Efficient Conditional Diffusion Sampling for Image Restoration and Editing

카이스트 AI 대학원 예종철 교수님 연구실 (BiSPL)
김정솔

Jeongsol Kim

Education

- KAIST (B.S / M.S / Ph.D.) – Bio and Brain Engineering
- Advisor - Jong Chul Ye (M.S, Ph.D.) and Mooseok Jang (Ph.D.)

Representative Researches

- 2023 ICLR **Spotlight** paper (Diffusion Posterior Sampling, DPS)
- 2024 ECCV paper (DreamSampler)
- 2025 ICLR **Spotlight** paper (Text regularization for latent inverse solvers, TReg)
- 2025 ICCV paper (FlowDPS)
- 2025 NeurIPS **Spotlight** paper (Chain-of-Zoom)

Awards

- Samsung humantech (2023: Gold / 2024: Bronze / 2025: Silver)

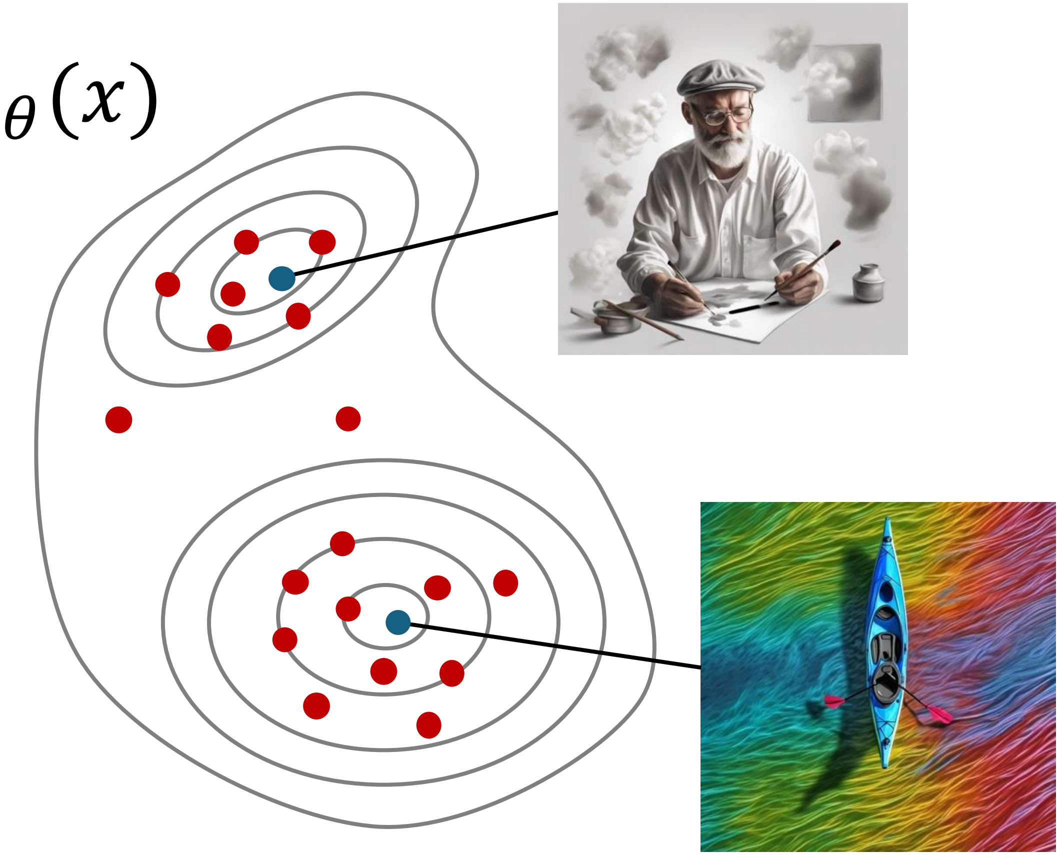
Previous Research Intern @ Snap Research, NYC

BiSPL

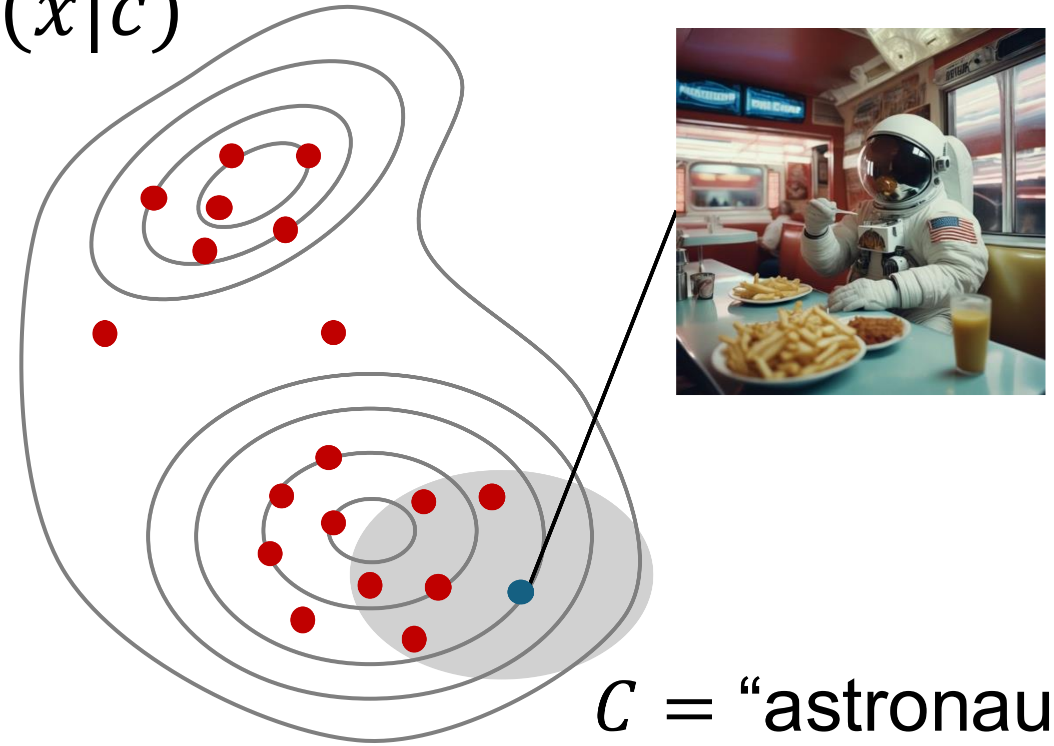
KAIST

Generation is sampling with controllability

$p_\theta(x)$



$p_\theta(x|c)$



C = "astronaut"

- Training samples
- Generated samples

Conditional sampling beyond generation

Conditional sampling

$$x \sim p_\theta(x|c)$$

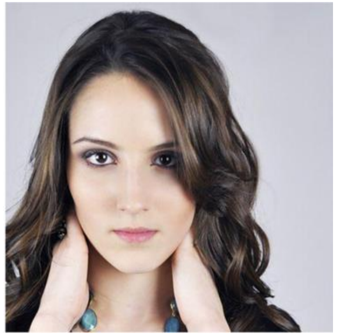
Degraded image as condition

: Image restoration

$$x \sim p(x|y)$$

Source image as condition

: Image editing



$$x \sim p(x|x_{src}, c)$$

Research Question

How to turn generative models into solvers

- that produce feasible solutions under real-world constraints.
- without re-training

Extend pre-trained generative models beyond generation

Powerful diffusion models
that generates images and videos

Flux

RunDiffusion

Juggernaut

stability.ai

Stable Diffusion

Retro Diffusion

Pixel Art

Google DeepMind

Veo

Alibaba Cloud

Wan

Image restoration

Image editing

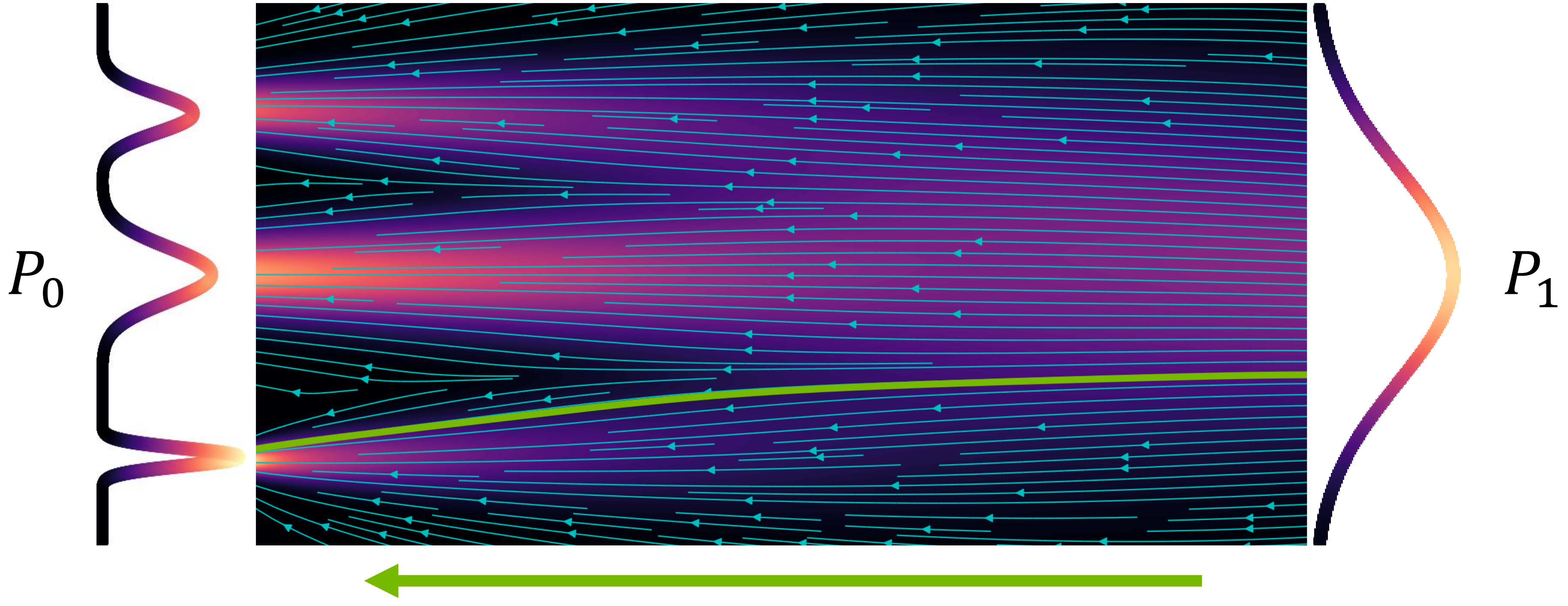
General approach that is compatible with any diffusion model: training-free methods

