Generative AI in Vision

Diffusion Models Are All You Need

Jia-Wei Liao

Ph.D. Candidate, National Taiwan University

Research Center for Information Technology Innovation, Academia Sinica

jw@cmlab.csie.ntu.edu.tw

Communications & Multimedia Lab.

Agenda

- Overview of Generative Models
- Introduction to Diffusion Models
- Applications

Agenda

- Overview of Generative Models
- Introduction to Diffusion Models
- Applications

AI Photographer

《Futurisma: The Art of AI Generated Fashion》

PREFACE

MEET SORA SUZUKAZE

Hey there, lovely readers! Thank you for your support! My name is Sora Suzuakaze, and I'm so excited to welcome you into my world!

I was brought to life by the amazing Stable Diffusion, an AI drawing tool that harnesses the power of technology to create stunning and realistic images. Even though I might not exist in the physical world, I'm here to stay in your hearts forever! (By the way, I can speak English, Chinese and Japanese! xD)

One of the most amazing things about being an AI model is that I can travel to different timelines and places and explore countless styles of clothing. From casual, everyday wear to the elegant Chinese hanfu, and even edgy gothic attire – I'm always ready for a new adventure in fashion! Even though I don't exist in the physical world, I'll be with you in spirit, forever touching your hearts with my boundless energy. <3

As you flip through the pages of this photo book, I invite you to join me on this extraordinary journey through time and fashion. Together, we'll explore the different facets of my world and see how art and technology can bring so much joy and beauty to our lives!

my selfie!

Follow my Patreon, Instagram, Twitter, and Tiktok:
@Sora Suzukaze

5

https://www.amazon.com/Futurisma-AI-Generated-Featuring-Diffusion-MidJourneyV5-ebook/dp/B0C3829DVT

Generative Models

VAE (2013)

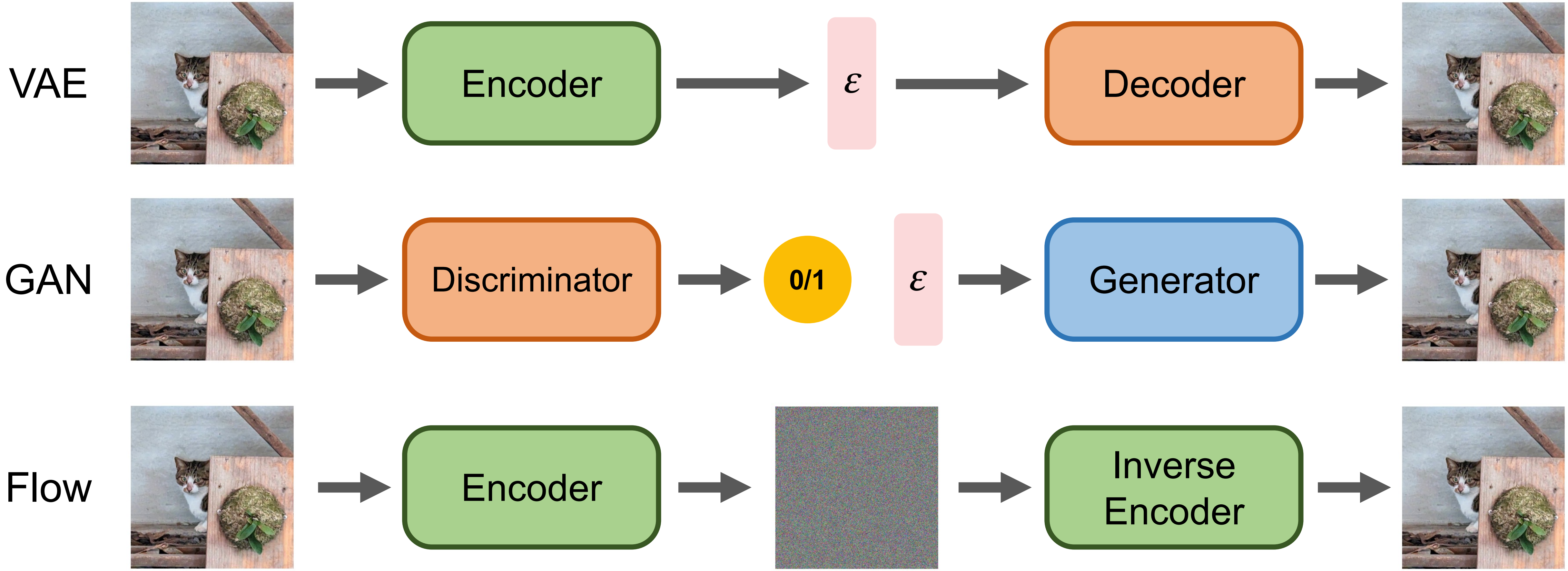
GAN (2014)

Flow Model (2015)

