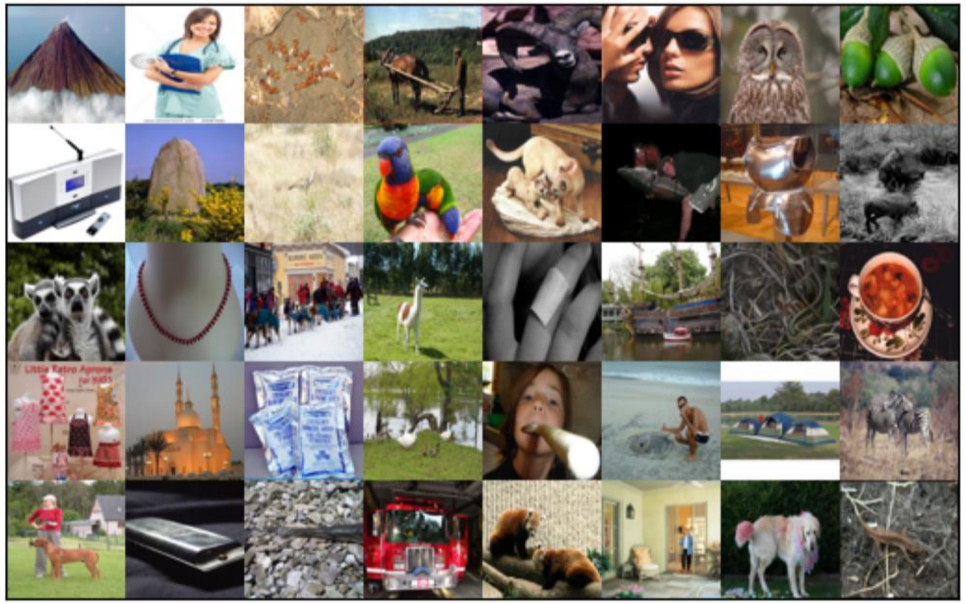
Value Gradient Sampling

KIAS CAINS Fall Workshop

2024-11-08

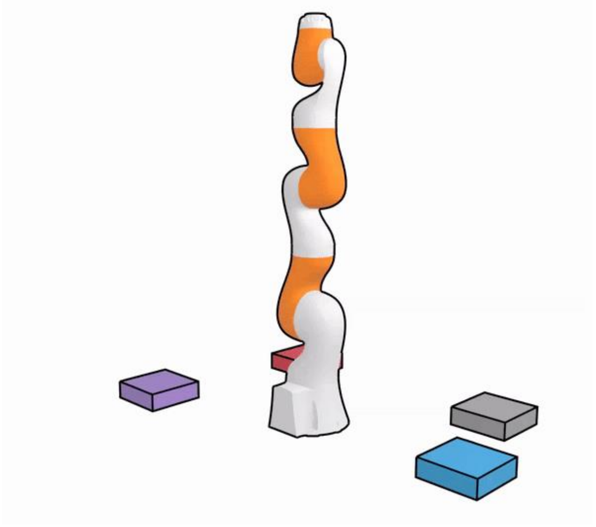
Sangwoong Yoon

Generative Modeling



$$\min_{\color{orange}{\pi \in \Pi}} D(\color{orange}{\pi}, p)$$

Reinforcement Learning



$$\max_{\color{orange}{\pi \in \Pi}} \mathbb{E}_{\color{orange}{\tau \sim \pi}}[R(\tau)]$$

https://anuragajay.github.io/decision-diffuser/

Problem: Sampling from Unnormalized Density

$$q(x) = \frac{1}{Z} e^{-E(x)/\tau}$$

- Energy $E(x)$
- Temperature τ
- Normalization constant Z

Task: Given $E(x)$ (and τ), draw $x \sim q(x)$

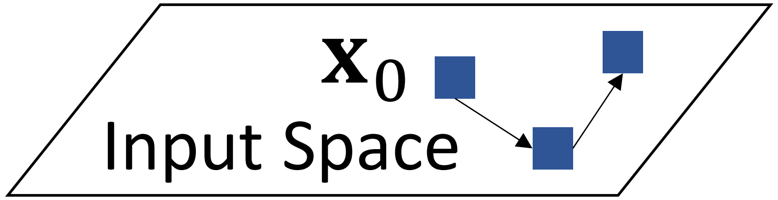
Markov Chain Monte Carlo (MCMC)