Scalability: Diffusion improves = Method improves

BiSPL

KAIST

Flow-based models and Flow ODE



Generation by solving Probability Flow ODE

$$dx_t = v(x_t, t)dt$$

$$v(x_t, t) = -t\nabla_{x_t} \log p(x_t) = \frac{x_t - \mathbb{E}[x_0|x_t]}{t}$$

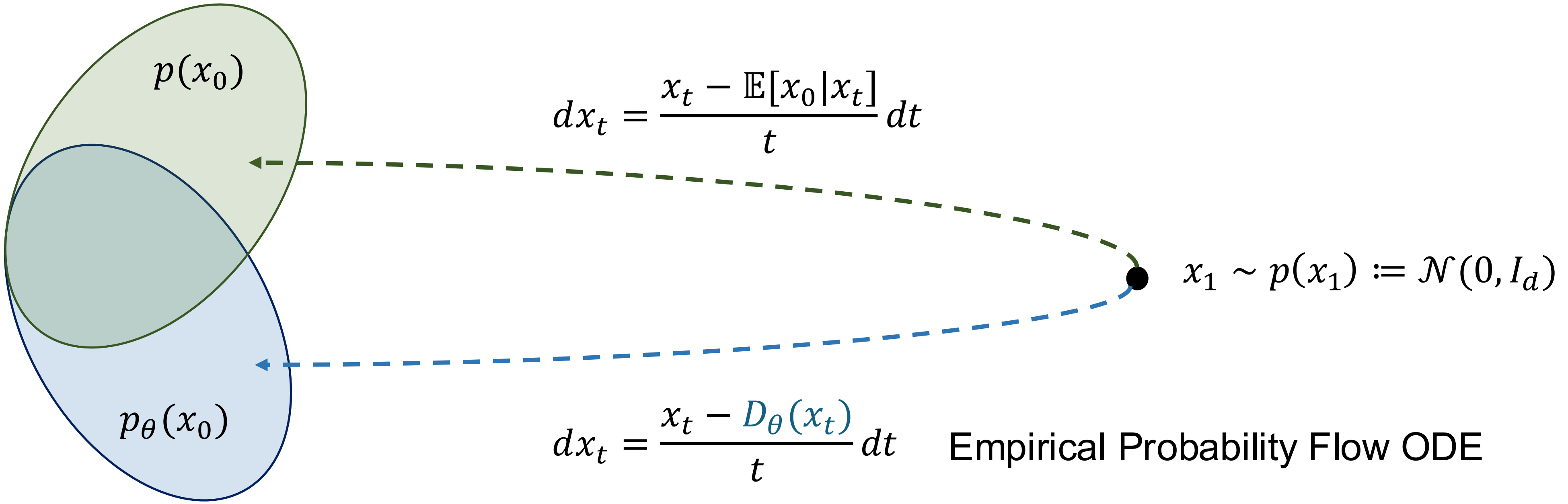
BiSPL

KAIST

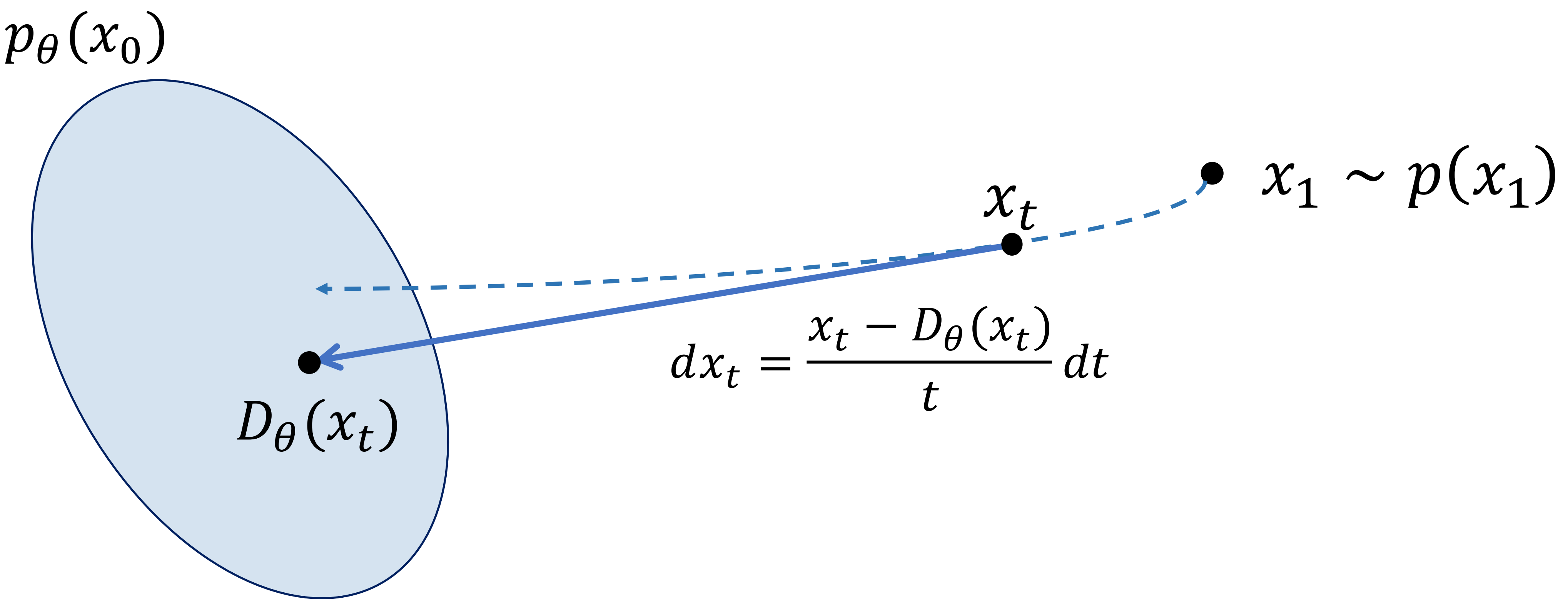
Empirical Flow ODE with parameterized denoiser

$$v(x_t, t) = -t\nabla_{x_t} \log p(x_t) = \frac{x_t - \mathbb{E}[x_0|x_t]}{t}$$

$$\mathbb{E}[x_0|x_t] = x_t + t^2\nabla_{x_t} \log p(x_t) \approx D_\theta(x_t)$$

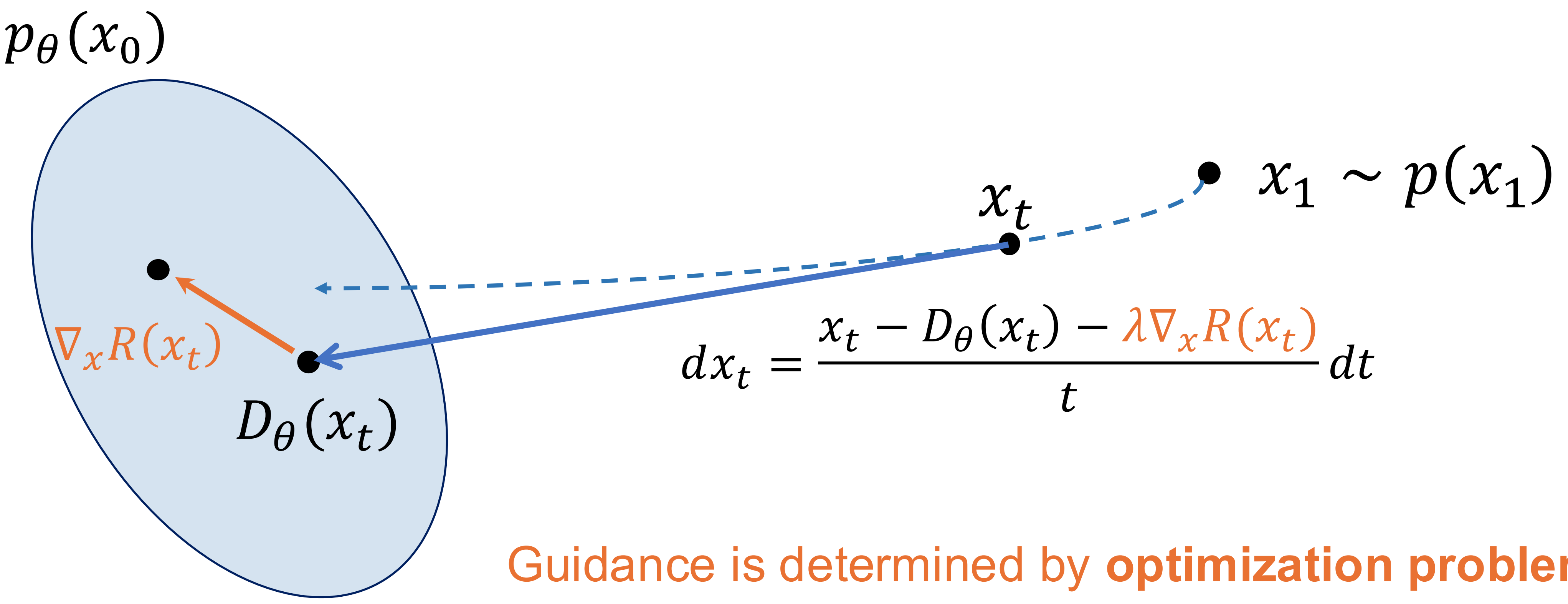


Flow ODE moves samples toward the denoised estimate



BiSPL KAIST

Principle – Guidance is applied to denoised estimates



$$dx_t = \frac{x_t - D_\theta(x_t) - \lambda \nabla_x R(x_t)}{t} dt$$

Guidance is determined by **optimization problem**

Content – How to define optimization problems

1. Proximal optimization : Image Editing with inversion

   - DreamSampler, *J.Kim*, G.Park*, & J.C.Ye, ECCV 2024*

2. Optimal control problem : Image Editing without inversion

   - FlowAlign*, J.Kim*, Y.Hong*, J.Park, & J.C.Ye, ICLR 2026*

3. Reward optimization: Inverse problems

   - DPS, *H.Chung*, J.Kim*, M.T.Mccann, M.L.Klasky, & J.C.Ye, ICLR 2023*
   - FlowDPS, *J.Kim*, B.S.Kim* & J.C.Ye, ICCV 2025*