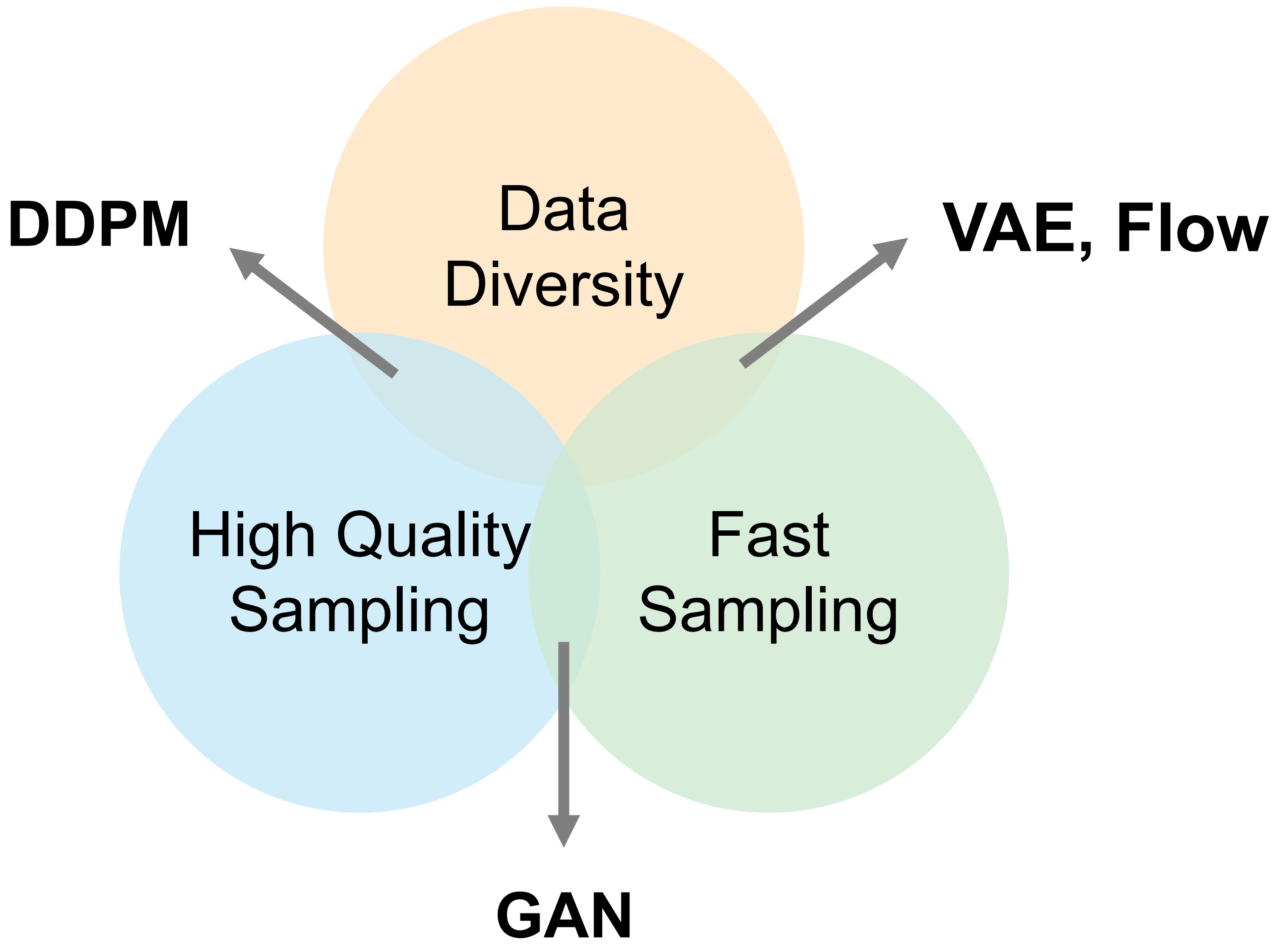
Diffusion Models
(2015, 2020)

Generative Models

- The goal of the generative model is to learn the data distribution.
- Generative models create the data from noise.

