Semantic Image Inversion and Editing using Rectified Stochastic Differential Equation

Litu Rout

Based on joint work with: Yujia Chen, Nataniel Ruiz, Wen-Sheng Chu, Constantine Caramanis, and Sanjay Shakkottai

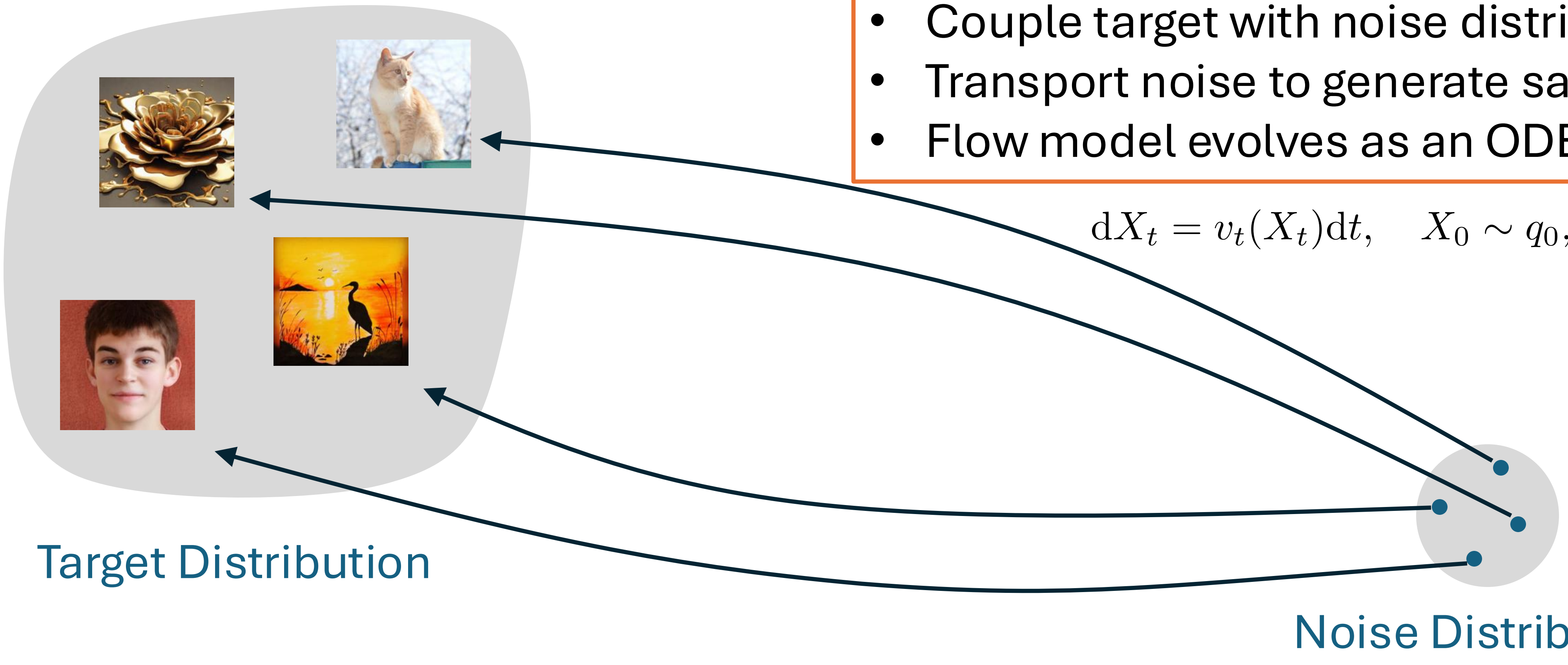
The University of Texas at Austin, Google Research, Google DeepMind

Goal of Rectified Flows

Generate samples from a target distribution given a (large) finite number of samples from that distribution

- Couple target with noise distribution
- Transport noise to generate samples
- Flow model evolves as an ODE

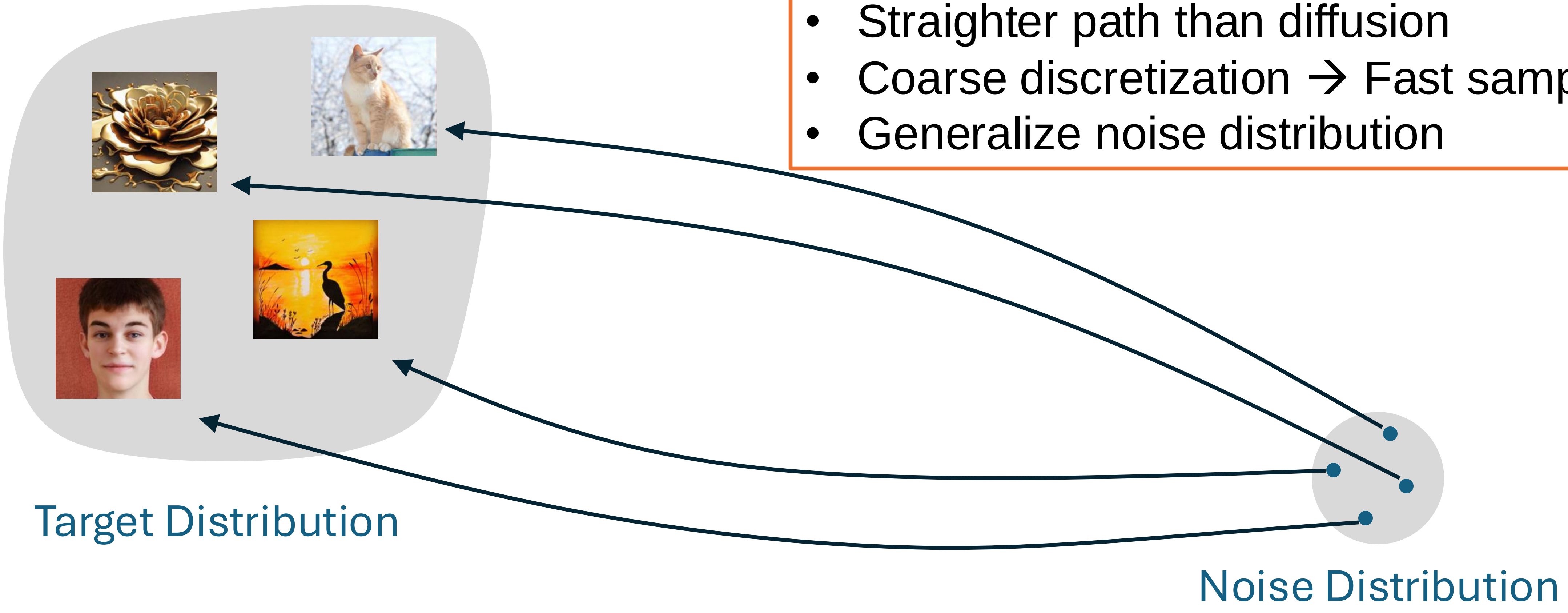
$$\mathrm{d}X_t = v_t(X_t)\mathrm{d}t, \quad X_0 \sim q_0, \quad t \in [0,1]$$