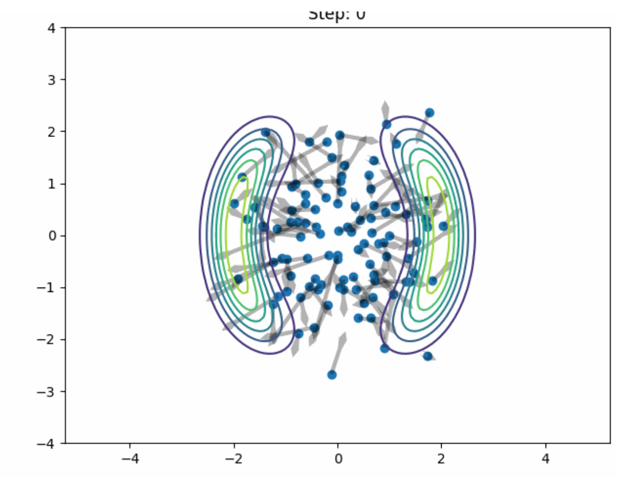
Langevin Monte Carlo (LMC) Sampling

$t = 0, \cdots, T$

$\mathbf{x}_0 \sim$ Noise distribution

$$\mathbf{x}_{t+1} = \mathbf{x}_t - \frac{\lambda_1}{2} \nabla_{\mathbf{x}} E(\mathbf{x}_t) + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \lambda_2)$$





- $\mathbf{x}_T$: A sample from $q_\theta(\mathbf{x})$
- λ_1, λ_2: Step size and noise strength.

Parametric Sampler

$$\min_{\phi} KL\left(\pi_{\phi} || q\right)$$

- Faster (at the expense of training time)
- Adaptive – Parameter ϕ is optimized for given $q(x)$
- Re-usable

Sampling is MaxEnt RL

$$\min_{\pi} KL(\pi(x)||q(x))$$
$$= \min_{\pi} \mathbb{E}_{\pi}[-\log q(x)] - \mathcal{H}(\pi(x))$$

- Policy is $\pi(x)$
- Reward is $\log q(x)$ (cost is $-\log q(x) = E(x)$)
- MaxEnt regularization: $-\mathcal{H}(\pi(x))$

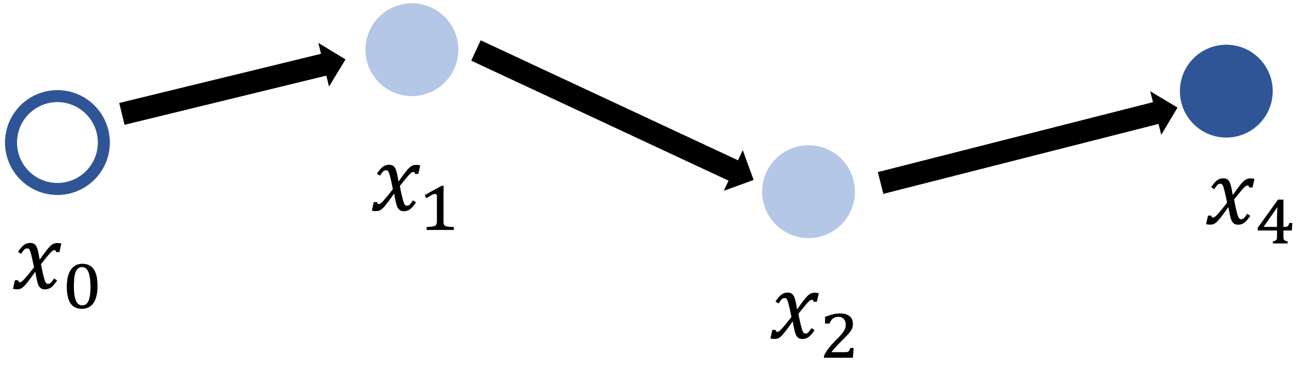
$x_t \in \mathbb{R}^D$

Langevin-like Parametric Sampler

$$x_0 \sim \mathcal{N}(0, I)$$
$$x_{t+1} = a_t x_t + f_\theta(x_t, t) + \sigma_t \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, I)$$

$$\pi(x_{t+1} | x_t) = \mathcal{N}(\mu_t = a_t x_t + f_\theta(x_t, t), \sigma_t^2 I)$$

The last step x_T $(= x)$ is taken as a sample.