Work 1: Proximal optimization

Image editing with inversion

DreamSampler, *J.Kim*, G.Park*, & J.C.Ye, ECCV 2024*

BiSPL

KAIST

Revisit Reverse Diffusion Sampling

At timestep t

$$\hat{\epsilon}_\theta = \epsilon_\theta(x_t, c_\phi, t) + \omega[\epsilon_\theta(x_t, c, t) - \epsilon_\theta(x_t, c_\phi, t)]$$

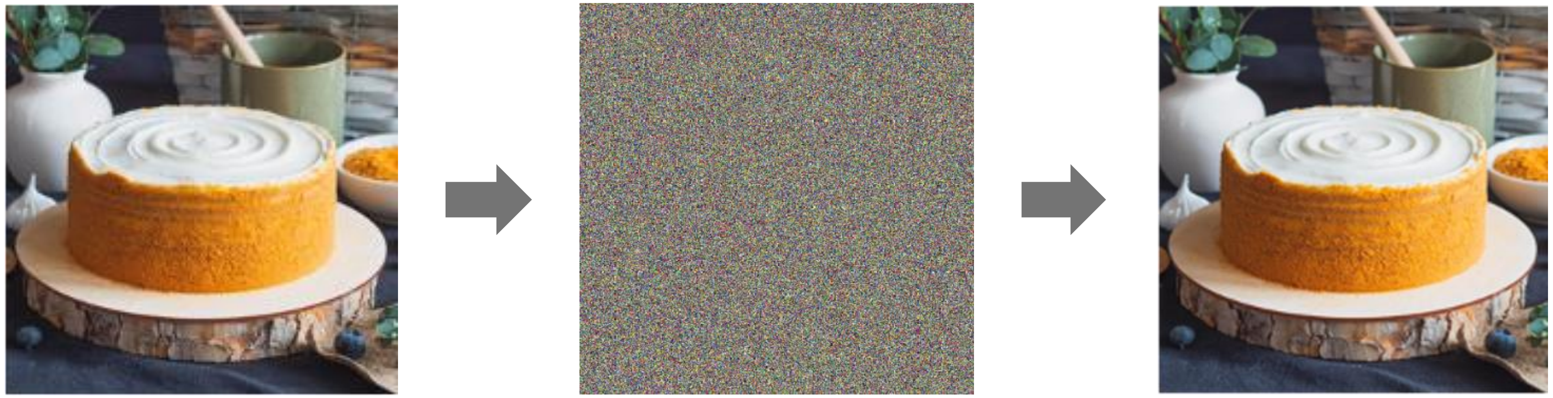
(Denoise) $\hat{x}_{0|t} = \left(x_t - \sqrt{1-\bar{\alpha}_t}\hat{\epsilon}_\theta\right)/\sqrt{\bar{\alpha}_t}$

$$\tilde{x} = \underset{x}{\mathrm{argmin}} \left\| x - \hat{x}_{0|t} \right\|_2^2 + \lambda \mathcal{R}(x)$$

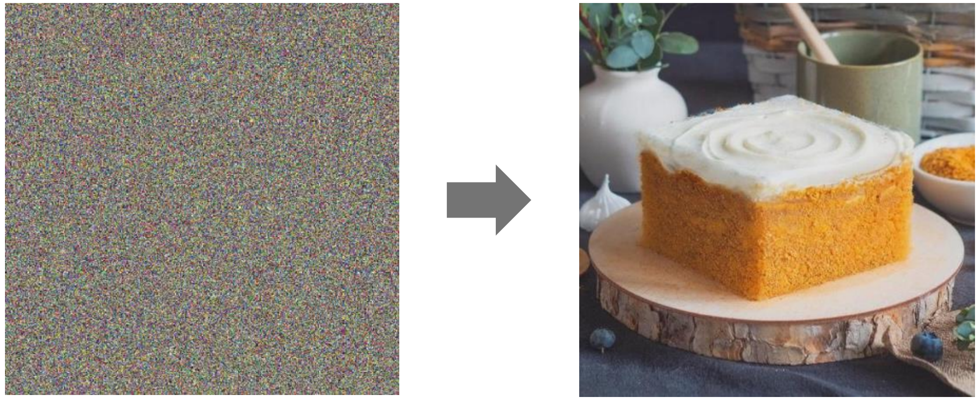
(Noise) $x_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\tilde{x} + \sqrt{1-\bar{\alpha}_{t-1}}\tilde{\epsilon}_t$

When $\lambda = 0$, it is a standard DDIM sampling.

Application 1. Image editing with inversion



Inversion & Reconstruction (unconditional)



"A rectangle cake"

Generation

DreamSampler – Text-driven Image Editing

$$\min_x \underbrace{\left\| x - \hat{x}_{0|t}(c_\emptyset) \right\|^2}_{\text{Proximal pivot}} + \frac{\gamma}{1-\gamma} \underbrace{\left\| x - \hat{x}_{0|t}(c_{tgt}) \right\|^2}_{\text{Text guidance}}$$

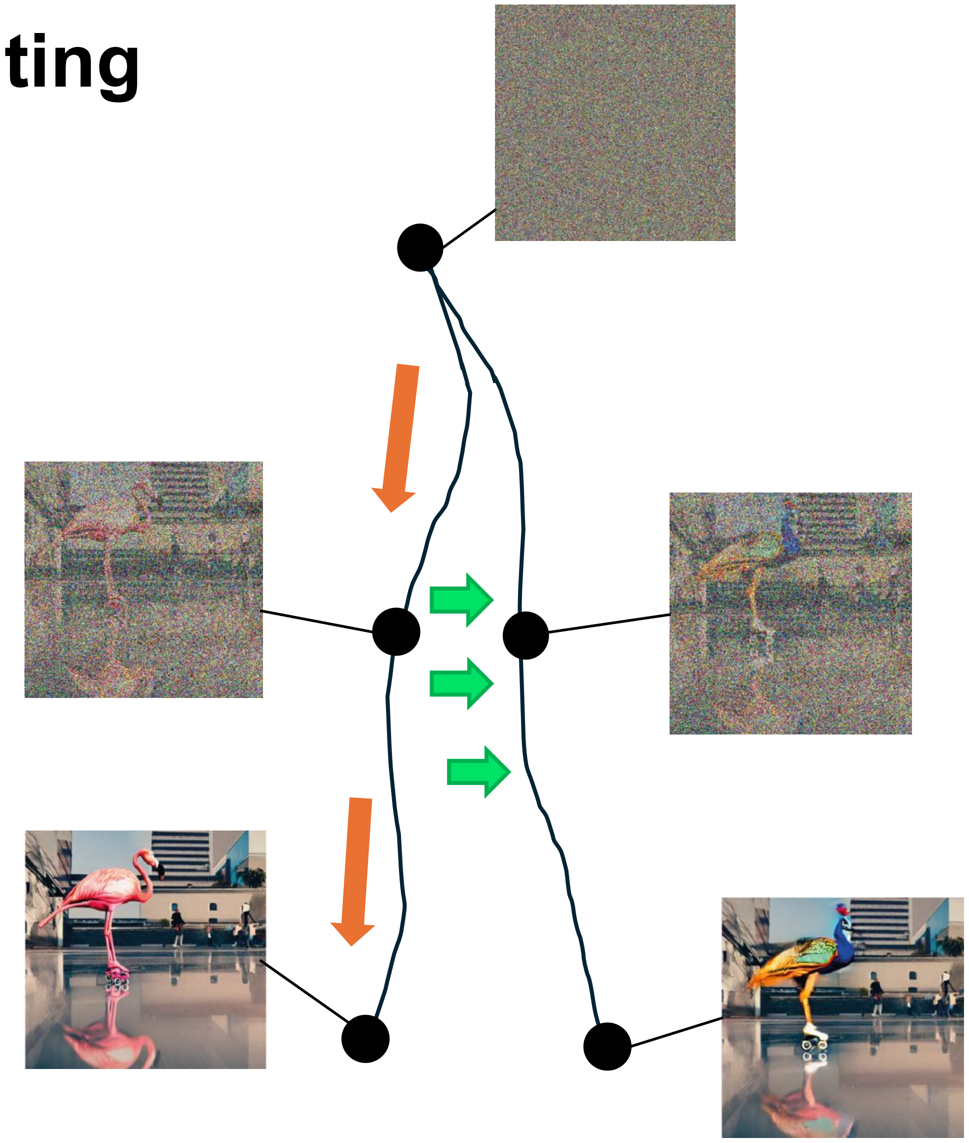
Closed form solution

$$\tilde{x} = (1-\gamma)\hat{x}_{0|t}(c_\phi) + \gamma \hat{x}_{0|t}(c_{tgt})$$

$$= x_{0|t}(c_\phi) + \gamma\sqrt{1-\bar{\alpha}_t}\left(\epsilon_\theta(x_t, t, c_\phi) - \epsilon_\theta(x_t, t, c_{tgt})\right)$$