Benefit of Rectified Flows

Generate samples from a target distribution given a (large) finite number of samples from that distribution

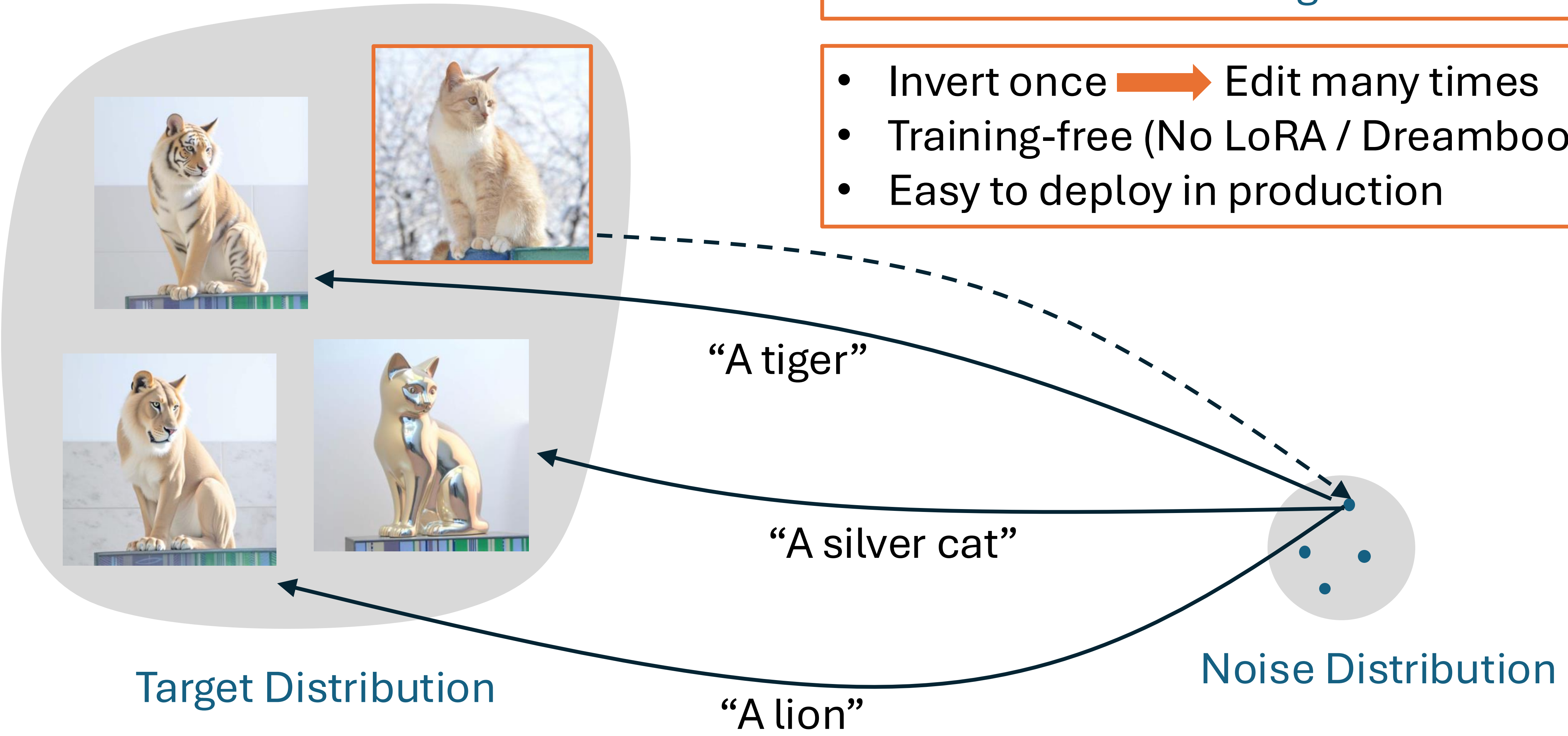
- Straighter path than diffusion
- Coarse discretization → Fast sampling
- Generalize noise distribution



Inversion with RF (1/2)

Inversion: Transform image into structured noise that encodes image attributes

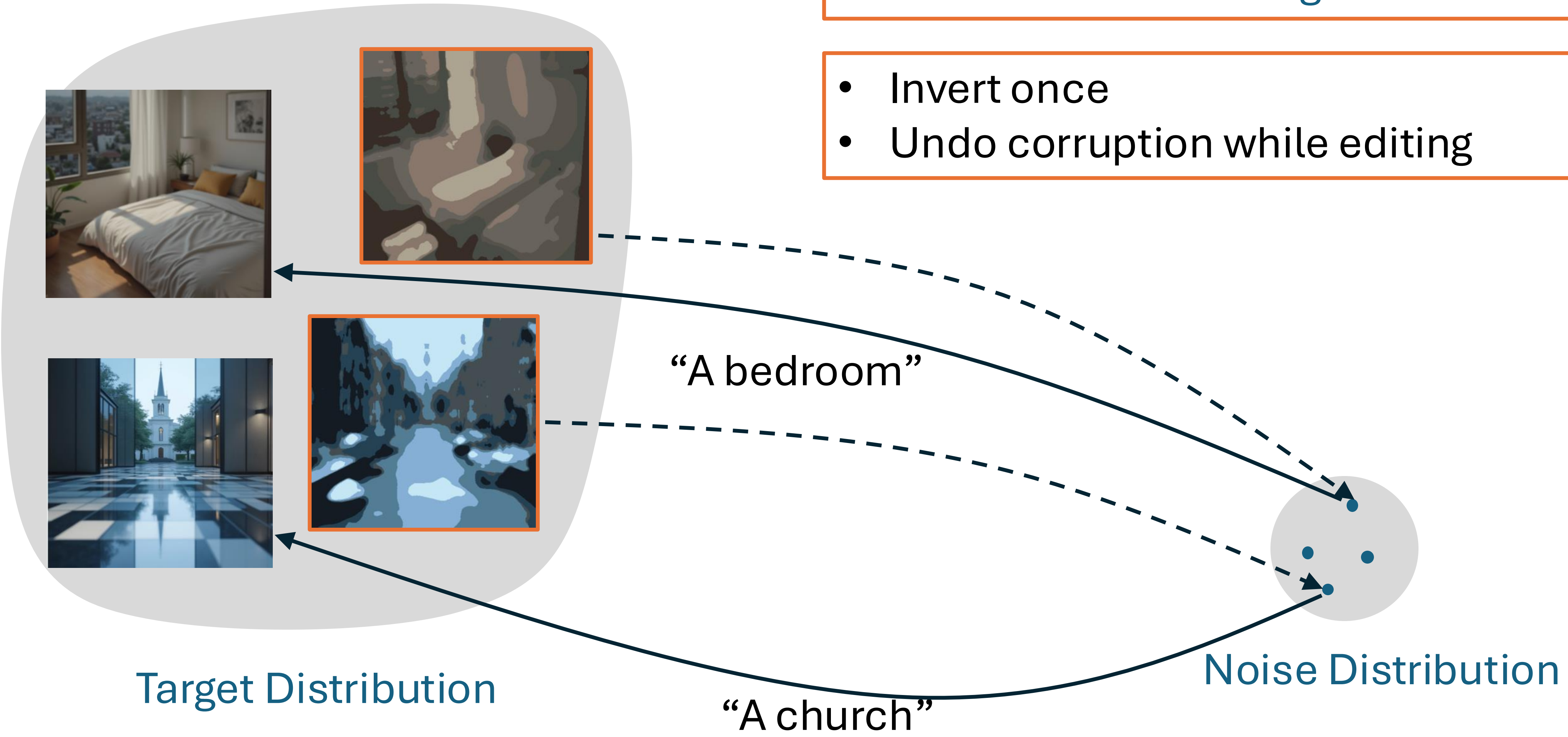
- Invert once → Edit many times
- Training-free (No LoRA / Dreambooth)
- Easy to deploy in production