A sequential decision making problem

$x_t \in \mathbb{R}^D$

Example: Diffusion Model

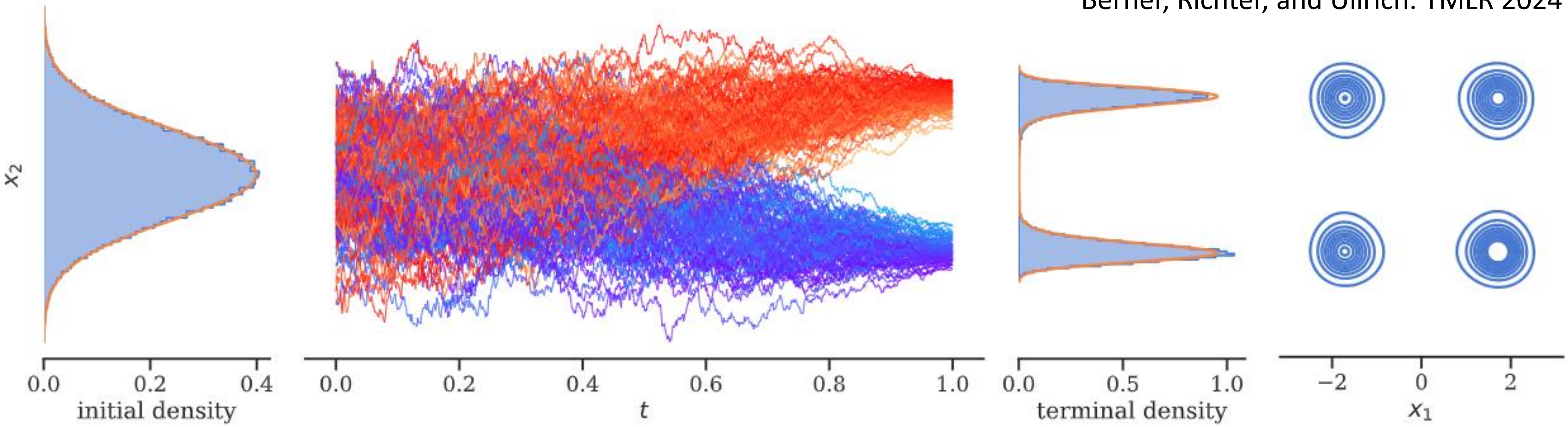
$$x_0 \sim \mathcal{N}(0, I)$$
$$x_{t+1} = x_t + f_\theta(x_t, t) + \sigma_t \epsilon_t, \qquad \epsilon_t \sim \mathcal{N}(0, I)$$

$$f_\theta(x_t, t) = \sigma_t^2 \nabla_x \log p_t(x)$$

- $p_t(x)$: Diffused data distribution
- We don't have access to data points – only energy.

Example: SDE-Based Samplers

Zhang and Chen. ICLR 2022
Berner, Richter, and Ullrich. TMLR 2024



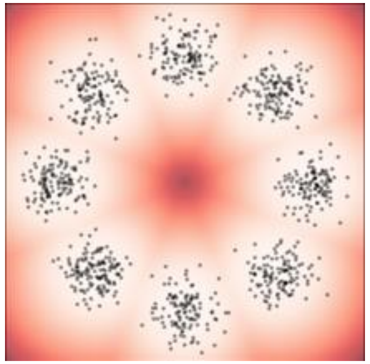
- Formulated in continuous time.
- Requires fine-grained simulation.

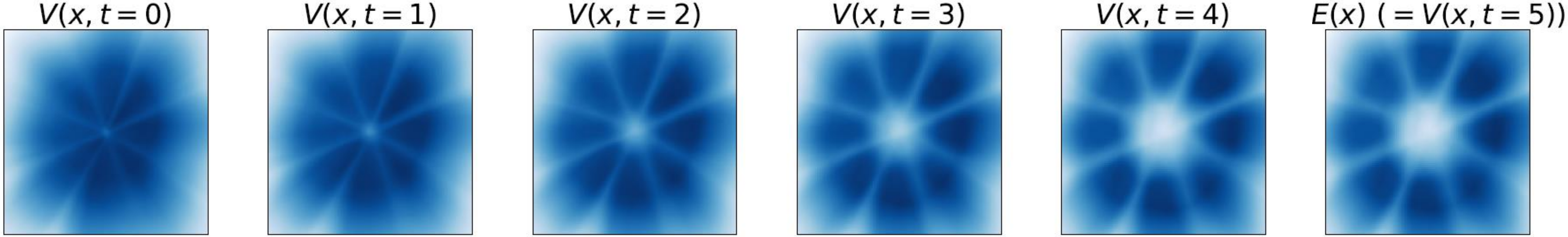
Proposal: Value Gradient Sampling

$$x_0 \sim \mathcal{N}(0, I)$$

$$x_{t+1} = a_t x_t - \frac{\sigma_t^2}{\tau} \nabla_{x_{t+1}} V_{t+1}\left(x_{t+1}\right) + \sigma_t \epsilon_t$$

Target Density





Value Function $V_t(x)$

"How good is the current state? "

Value Function $V_t(x)$

Expected reward/cost from the current state x_t

$$V(x_t) = \mathbb{E}_{\pi(x_{t+1:T}|x_t)}\left[\sum_{t'=t+1}^{T} R(x_{t'})\right]$$

Optimal action:

$$\pi^*(x_{t+1}|x_t) = \operatorname{argmax}_{\pi} \mathbb{E}_{\pi(x_{t+1}|x_t)}[V(x_{t+1})]$$

Difficulty

$$\min_{\pi} KL(\pi(x)||q(x))$$
$$= \min_{\pi} \mathbb{E}_{\pi}[-\log q(x) + \log \pi(x)]$$

1. We can not evaluate $\log \pi(x)$ analytically.
2. Backprop through time is difficult.

Data Processing Inequality

$$KL(P(X)||Q(X)) \leq KL(P(X, Y)||Q(X, Y))$$

$$\min_{\pi} KL(\pi(x_T)||q(x_T))$$

Auxiliary distribution

$$KL\big(\pi(x_T)||q(x_T)\big) \leq \underline{KL(\pi(x_{0:T})||q(x_T)\tilde{q}(x_{0:T-1}|x_T))}$$