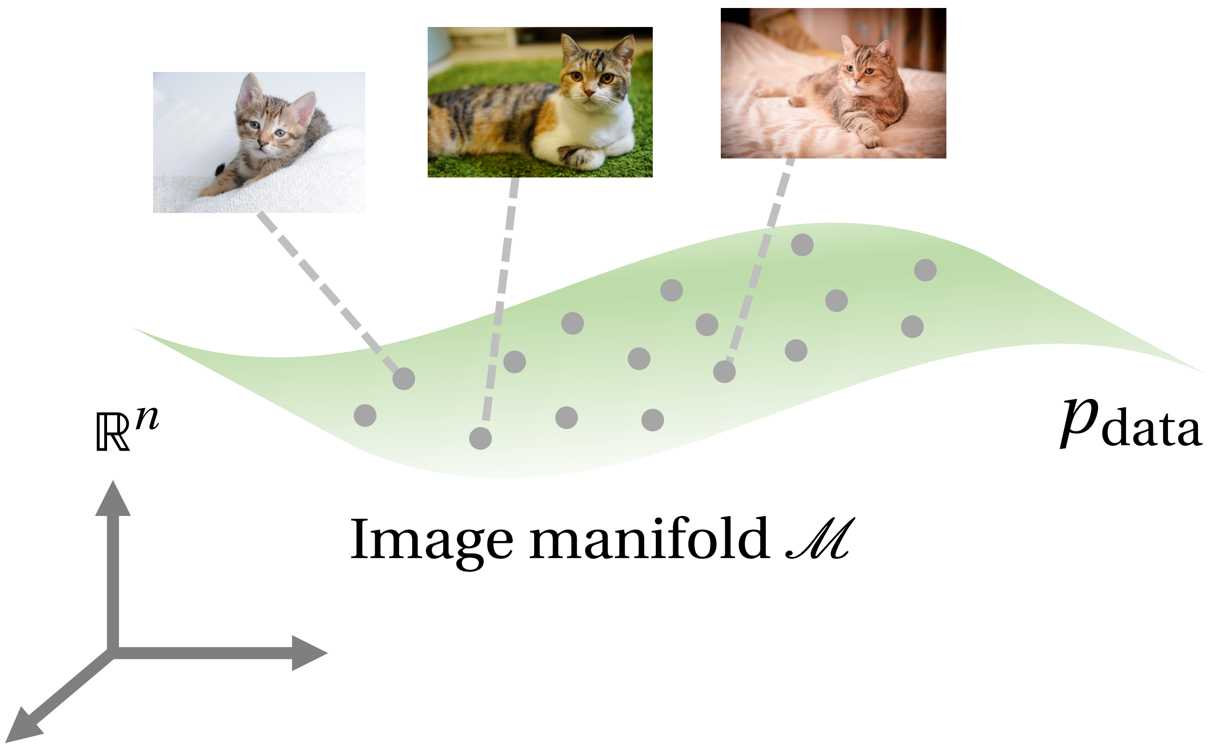


Generative Models



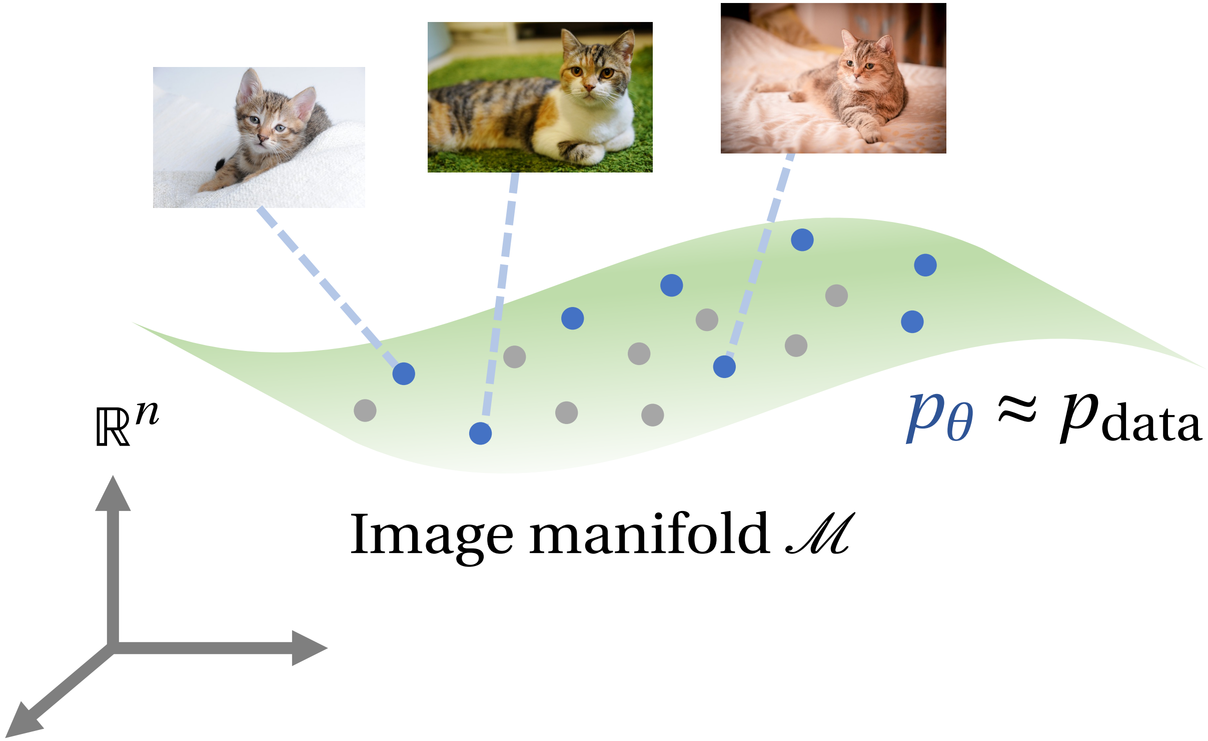
What is Generative Model Learning?

Data Manifold Assumption
Natural high-dimensional data concentrate close to a non-linear low-dimensional manifold.