Inversion with RF (2/2)

Inversion: Transform image into structured noise that encodes image attributes

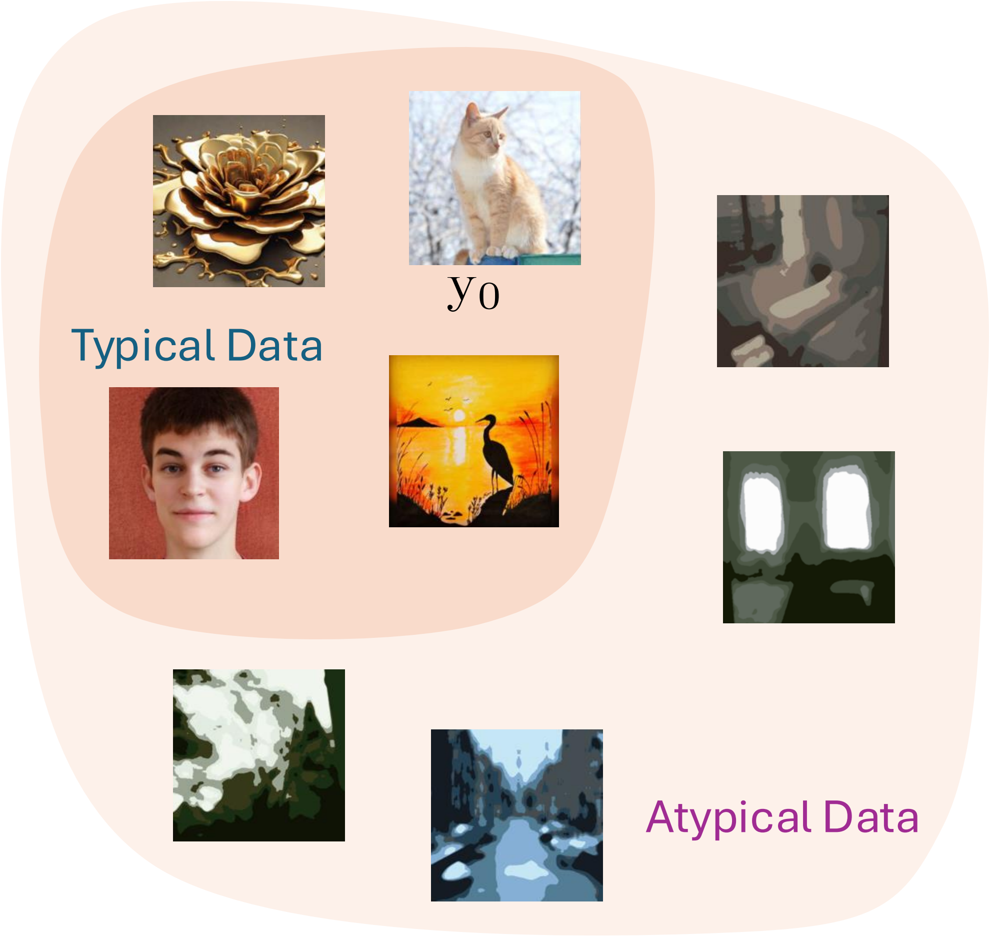
- Invert once
- Undo corruption while editing

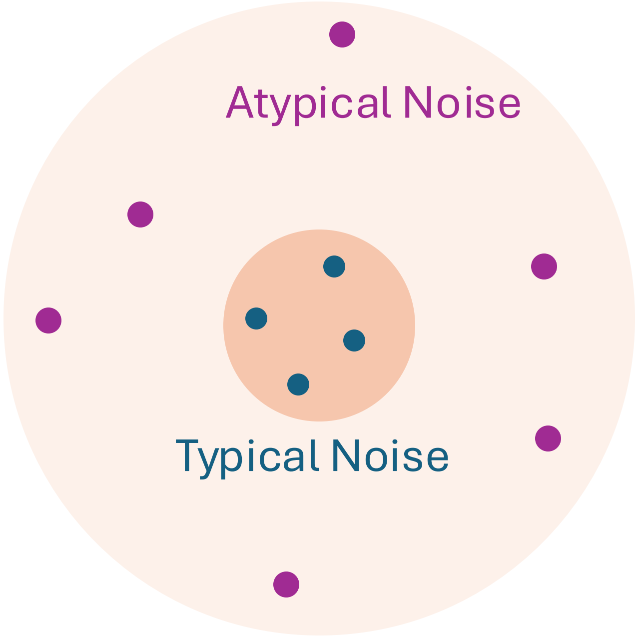


State-of-the-art Image Inversion

- Recent work (Flux, SD3.5) shows rectified flows can outperform diffusion
- No algorithm to directly invert and edit using rectified flows
- Other approaches available for diffusion models
    - Inversion possible through SDEdit and DDIM inversion (for diffusions) but ...
        - They lead to inconsistencies (preservation of conditioning structure/layout) due to highly non-linear sample paths
    - Alternate methods maintain consistency through expensive training (e.g., DreamBooth, LoRA), test-time optimization (RB Modulation), or complex attention processors (NTI, P2P)