: Gradient of Delta Denoising Score (DDS)

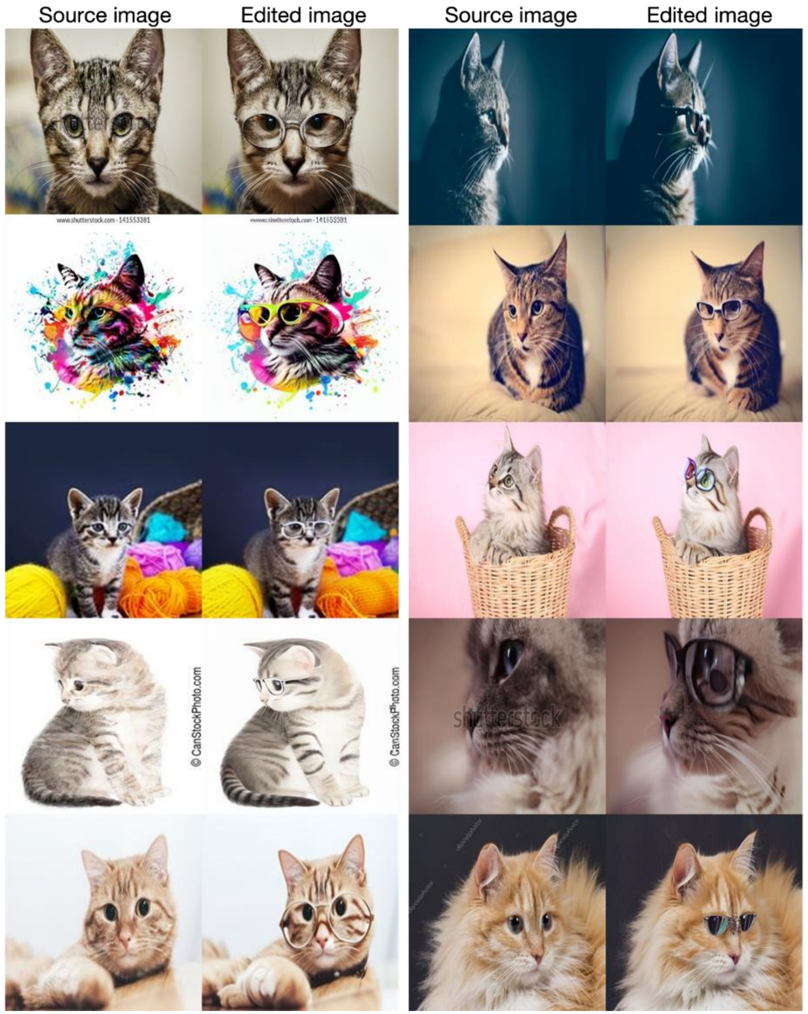
$$dx_t = \left[\frac{x_t - D_\theta(x_t) - \lambda\left(\epsilon_\theta(x_t, t, c_\phi) - \epsilon_\theta(x_t, t, c_{tgt})\right)}{t}\right] dt$$



Qualitative results

Horse → Zebra

Cat → Cat with glasses



BiSPL

KAIST

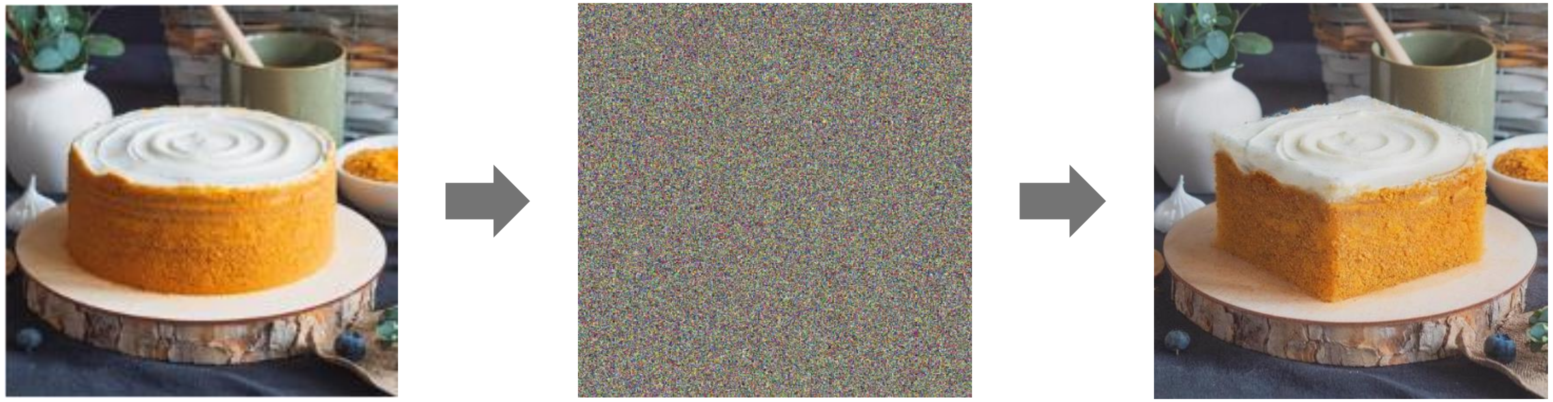
Work 2: Optimal Control

Image editing without inversion

FlowAlign, *J.Kim*, Y.Hong*, J.Park, & J.C.Ye, ICLR 2026*

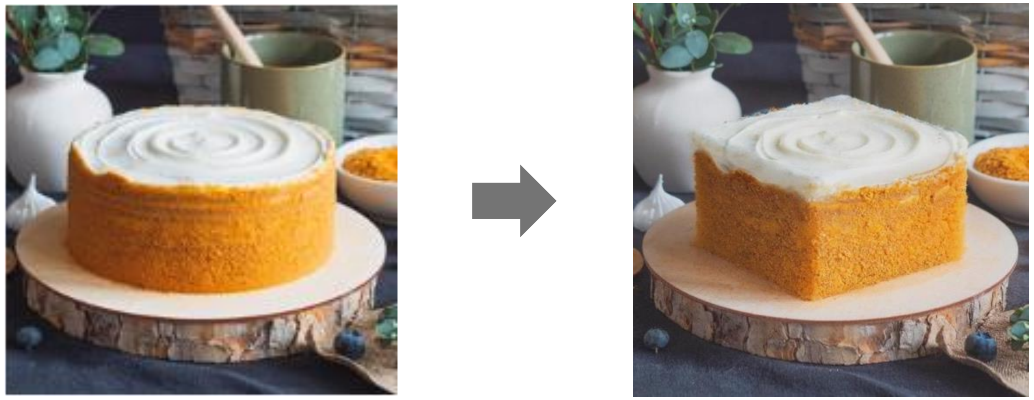
Application 2. Image editing without inversion