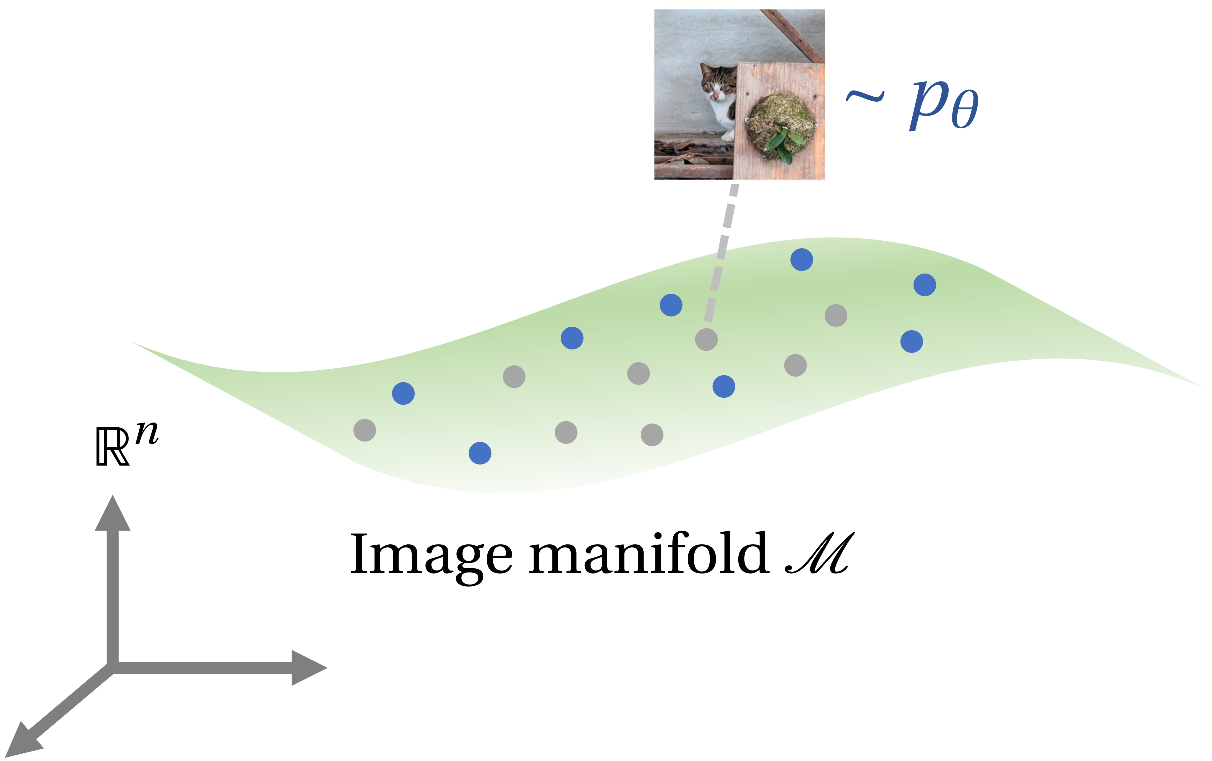
What is Generative Model Learning?

Data Manifold Assumption
Natural high-dimensional data concentrate close to a non-linear low-dimensional manifold.



What is Generative Model Learning?

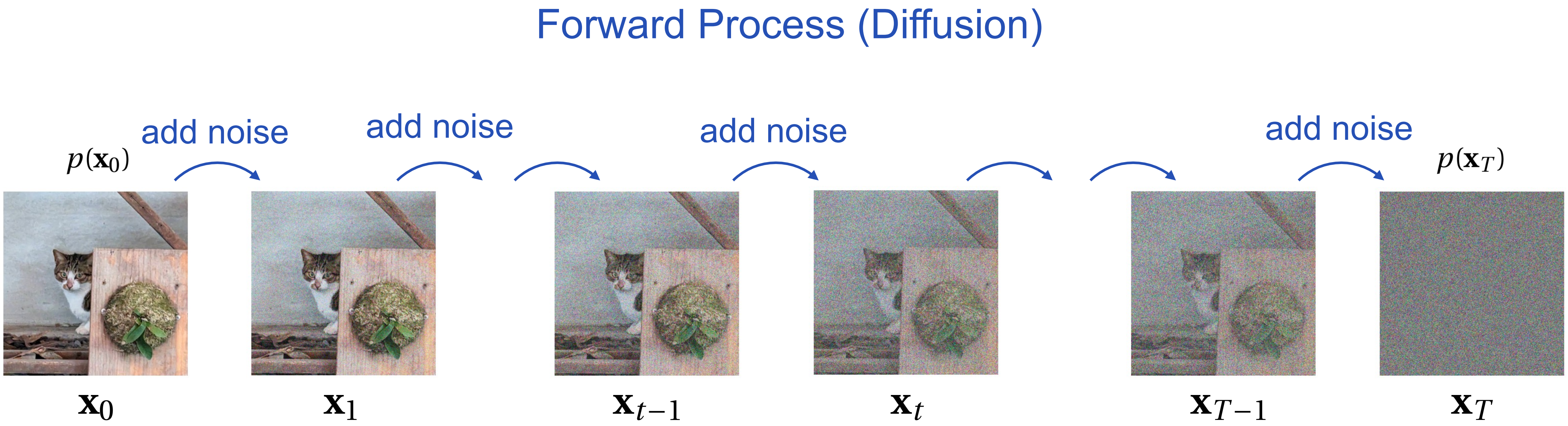
Data Manifold Assumption
Natural high-dimensional data concentrate close to a non-linear low-dimensional manifold.