Minimize THIS

Choice of Auxiliary Distribution $\tilde{q}(x_{0:T-1}|x_T)$

Let each $\tilde{q}(x_t|x_{t+1})$ be Gaussian:

$$\tilde{q}(x_{0:T-1}|x_T) = \prod_{t=0}^{T-1} \tilde{q}(x_t|x_{t+1})$$

$$\tilde{q}(x_t|x_{t+1}) = \mathcal{N}(x_{t+1}, s_t^2 I)$$

Optimal Control Formulation

$$\min_{\pi} KL(\pi(x_{0:T})||q(x_T)\tilde{q}(x_{0:T-1}|x_T))$$

becomes an optimal control problem:

$$\min_{\pi} \mathbb{E}_{\pi}\left[\underbrace{E(x_T)}_{\text{Terminal Cost}} + \tau \underbrace{\sum_{t=0}^{T-1} \log \pi(x_{t+1}|x_t) + \sum_{t=0}^{T-1} \frac{\tau}{2s_t^2} \|x_{t+1} - x_t\|^2}_{\text{Running Cost}}\right]$$

Value Function $V_t(x_t)$

Expected future return at (x_t, t)

$$V_t(x_t) = \min_{\pi} \mathbb{E}_{\pi}\left[E(x_T) + \overbrace{\tau \sum_{t'=t}^{T-1} \log \pi(x_{t'+1}|x_{t'}) + \sum_{t'=t}^{T-1} \frac{\tau}{2s_{t'}^2} \|x_{t'+1} - x_{t'}\|^2}^{\text{Running Cost}} \right]$$

$V_t(x_t)$ is parametrized as a neural network.

Value Function Learning

Temporal Difference Learning

$$\min_{V_t} \mathbb{E}_{x_t, x_{t+1} \sim \pi} \left[\left(V_{t+1}(x_{t+1}) + \tau R(x_t, x_{t+1}) - V_t(x_t) \right)^2 \right]$$

Running cost

$$R(x_t, x_{t+1}) = \log \pi(x_{t'+1} | x_{t'}) + \frac{1}{2s_{t'}^2} \|x_{t'+1} - x_{t'}\|^2$$

s_t: std of aux dist

(Approximate) Optimal Policy

$$\pi(x_{t+1}|x_t) = \mathcal{N}(x_t + \mu_t, \sigma_t^2 I)$$

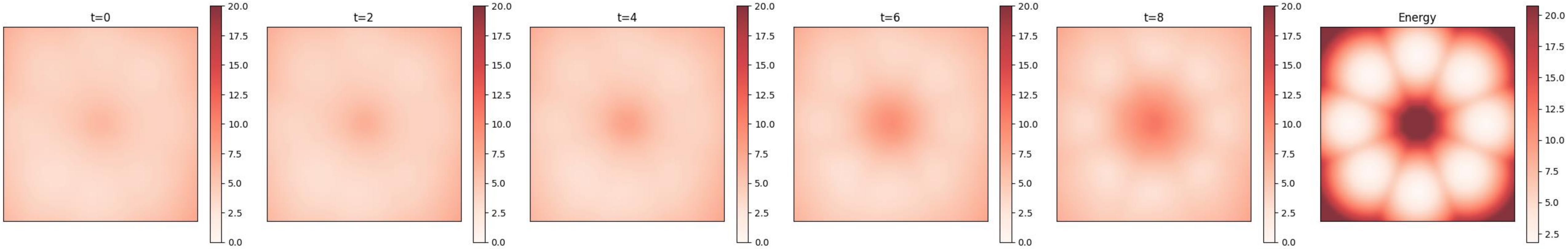
$$\min_{\pi(x_{t+1}|x_t)} \mathbb{E}_{x_t, x_{t+1} \sim \pi}\left[V_{t+1}(x_{t+1}) + R(x_t, x_{t+1})\right]$$

(Under some assumptions)

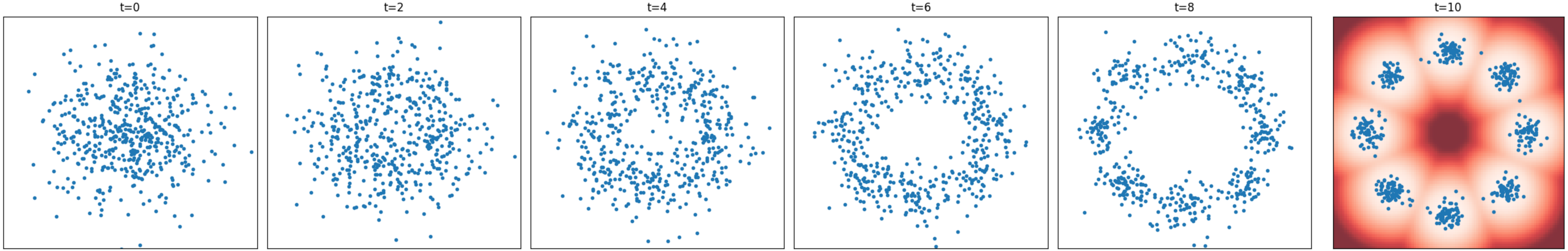
$$\mu_t = -\frac{s_t^2}{\tau} \nabla_{x_{t+1}} V_{t+1}(x_{t+1})$$