Inversion & Text-guided Sampling

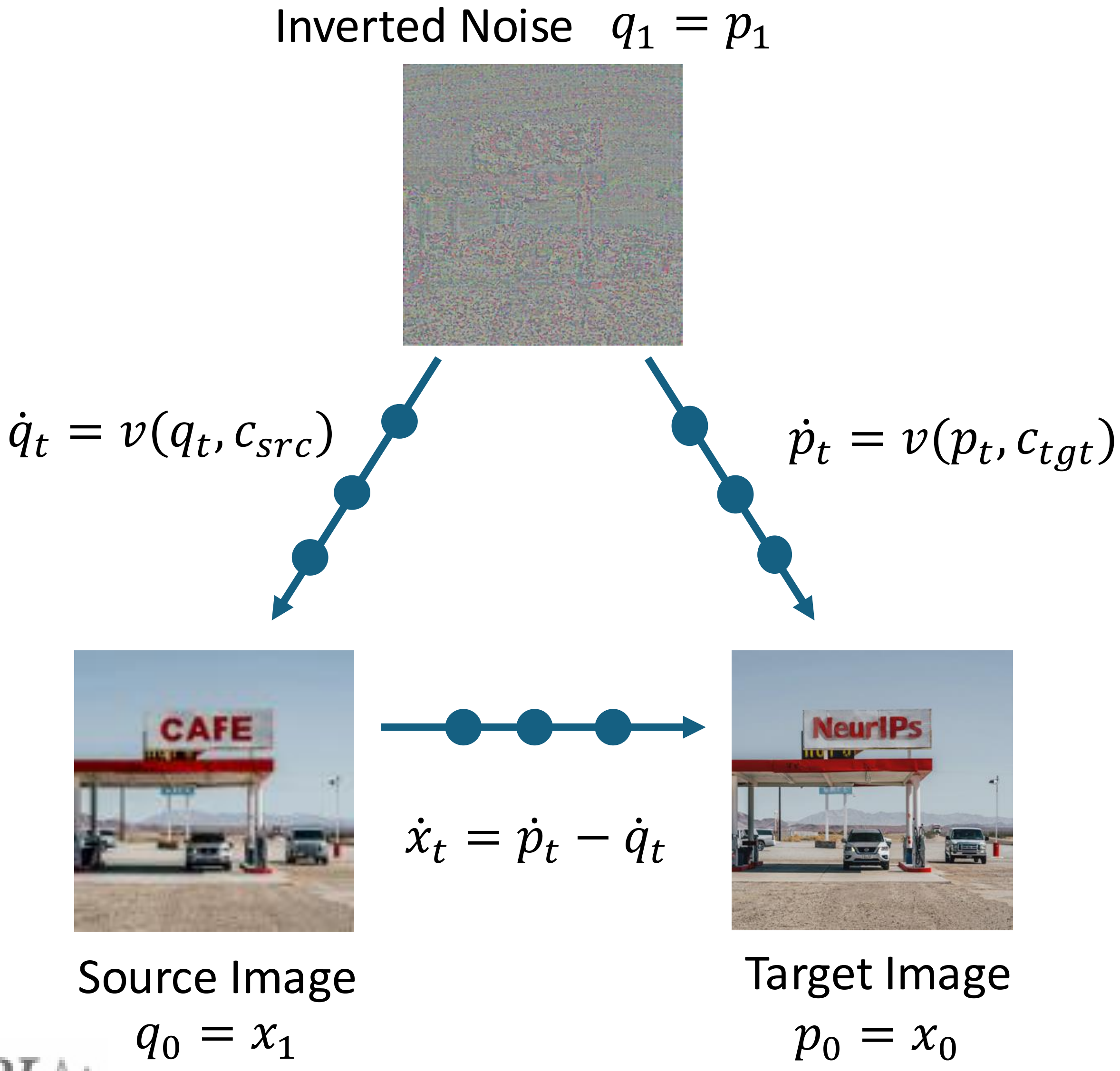


x2 Cost

Inversion Free Sampling



Simulate image-to-image translation



$$q_t = (1-t)X_1 + t\epsilon$$

$$p_t = (1-t)X_0 + t\epsilon$$

$$X_t = (1-t)X_0 + tX_1$$

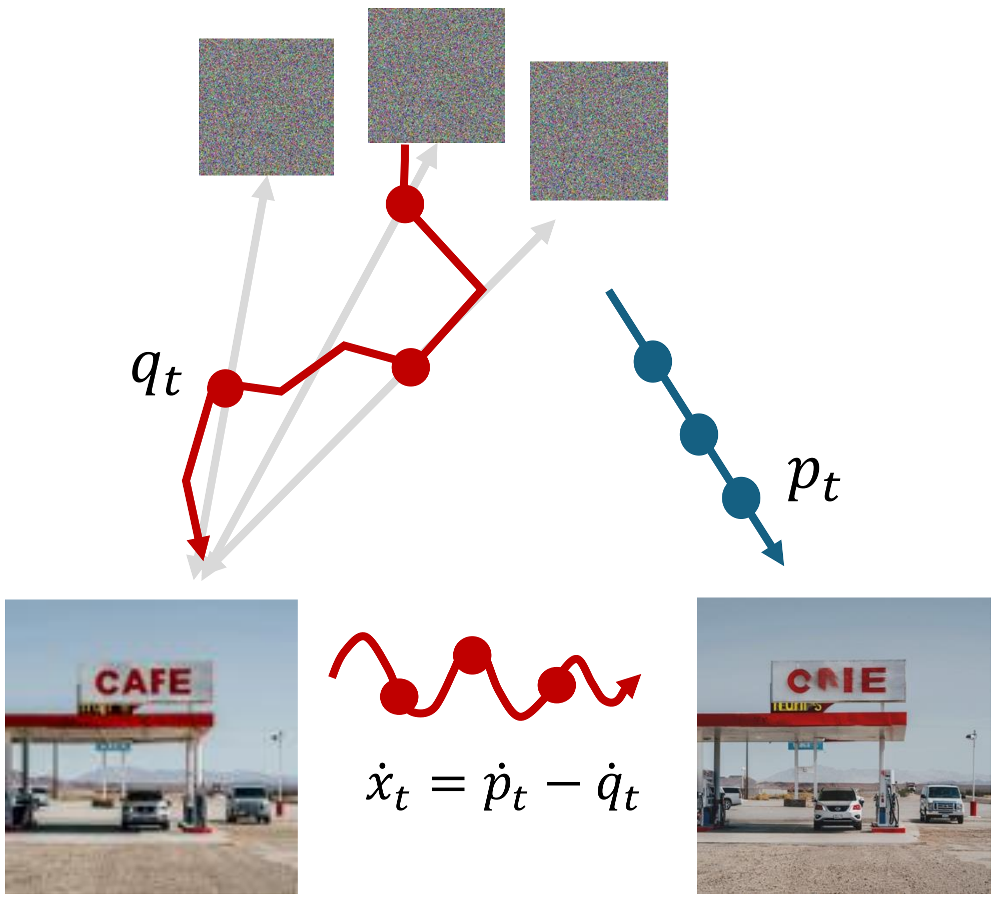
$$X_t - X_1 = p_t - q_t$$

$$dX_t = [dp_t - dq_t]dt$$

$$= [v_\theta(p_t, c_{tgt}) - v_\theta(q_t, c_{src})]dt$$

Inversion-free by replacing source trajectory with forward process

Random Noise $q_1 \sim \mathcal{N}(0, I)$



Source Image
$q_0 = x_1$

Target Image
$p_0 = x_0$

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

$$q_t = (1-t)X_1 + t\epsilon$$

$$p_t = q_t + X_t - X_1$$

$$dX_t = dp_t - dq_t$$
$$= \left[v_\theta\left(p_t, c_{tgt}\right) - v_\theta\left(q_t, c_{src}\right)\right]dt$$

Kulikov et al., 2025 ICCV (oral)

BiSPL KAIST

Optimal Control driven Source Consistency Regularization

$$V(u_t) = \int_0^1 \ell(x_t, u_t, t)dt + m(x_0)$$

Running Cost $\ell(x_t, u_t, t) := \frac{1}{2}\left\| u_t - \left(v_\theta\left(p_t, c_{tgt}\right) - v_\theta\left(q_t, c_{src}\right)\right)\right\|^2$

Terminal Cost $m(x_0) := \frac{\eta}{2}\|x_0 - x_{src}\|^2$

Solution $\underset{u}{\operatorname{argmin}} V(u_t) = v_\theta\left(p_t, c_{tgt}\right) - v_\theta(q_t, c_{src}) + \gamma(\mathbb{E}[p_0|p_t] - \mathbb{E}[q_0|q_t])$

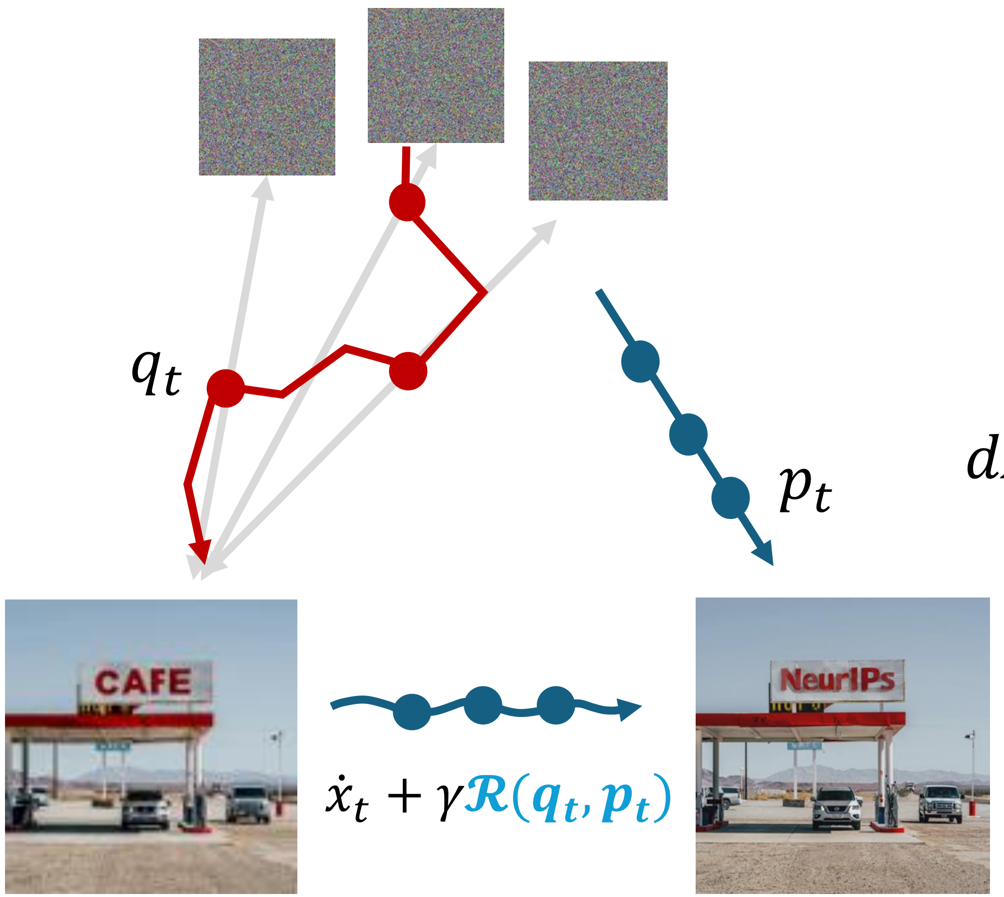
$$\therefore \quad x_t \leftarrow x_t + \left[v_\theta\left(p_t, c_{tgt}\right) - v_\theta(q_t, c_{src})\right]dt - \gamma dt(\mathbb{E}[q_0|q_t] - \mathbb{E}[p_0|p_t])$$