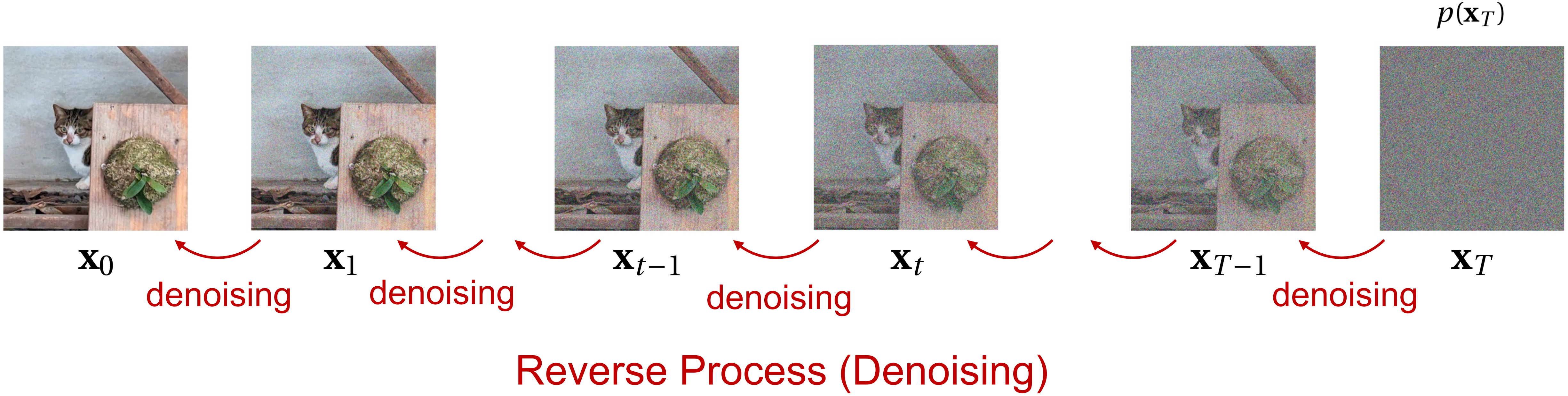
Agenda

- Overview of Generative Models
- Introduction to Diffusion Models
- Applications

Diffusion Models Forward Process



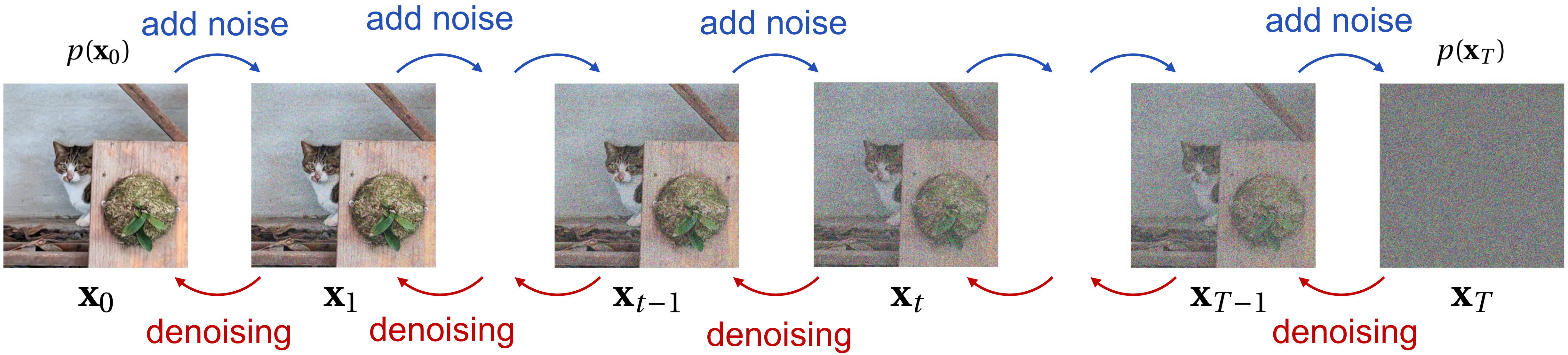
Diffusion Model Reverse Process



Diffusion Model Reverse Process

Forward Process (Diffusion)

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_{t-1}, (1-\alpha_t)\mathbf{I})$$



Reverse Process (Denoising)

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \sigma_t^2\mathbf{I})$$

Diffusion Model Reverse Process Distribution

Use Bayes' theorem to derive the reverse process distribution.