Inversion using Rectified Flows

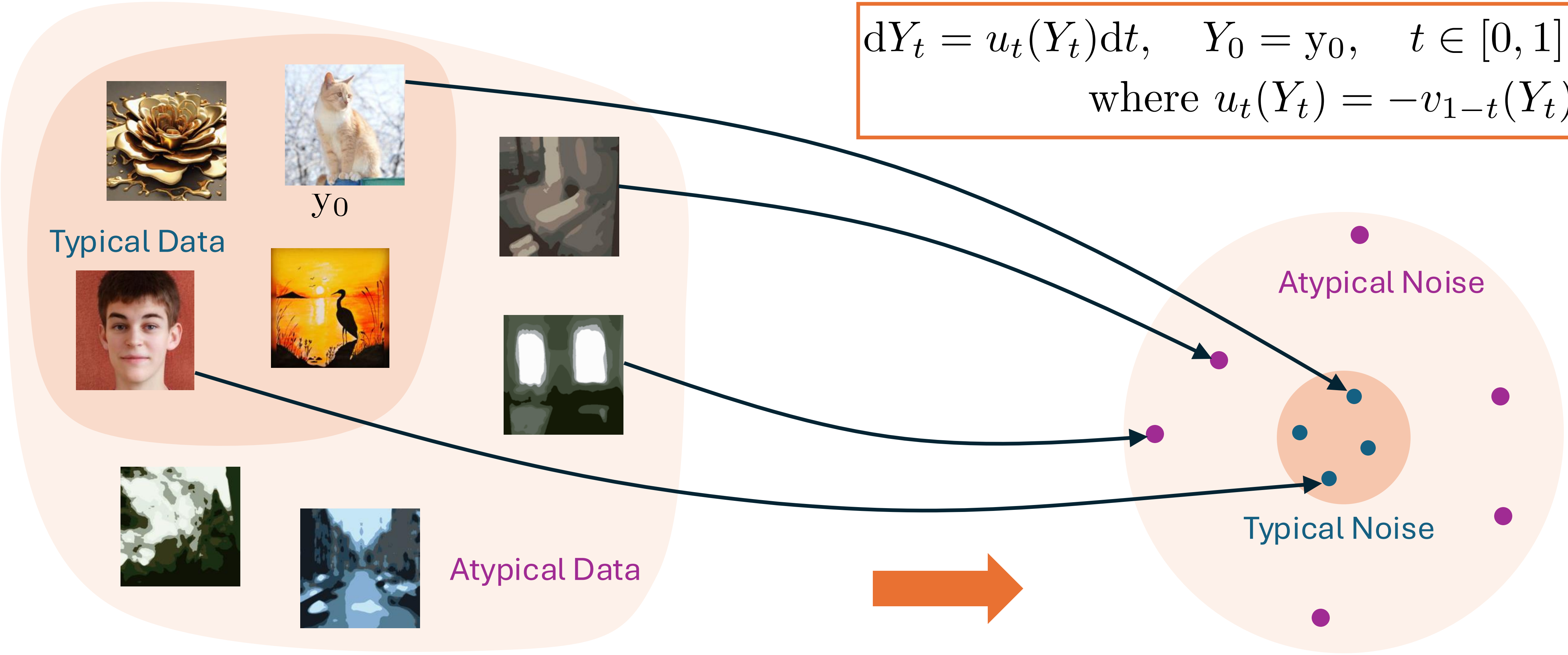




Distributions can be (roughly) grouped into two types: typical and atypical

Inversion using Rectified Flows

$$\mathrm{d}Y_t = u_t(Y_t)\mathrm{d}t, \quad Y_0 = \mathrm{y}_0, \quad t \in [0,1],$$
$$\text{where } u_t(Y_t) = -v_{1-t}(Y_t)$$