$$\sigma_t^2 = s_t^2$$

Value Functions



Samples



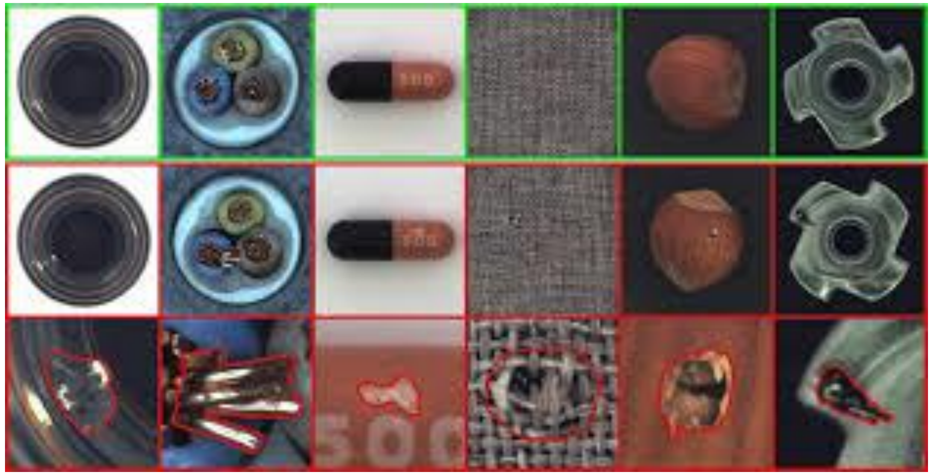
Application: Training Energy-Based Model

Maximum Likelihood Training of EBM

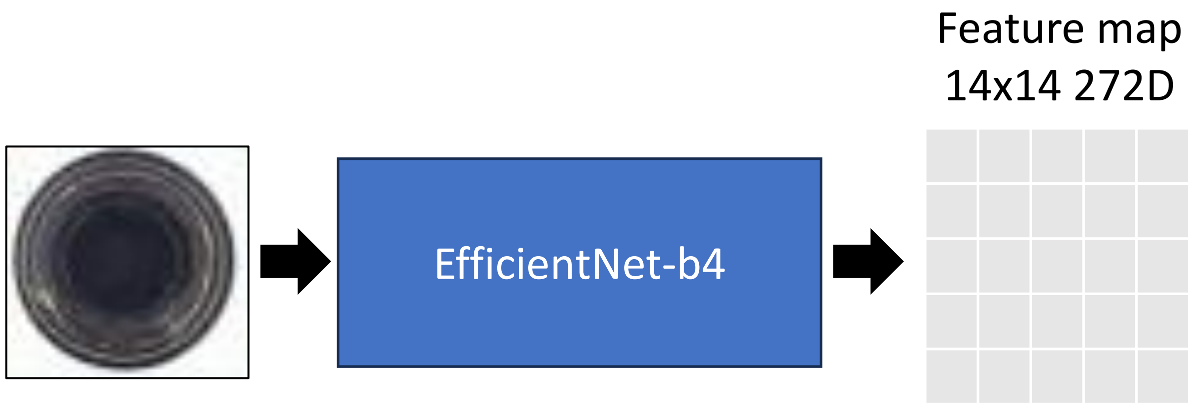
$$\nabla_\theta \log p_\theta(\mathbf{x}) = -\nabla_\theta E_\theta(\mathbf{x}) + \boxed{\mathbb{E}_{\mathbf{x}^- \sim p_\theta(\mathbf{x})}}[\nabla_\theta E_\theta(\mathbf{x}^-)]$$

"Negative" samples from EBM

Anomaly Detection



Model	DET	LOC
DRAEM [56]	88.1	87.2
MPDR [57]	96.0	96.7
UniAD [58]	96.5±0.08	96.8±0.02
EBM-VGS	**97.0** ±0.11	**97.1**±0.02



- $E(x)$ is used as anomaly score

Maximum Entropy Inverse Reinforcement Learning of Diffusion Models with Energy-Based Models