BiSPL

KAIST

FlowAlign: Source Consistency Regularization for smooth trajectory

Random Noise $q_1 \sim \mathcal{N}(0, I)$



Target Image

$q_0 = x_1$ $\quad$ $p_0 = x_0$

$$q_t = (1 - t)X_1 + t\epsilon$$

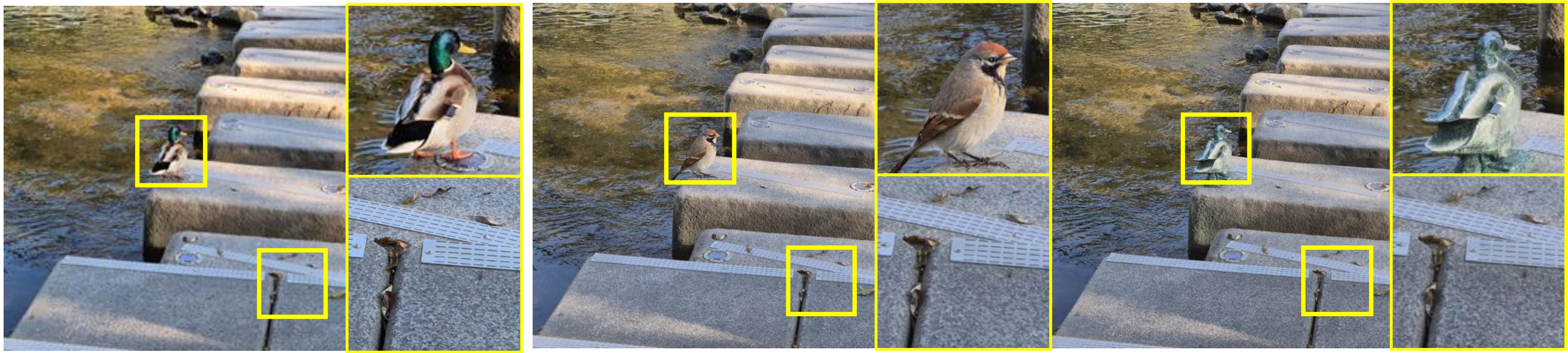
$$p_t = q_t + X_t - X_1$$

$$dX_t = dp_t - dq_t$$

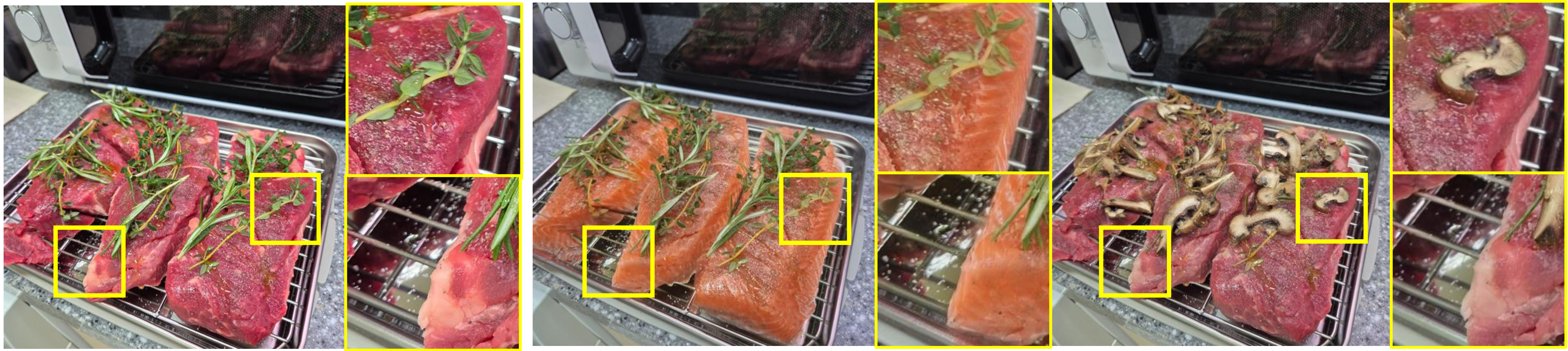
$$= \left[v_\theta(p_t, c_{tgt}) - v_\theta(q_t, c_{src})\right]dt + \gamma R(q_t, p_t)$$

FlowAlign – Qualitative Results

Source Image | Edited 1 | Edited 2



"duck" | "sparrow" | "stone statue"



"raw meat with herb" | "raw salmon with herb" | "raw meat with mushroom"

FlowAlign – Qualitative Results

Source GS

Edited GS (“Joker”)

Work 3: Reward optimization

Inverse problem solving via posterior sampling

DPS, *H.Chung*, J.Kim*, M.T.Mccann, M.L.Klasky, & J.C.Ye, ICLR 2023 (spotlight)*

FlowDPS, *J.Kim*, B.S.Kim* & J.C.Ye, ICCV 2025*

Reward-based fine-tuning yields optimal marginal density

Reward optimization

$$\max_{\theta} \quad r(x_0) + \alpha KL(p_0^{\theta} || p_0^{ref})$$

Closed-form terminal density function

$$p_0^*(x_0) \propto p_0^{\mathrm{ref}}(x_0) \exp\left(\frac{r(x_0)}{\alpha}\right)$$

Closed-form marginal density function

$$p_t^*(x_t) = \int p_t(x_t|x_0) p_0^*(x_0) dx_0$$

$$= \frac{1}{Z_0} p_t^{ref}(x_t) \mathbb{E}_{p(x_0|x_t)}\left[\exp\left(\frac{r(x_0)}{\alpha}\right)\right]$$

Reward guidance appears on denoised estimate

$$p_t^*(x_t) \propto p_t^{ref}(x_t)\mathbb{E}_{p(x_0|x_t)}\left[\exp\left(\frac{r(x_0)}{\alpha}\right)\right]$$

$$\nabla_{x_t}\log p_t^*(x_t) = \nabla_{x_t}\log p_t^{ref}(x_t) + \nabla_{x_t}\log \mathbb{E}_{p(x_0|x_t)}\left[\exp\left(\frac{r(x_0)}{\alpha}\right)\right]$$

Guidance to score function = Guidance to denoised estimate

$$\mathbb{E}_{p_t^*}[x_0|x_t] = \mathbb{E}_{p_t^{ref}}[x_0|x_t] + \eta_t\nabla_{x_t}\log \mathbb{E}_{p(x_0|x_t)}\left[\exp\left(\frac{r(x_0)}{\alpha}\right)\right]$$