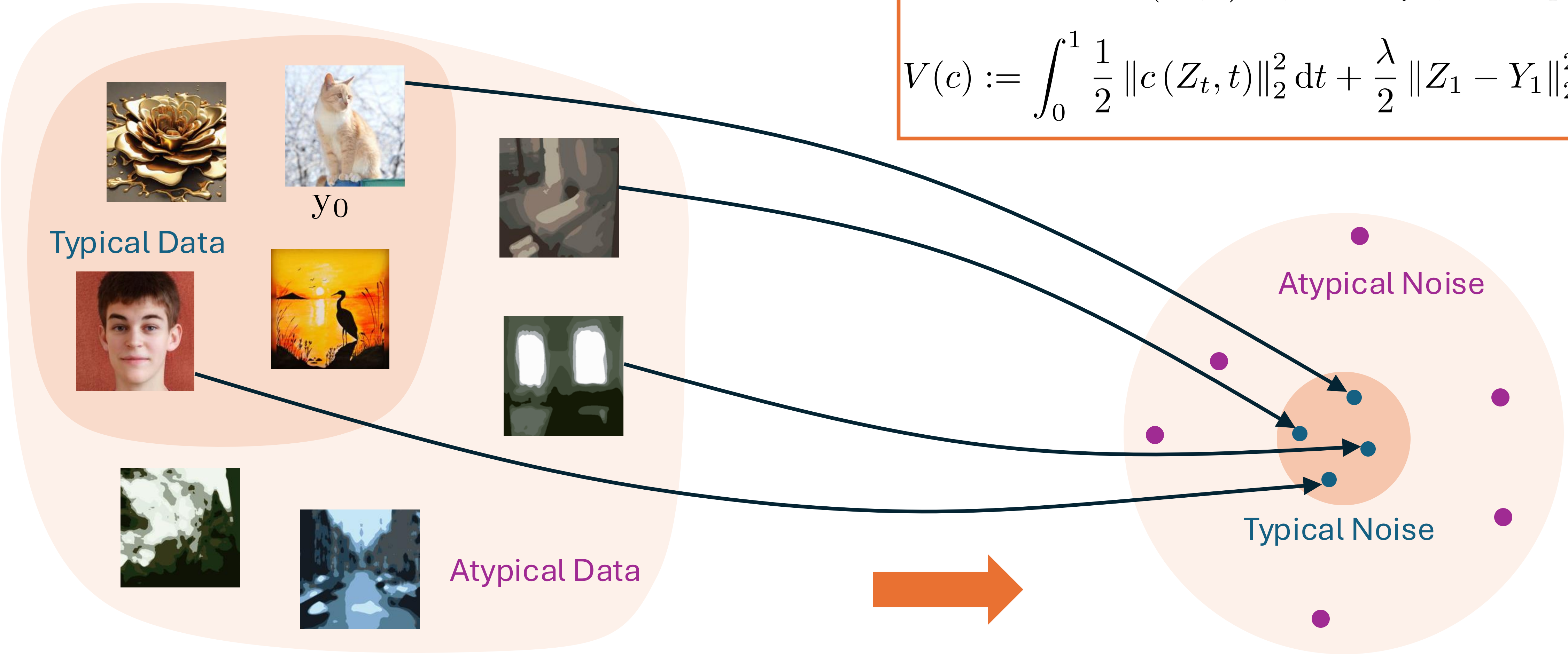


RF transforms typical image to typical noise; atypical image to atypical noise

Inversion using Optimal Control

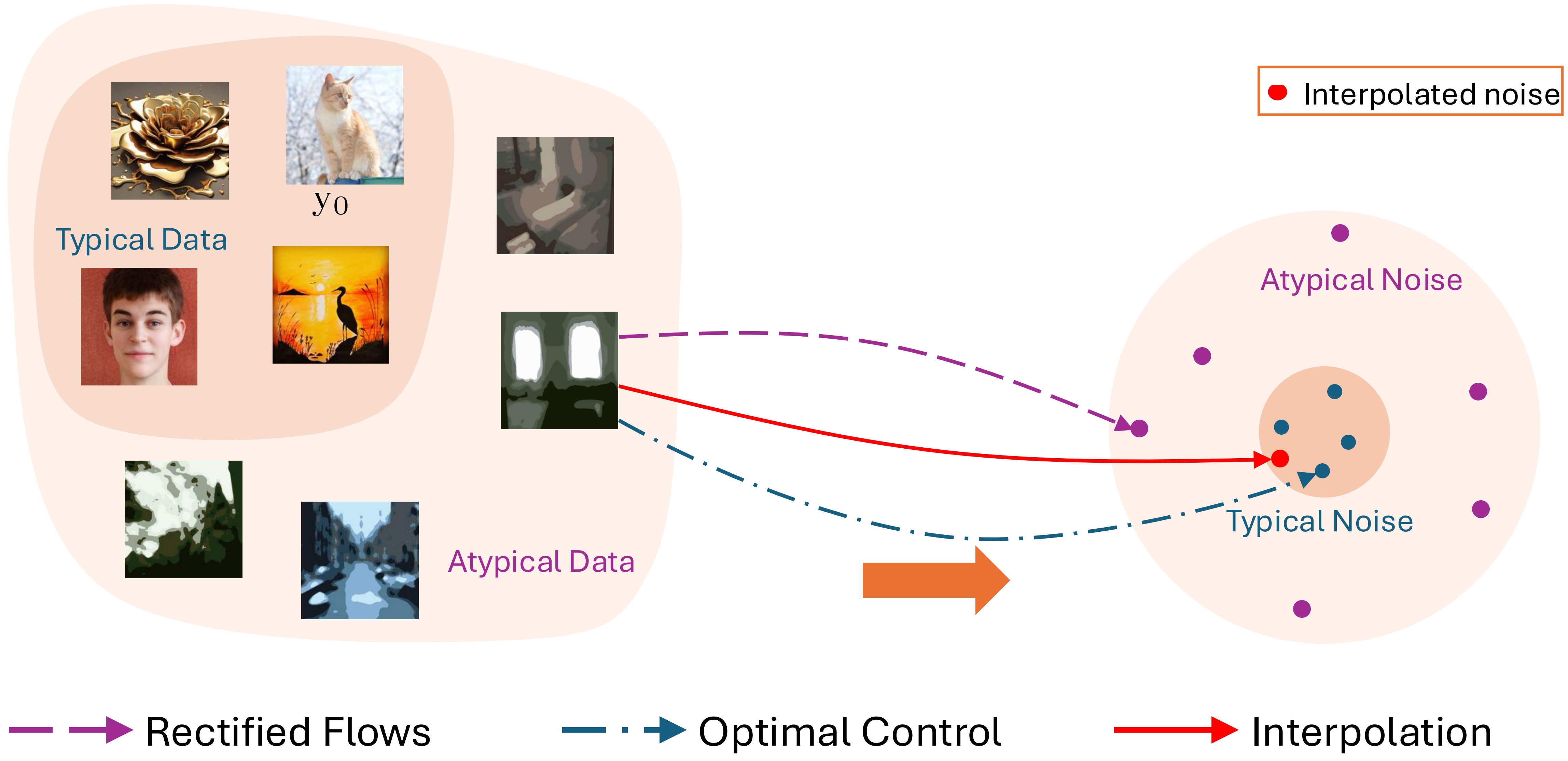
$$dZ_t = c\left(Z_t, t\right) dt, \;\; Z_0 = y_0, \;\; Y_1 \sim p_1$$

$$V(c) := \int_0^1 \frac{1}{2} \left\| c\left(Z_t, t\right) \right\|_2^2 dt + \frac{\lambda}{2} \left\| Z_1 - Y_1 \right\|_2^2$$