NeurIPS 2024 Oral Presentation





Sangwoong Yoon
KIAS

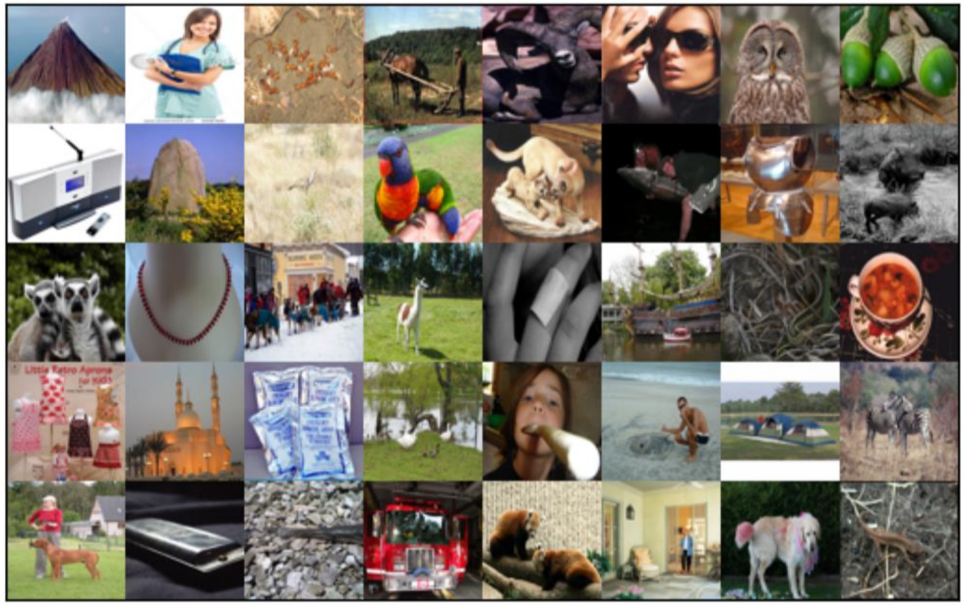
Himchan Hwang
Seoul National Univ.