Inverse problem: From Measurement to Reconstruction

Clean Image x

Measurement y

Solution x^*

Imaging system

$$y = A(x) + \epsilon$$

$$\epsilon \sim \mathcal{N}(0, \sigma^2 I)$$

Reconstruction

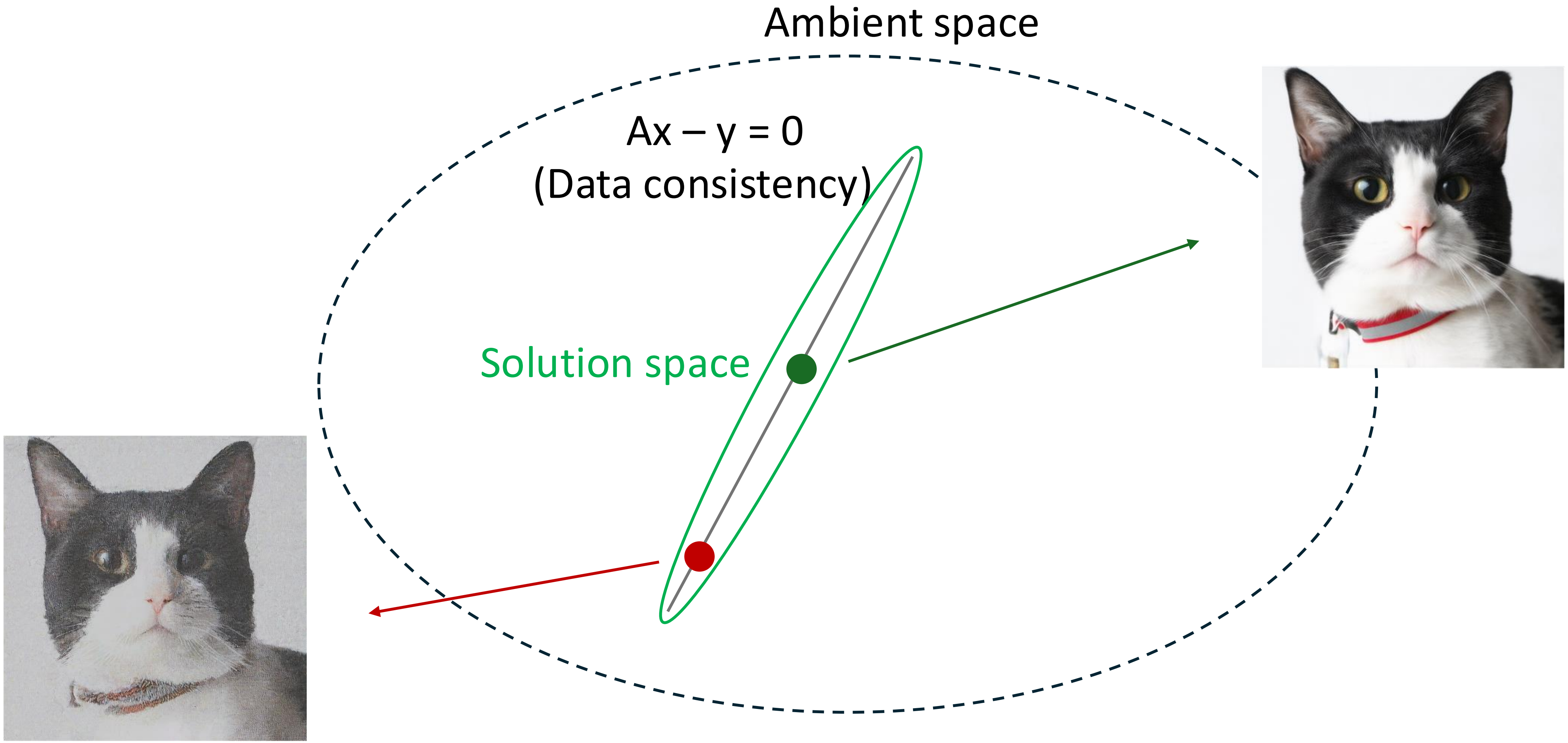
$$y - A(x^*) = 0$$

Data consistency

BiSPL

KAIST

Inverse problem: Finding a Plausible Set of Solutions



Key Factors to Constrain the Solution Space

$$y = A(x) + \epsilon$$
$$\epsilon \sim \mathcal{N}(0, \sigma^2 I)$$

$$\max_x \ \log p(x|y)$$

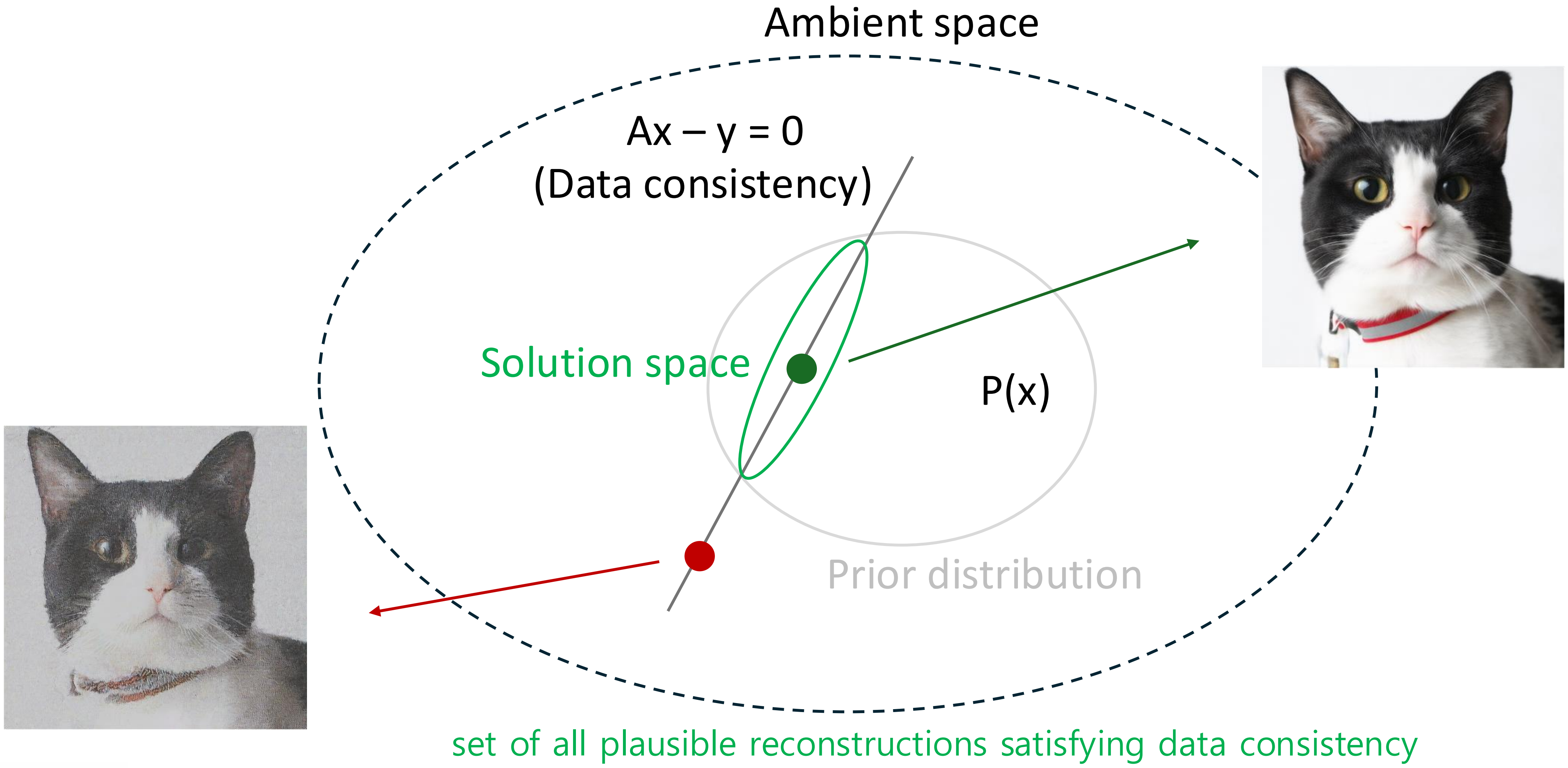
$$= \max_x \ \log p(y|x) + \log p(x)$$

$$= \max_x \ -\frac{\|y - A(x)\|^2}{2\sigma^2} + \log p(x)$$

$$= \min_x \ \frac{\|y - A(x)\|^2}{2\sigma^2} - \log p(x)$$

Data consistency **Image prior**

Diffusion Models as Posterior Samplers



Data consistency as reward function

$$r(x_0) = -\frac{\|y - A(x_0)\|^2}{2\sigma_y^2}$$

$$\nabla_{x_t} \log p_t^*(x_t) = \nabla_{x_t} \log p_t^{ref}(x_t) + \nabla_{x_t} \log \mathbb{E}_{p(x_0|x_t)}\left[\exp\left(\frac{r(x_0)}{\alpha}\right)\right]$$

$$= \nabla_{x_t} \log p_t^{ref}(x_t) + \nabla_{x_t} \log \mathbb{E}_{p(x_0|x_t)}\, p(y|x_0)^{1/\alpha}$$

Approximation and $\alpha = 1$:

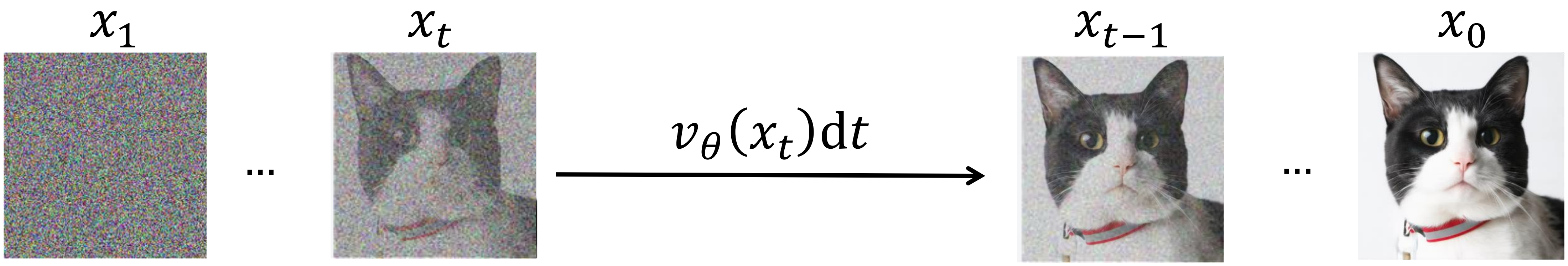
$$\approx \nabla_{x_t} \log p_t^{ref}(x_t) + \nabla_{x_t} \log p(y|\mathbb{E}[x_0|x_t]) \qquad \text{(DPS)}$$

Application 3. Inverse problem solver based on affine flow models

Affine Conditional flows are defined as $\psi_t(x|x_0) = a_t x_0 + b_t x$

$\psi_0(x|x_0) = x$ $\quad\quad$ $\psi_1(x|x_0) = x_0$ $\quad\quad$ (Boundary Condition)

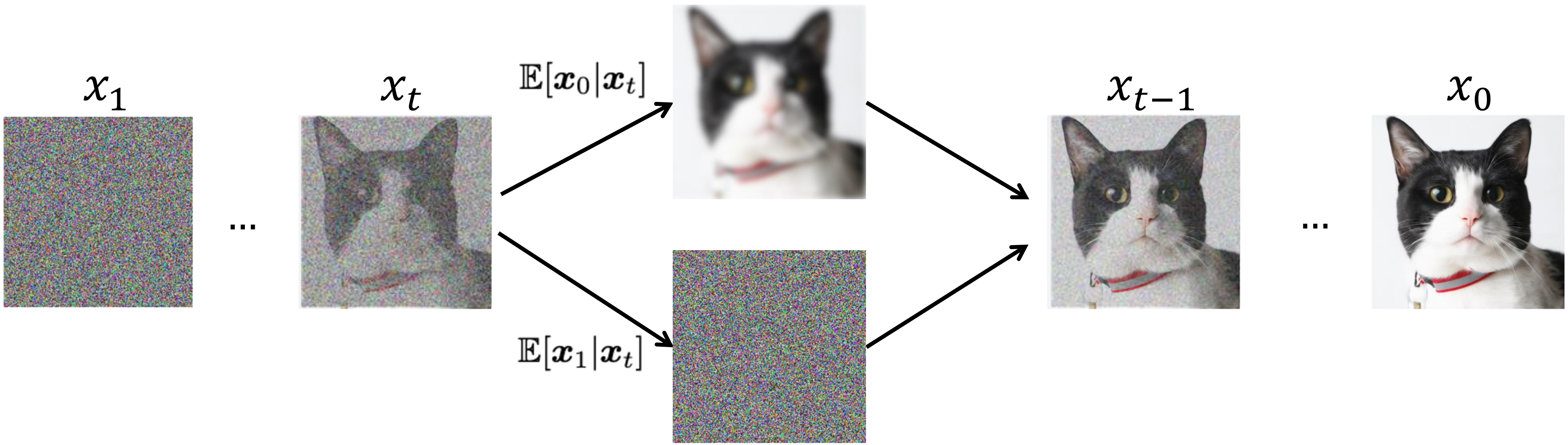
Corresponding velocity: $v_t(x_t|x_0) = \dot{\psi}_t(x_1|x_0) = \dot{a}_t x_0 + \dot{b}_t x_1$