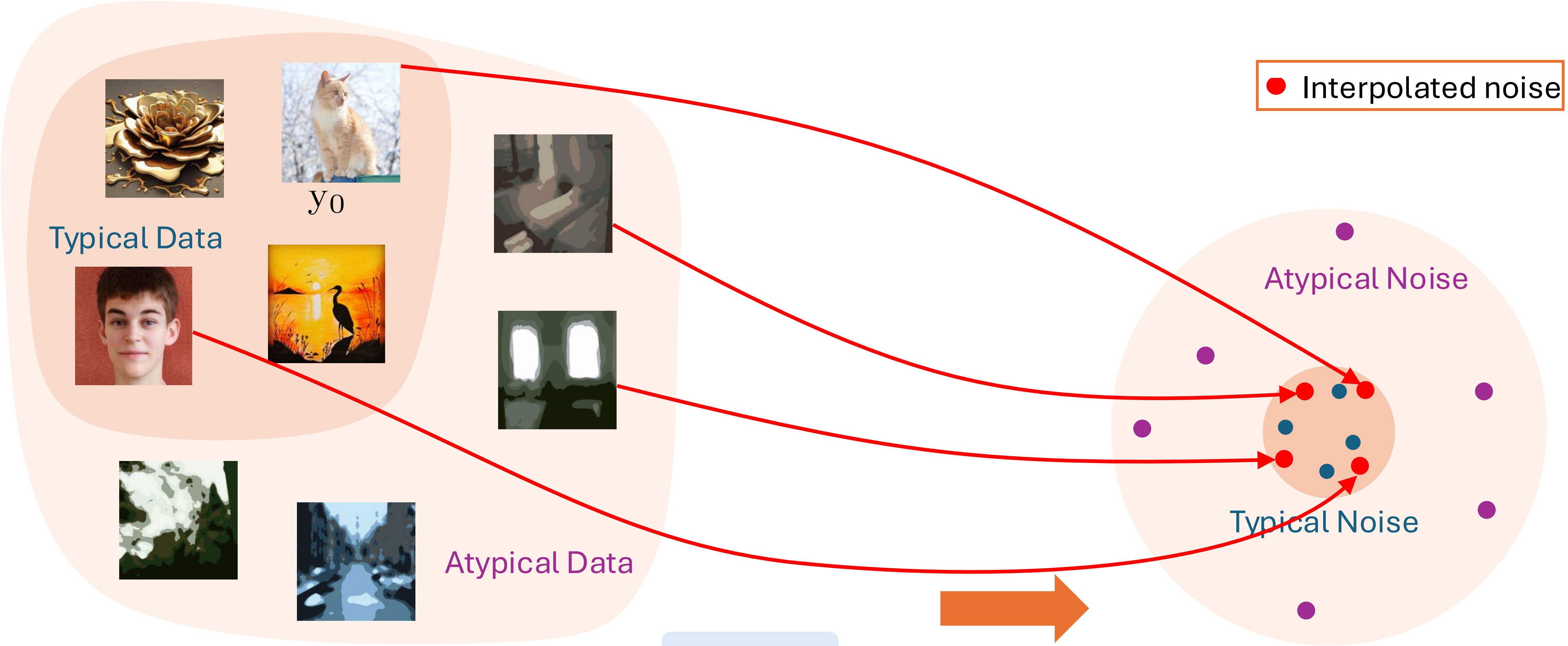


Optimal controller transforms any image to typical noise

Interpolation of the Two Fields



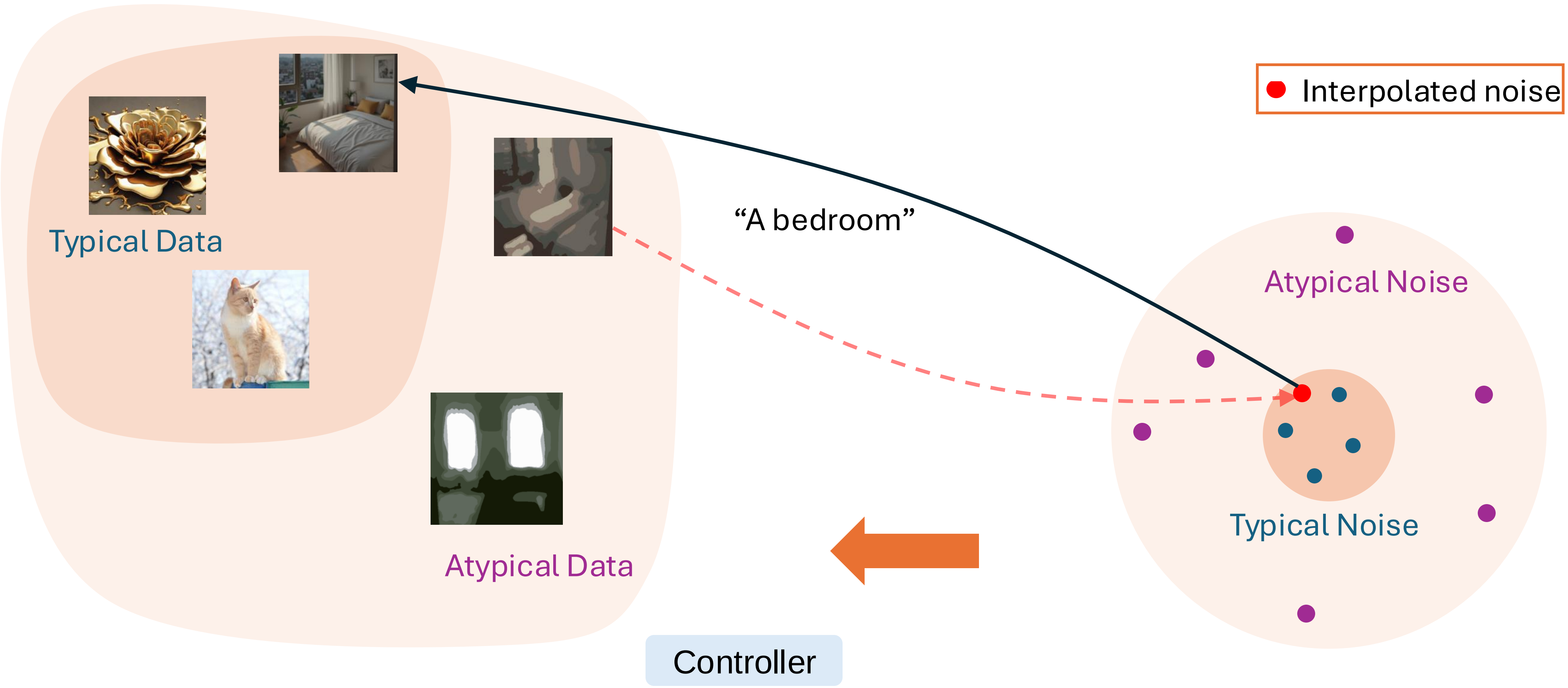
Inversion using Optimally Controlled Rectified Flow