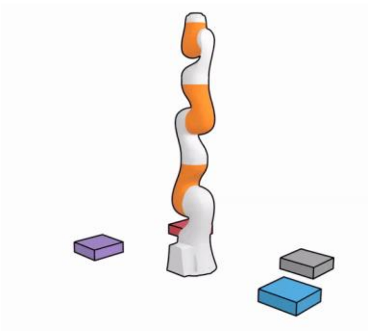
Dohyun Kwon
Univ. of Seoul
KIAS

Yung-Kyun Noh
Hanyang Univ.
KIAS

Frank C. Park
Seoul National Univ.
Saige Research

Generative Modeling is Imitation Learning





Generative Modeling is Imitation Learning

Data Generating Process ↔ Expert

Data ↔ Demonstration

Generation ↔ Action

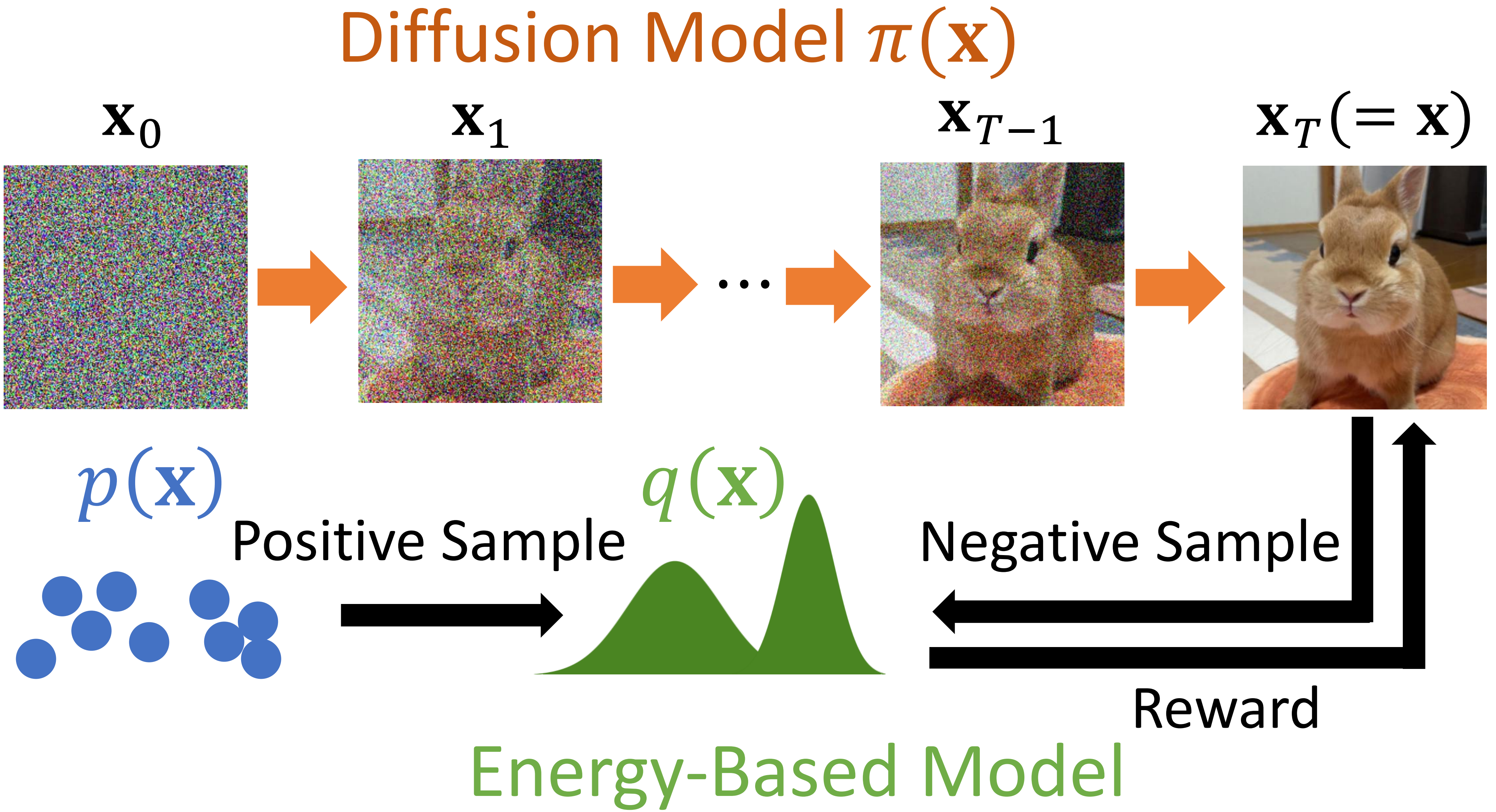
$\log p(x)$ **↔ Reward**

Fine-tuning Diffusion Models for Small T

Sample quality deteriorates significantly when fewer steps are used.

Diffusion model
trained on T=1000 &
generating on T=10

Diffusion by Maximum Entropy IRL (DxMI)



$\pi(x)$: Diffusion Model
$p(\mathbf{x})$: Data, $q(\mathbf{x})$: EBM

Generalized Contrastive Divergence

Inspired by Contrastive Divergence (Hinton, 2002)

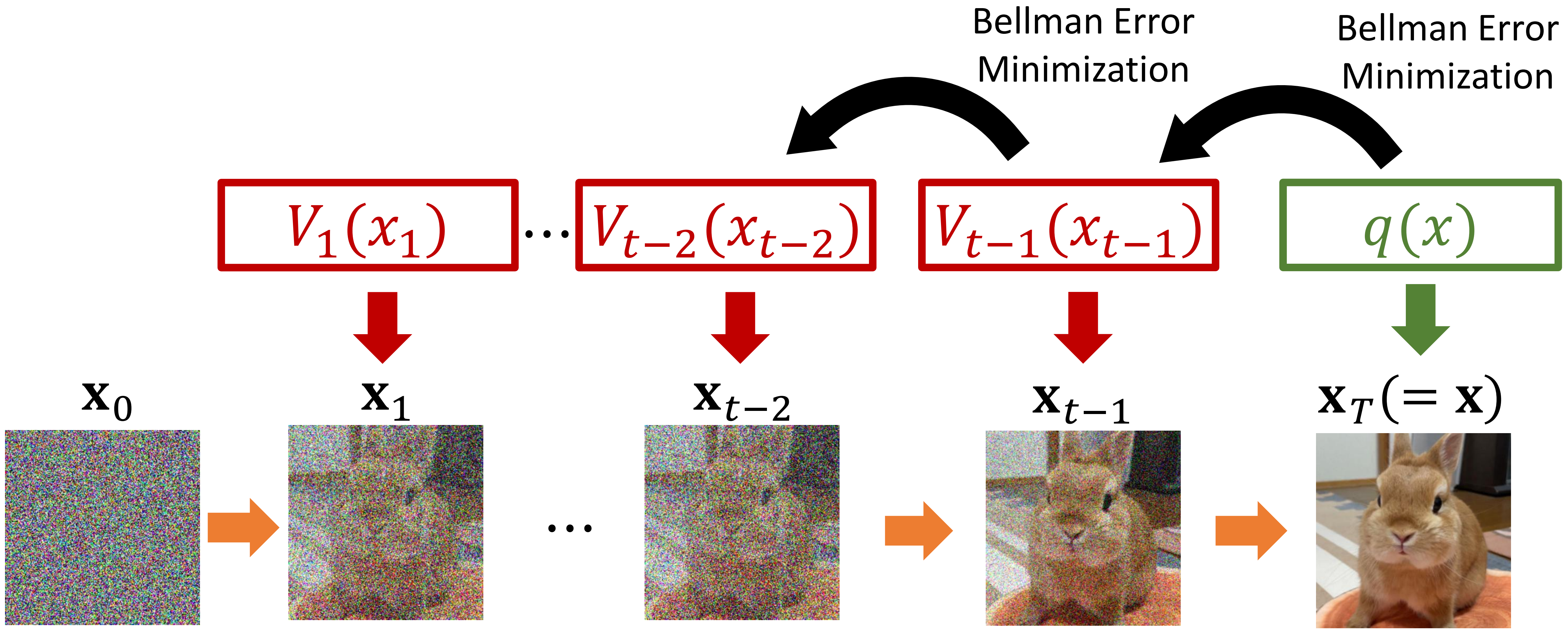
$$\min_{q} \max_{\pi} KL(p||q) - KL(\pi||q)$$

Normalization constant cancelled out

$$p(\mathbf{x}) \leftarrow q(\mathbf{x}) \leftarrow \pi(\mathbf{x})$$

- Same equilibrium: $p(\mathbf{x}) = q(\mathbf{x}) = \pi(\mathbf{x})$
- Equivalent to a previously known objective function in EBM literature but we are the first to use a diffusion model as $\pi(\mathbf{x})$.