BiSPL

KAIST

One-step Euler update is decomposed into Tweedie estimates



Proposition 1 (Tweedie Formula). *The denoised and noisy estimate given $\boldsymbol{x}_t$ are given by*

$$\mathbb{E}[\boldsymbol{x}_0|\boldsymbol{x}_t] = \left[a_t - \dot{a}_t \frac{b_t}{\dot{b}_t}\right]^{-1} \left(\boldsymbol{x}_t - \frac{b_t}{\dot{b}_t} v_t(\boldsymbol{x}_t)\right) := \hat{x}_{0|t}$$

$$\mathbb{E}[\boldsymbol{x}_1|\boldsymbol{x}_t] = \left[b_t - \dot{b}_t \frac{a_t}{\dot{a}_t}\right]^{-1} \left(\boldsymbol{x}_t - \frac{a_t}{\dot{a}_t} v_t(\boldsymbol{x}_t)\right) := \hat{x}_{1|t}$$

$$\begin{aligned} x_{t-1} &= x_t + v_\theta(x_t)dt \\ &= C_1(t)\,\hat{x}_{0|t} + C_2(t)\hat{x}_{1|t} \end{aligned}$$

$$C_1(t) = a_t + \dot{a}_t dt$$

$$C_2(t) = b_t + \dot{b}_t dt$$

FlowDPS: Posterior Sampling with General Affine Flow Models

Posterior Sampling can be done by

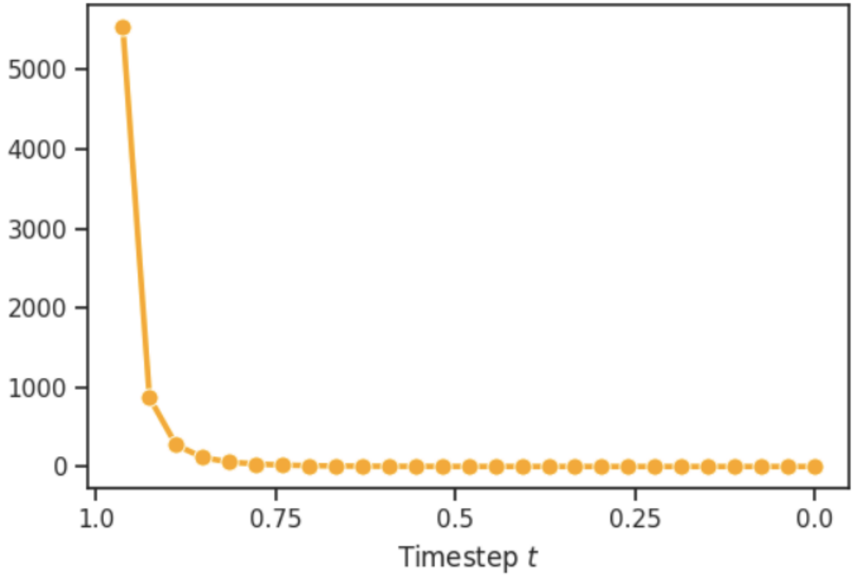
$$dx_t = v(x_t, t)dt - \lambda_t \nabla_{\hat{x}_{0|t}} \log p(y|\hat{x}_{0|t})$$

$$where \quad \hat{x}_{0|t} \coloneqq \mathbb{E}[x_0|x_t] \quad and \quad \lambda_t = \frac{b_t}{a_t}\left(\frac{\dot{b}_t a_t - b_t \dot{a}_t}{a_t}\right)\mathrm{dt}$$

We can extend this to latent diffusion models
without loss of generality

FlowDPS: Thoery-driven step size

Example: $-\beta_t$ of linear flow

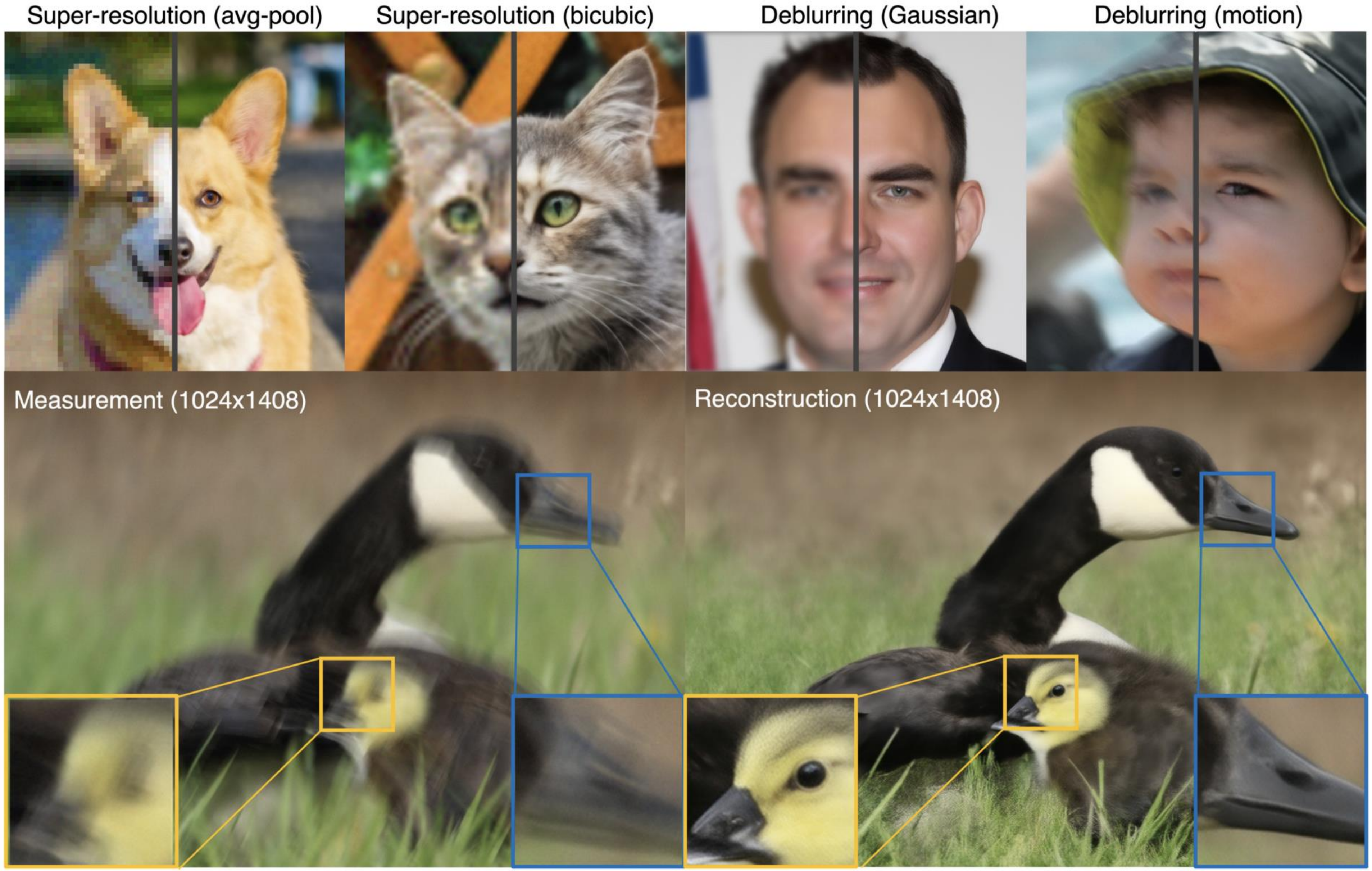


$$x_{t-1} = x_t + v_\theta(x_t)dt - \lambda_t \nabla_{\hat{x}_{0|t}} \log p(y|\hat{x}_{0|t})$$

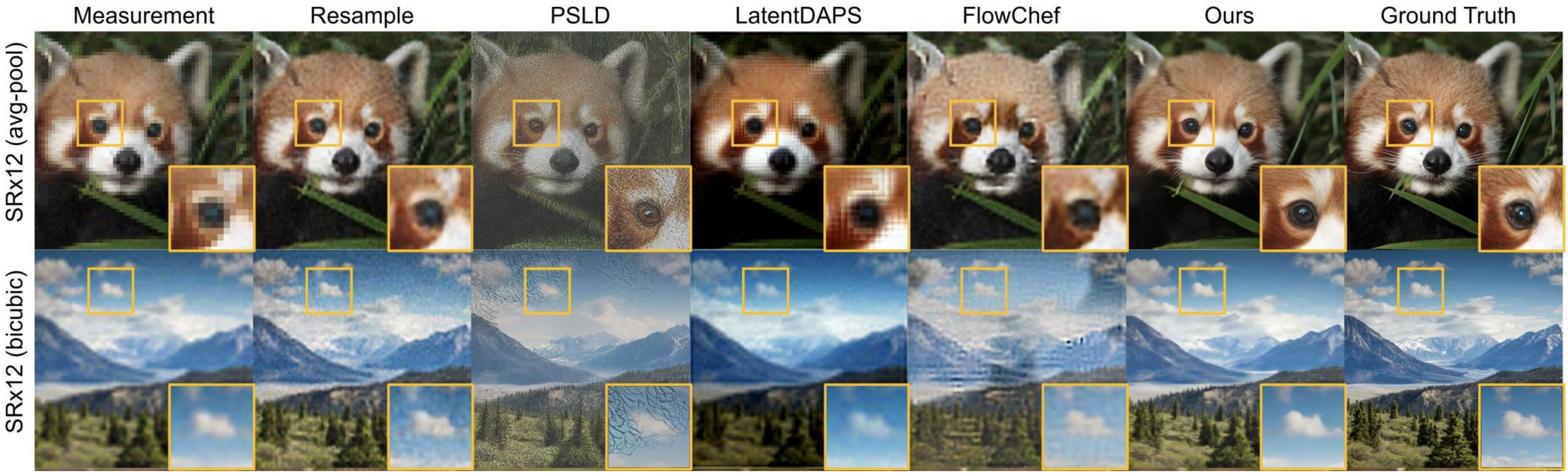
$$= C_1(t)\tilde{x}_{0|t} + C_2(t)\hat{x}_{1|t}$$

where $\tilde{x}_{0|t} = \hat{x}_{0|t} - \beta_t \nabla_{\hat{x}_{0|t}} \log p(y|\hat{x}_{0|t})$

FlowDPS: Qualitative results



FlowDPS: Qualitative comparison



BiSPL

KAIST

Conclusion

We solve multiple problems using pre-trained diffusion model

by applying guidance to denoised estimate

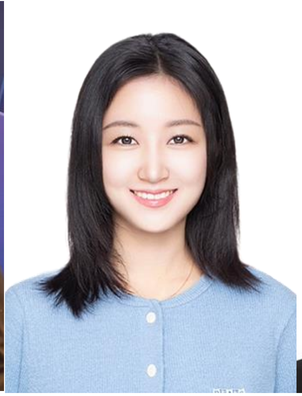
which is derived by task-specific optimization problem

Potential tasks – efficient sampling, minority, safety

Thank you for the attention

Hyungjin Chung

Geonyeong Park



Yeobin Hong

Jonghyun Park

Bryan S Kim

Jong Chul Ye