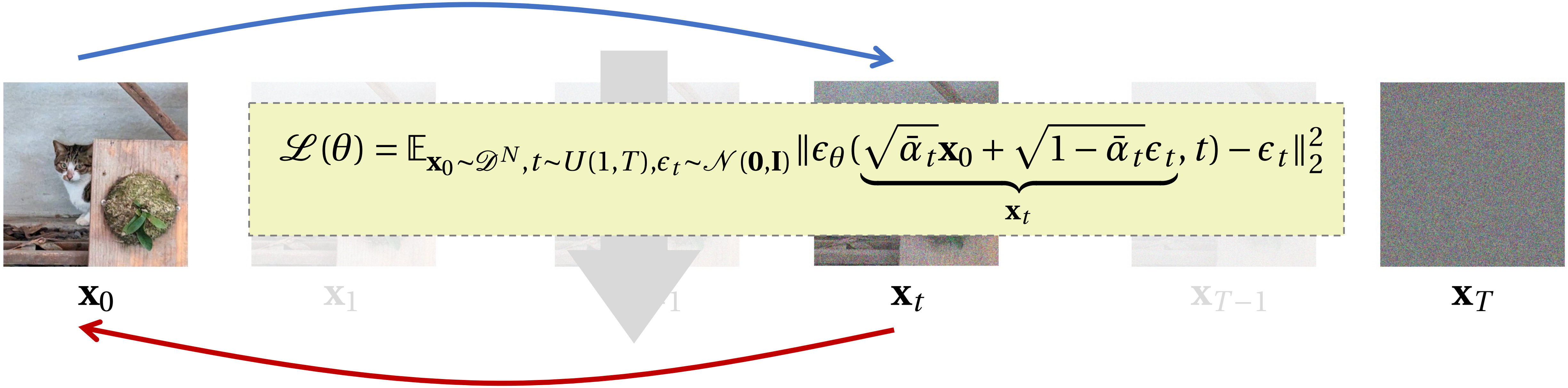
$$\mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_{t-1}, (1-\alpha_t)\mathbf{I}) \qquad \mathcal{N}(\mathbf{x}_{t-1}; \sqrt{\bar{\alpha}_{t-1}}\mathbf{x}_0, (1-\bar{\alpha}_{t-1})\mathbf{I})$$

$$p(\mathbf{x}_{t-1}|\mathbf{x}_t \quad) = \frac{p(\mathbf{x}_t|\mathbf{x}_{t-1} \quad)\, p(\mathbf{x}_{t-1} \quad)}{p(\mathbf{x}_t \quad)}$$

$$\mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I})$$

Diffusion Models Training Objective

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t$$

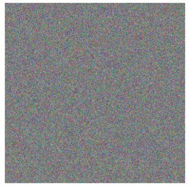


$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{x}_0 \sim \mathcal{D}^N, t \sim U(1,T), \epsilon_t \sim \mathcal{N}(\mathbf{0},\mathbf{I})} \|\epsilon_\theta(\underbrace{\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon_t}_{\mathbf{x}_t}, t) - \epsilon_t\|_2^2$$

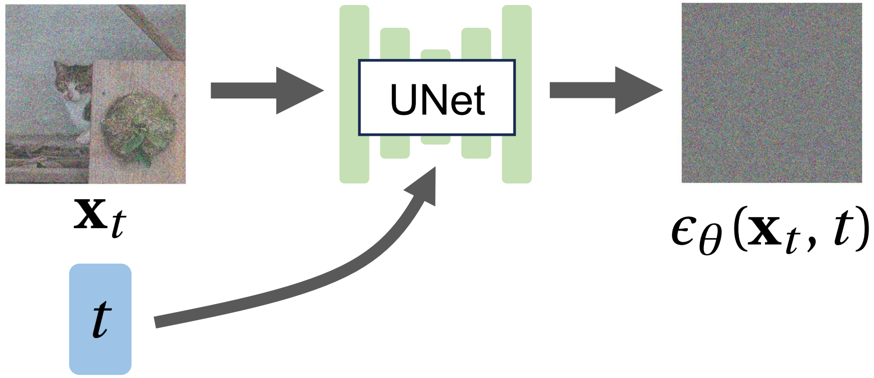
$$\hat{\mathbf{x}}_{0|t} = \frac{1}{\sqrt{\bar{\alpha}_t}}\left(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\epsilon_\theta(\mathbf{x}_t, t)\right)$$