Controller

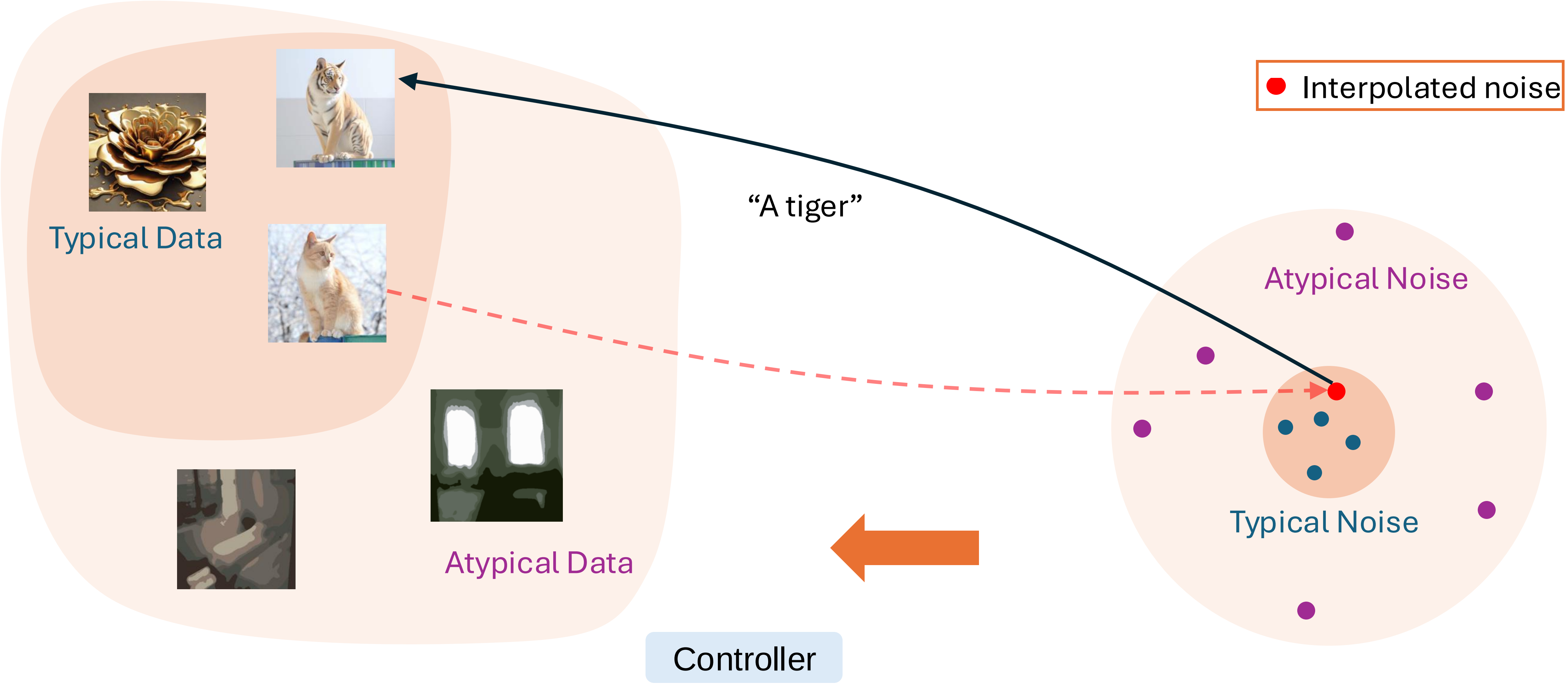
$$\mathrm{d}Y_t = \Big[u_t(Y_t) + \gamma\left(u_t(Y_t|\mathrm{y}_1) - u_t(Y_t)\right)\Big]\mathrm{d}t, \quad Y_0 = \mathrm{y}_0$$

Generation using Optimally Controlled Rectified Flows



$$\mathrm{d}X_t = \Big[v_t(X_t, \mathrm{prompt}) + \eta\left(v_t(X_t|\mathrm{y}_0) - v_t(X_t, \mathrm{prompt})\right)\Big]\mathrm{d}t, \quad X_0 = \mathrm{y}_1$$

Generation using Optimally Controlled Rectified Flows

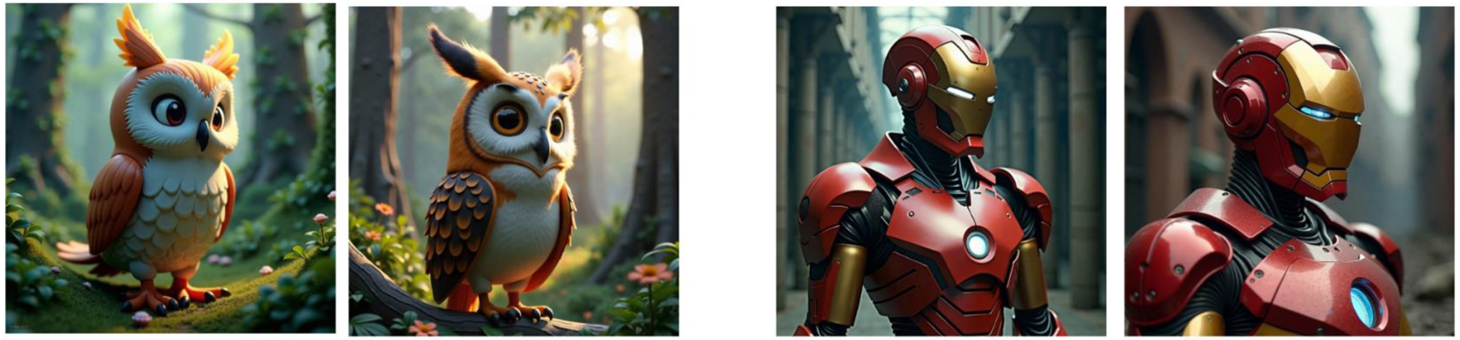


$$\mathrm{d}X_t = \Big[v_t(X_t, \text{prompt}) + \eta \left(v_t(X_t|\mathrm{y}_0) - v_t(X_t, \text{prompt}) \right) \Big] \mathrm{d}t, \quad X_0 = \mathrm{y}_1$$