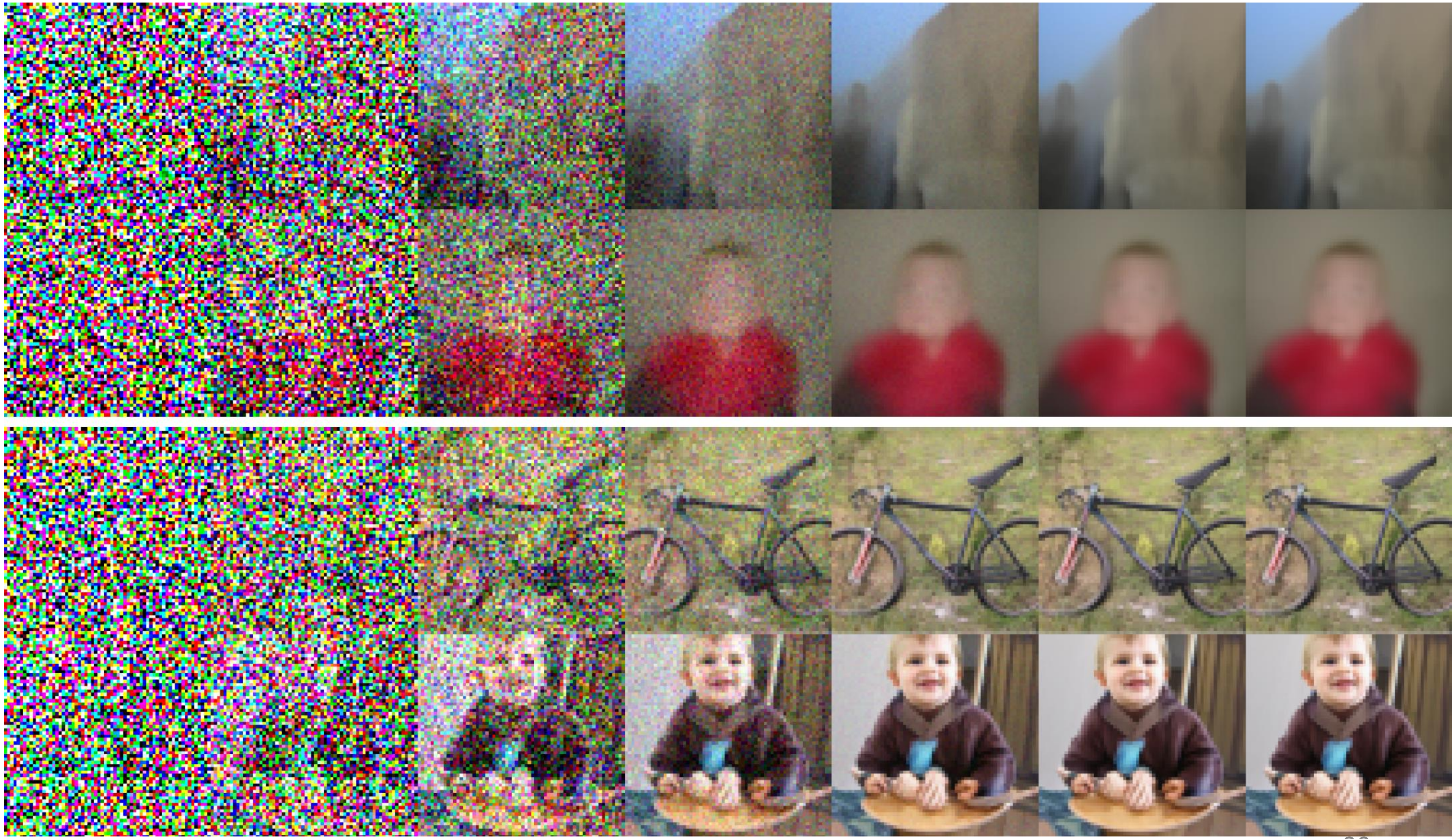
Diffusion by Dynamic Programming (DxDP)



Fine-tuning Diffusion Models for Small T

Diffusion model trained on T=1000 / generating on T=10

Fine-tuned T=10 with DxMI



ImageNet 64 Conditional Image Generation

	T	FID (↓)	Prec. (↑)	Rec. (↑)
EDM (Karras et al., 2022)	79	2.44	0.71	0.67
Consistency Model (Song et al., 2023)	2	4.70	0.69	0.64
	1	6.20	0.68	0.63
DxMI (Ours)	10	2.68	0.78	0.600
DxMI (Ours)	4	3.21	0.76	0.522

Thank you for listening!