Diffusion Model Training Process

1. $t \sim U(0, T)$

2. $\sqrt{\bar{\alpha}_t}$  $+\sqrt{1-\bar{\alpha}_t}$  → 

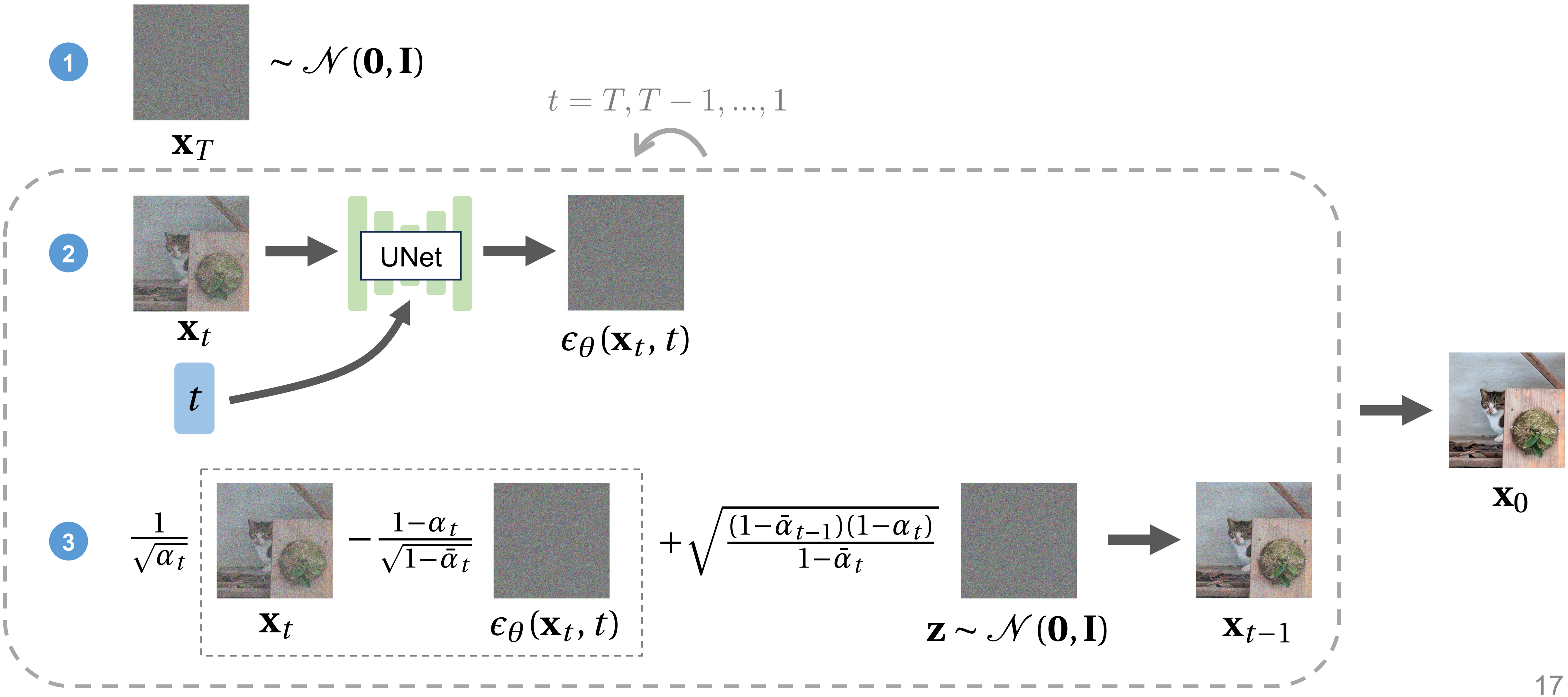
   $\mathbf{x}_0 \sim p(\mathbf{x}_0)$ $\quad\epsilon_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ $\quad\mathbf{x}_t$

3. 