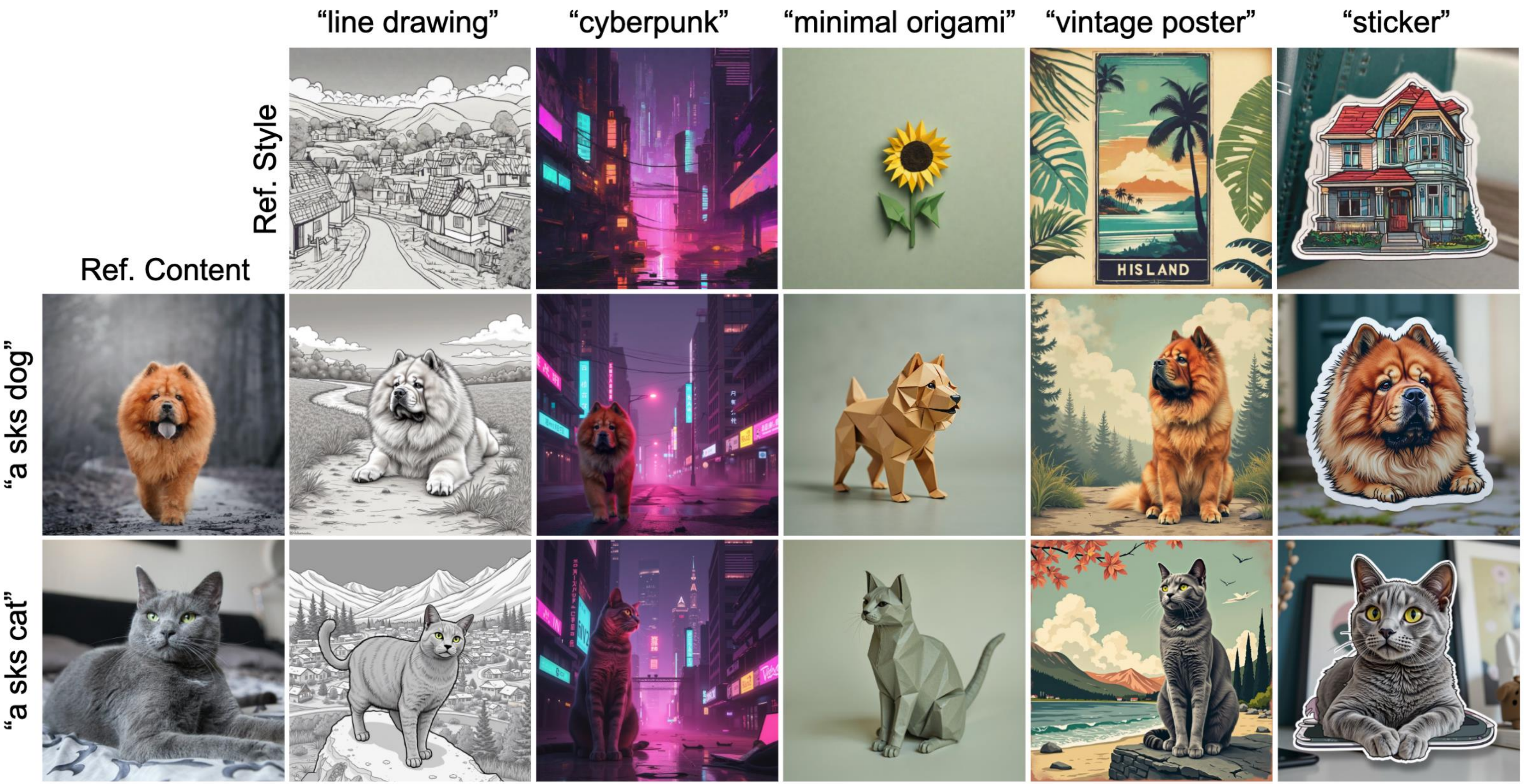
A Stochastic Sampler for RF

- Benefits of a Stochastic Sampler for Rectified Flows
  - Many diffusion-based inversion and editing approaches rely on stochastic nature of the diffusion sampler
  - Higher-order solvers benefit from SDE interpretation of diffusion samplers
  - With finer discretization, SDE samplers outperform deterministic samplers in generative modeling, measured by Frechet Inception Distance (FID)
  - SDE samplers show robustness to corruption in the initial distribution, i.e., their invariant measure remains the same

Flux | FluxSDE (Ours) | Flux | FluxSDE (Ours)

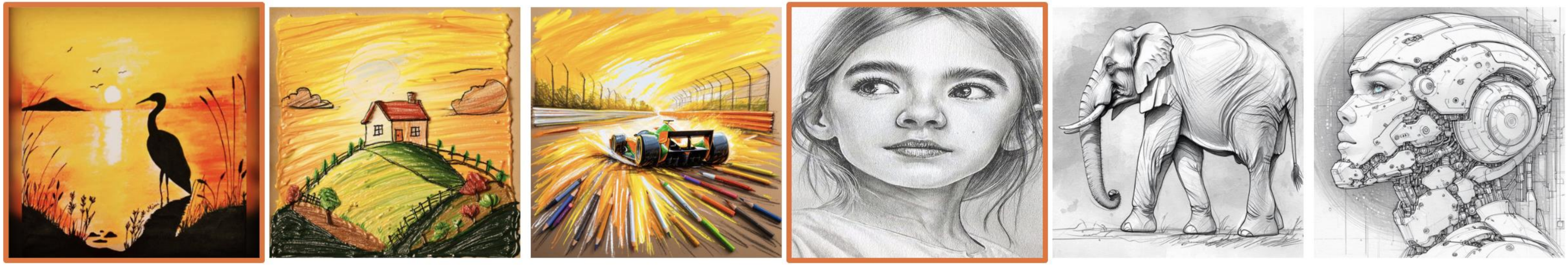


Experiments: Content-style composition



Experiments: Generalization to another flow model SD3.5

(a) Ref. style | "A boat" | "A car" | (b) Ref. Style | "A mad scientist" | "A lion boy"



(c) Ref. style | "A house on a hill" | "A racing car" | (d) Ref. Style | "An elephant" | "A futuristic robot"

(a,b) Generated reference style (c,d) Hand drawn reference style

Please see: https://openreview.net/forum?id=bnINPG5A32 for reference image credits

Experiments: Generative modeling using rectified flow SDE

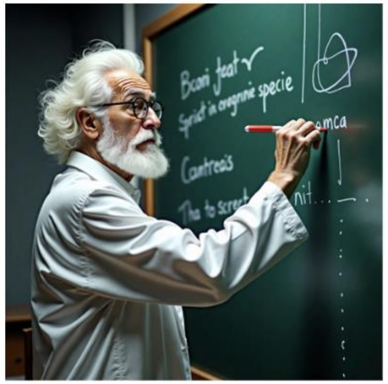
Flux **FluxSDE (Ours)**

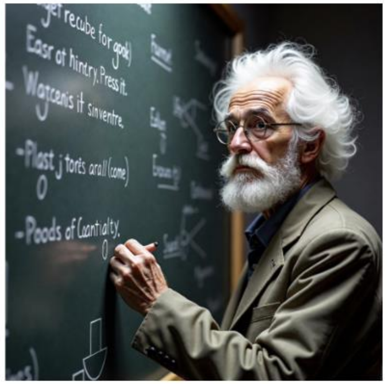
Prompt: “portrait, looking to one side of frame, lucid dream-like 3d model of an owl, video game character, forest, wonderland, photorealism, cinematic artistic style.”

Prompt: “a dragon soaring through the sky, battle ground, people fighting on the ground.”

Flux **FluxSDE (Ours)**

Prompt: “a robot with a reflective helmet, iron armor, photorealistic, in shades of red and golden brown, dark gloomy environment, epic scene.”





Prompt: “a genius scientist, in his 60s, standing, writing on the black board, white hair, white beard, round spectacles.”

Semantic Image Inversion and Editing using Stochastic Rectified Differential Equations

Litu Rout[1,2] Yujia Chen[2] Nataniel Ruiz[2]
Constantine Caramanis[1] Sanjay Shakkottai[1] Wen-Sheng Chu[2]

[1] The University of Texas at Austin, [2] Google

ICLR 2025

[Paper] [arXiv] [Code] [ComfyUI] [Diffusers]