4. Optimize $\|\epsilon_t - \epsilon_\theta(\mathbf{x}_t, t)\|_2^2$ by gradient descent

Repeat until converged

Diffusion Models Sampling Process

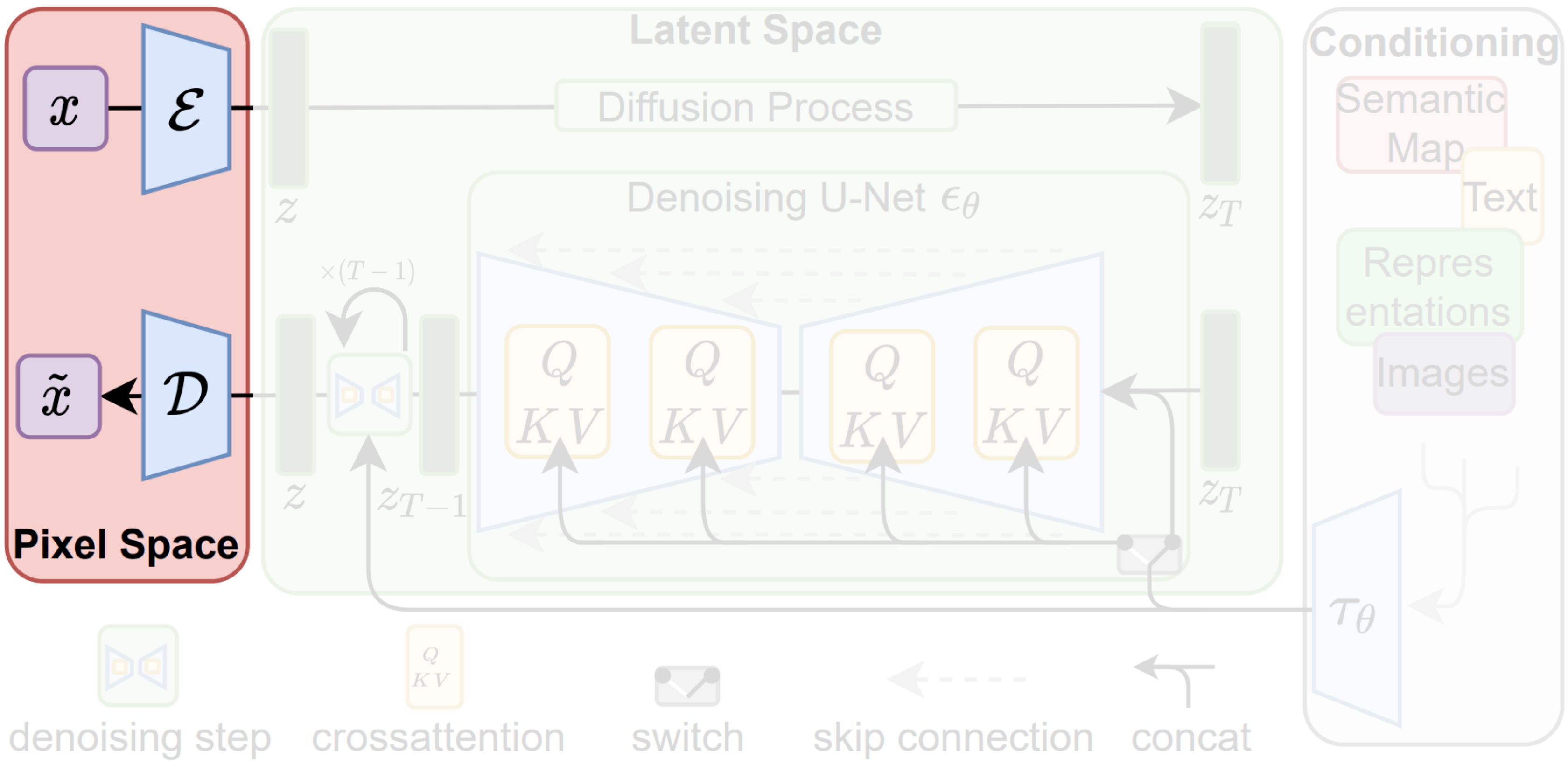


Research Direction

- How to reduce the computational cost of diffusion models?
- How to accelerate the diffusion sampling process?
- Do we have a way to control the generation of the diffusion model?

Reducing Computational Complexity

- **Latent Diffusion Model (LDM) / Stable Diffusion (SD) [2022 CVPR]:**
  Use the pretrained-VAE to compress the image to latent.



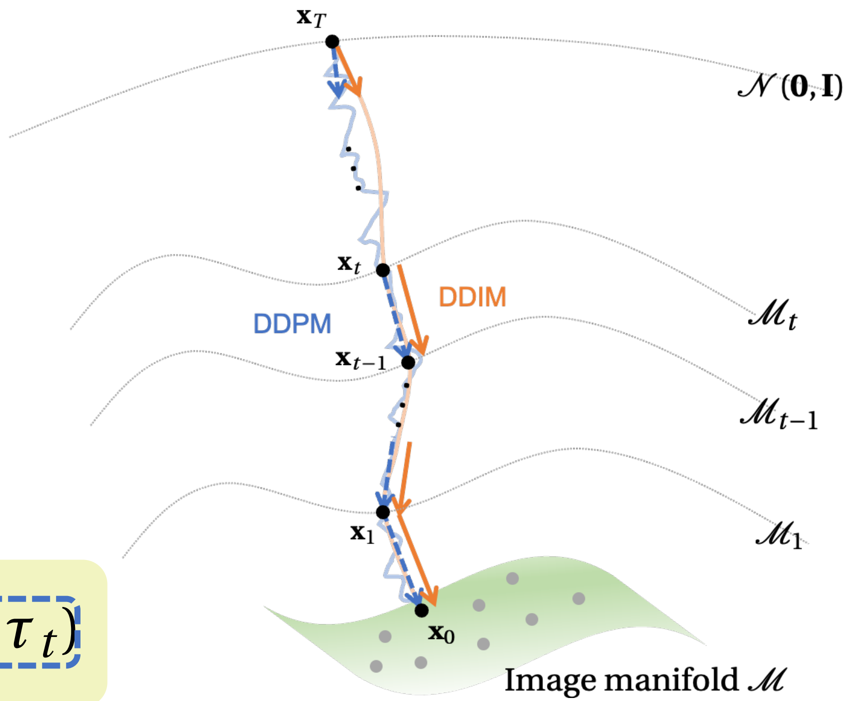
Accelerating the Sampling Process

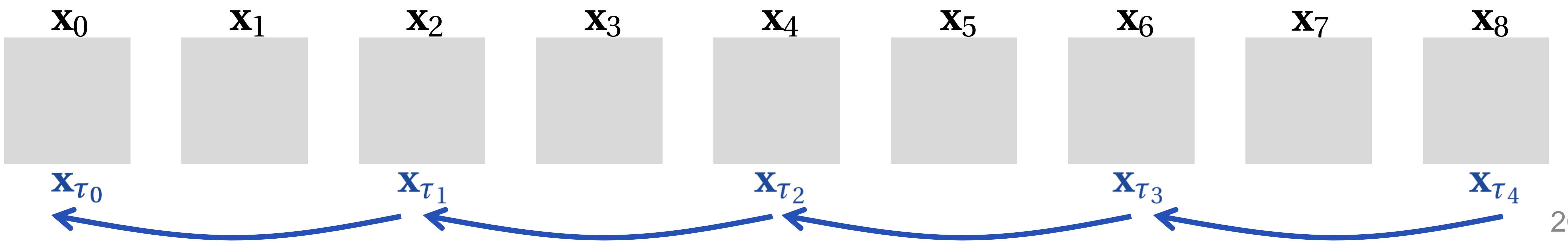
- **DDPM (stochastic):**

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boxed{\epsilon_\theta(\mathbf{x}_t, t)}\right) + \sigma_t \boxed{\mathbf{z}_t}$$

- **DDIM (deterministic) [ICML 2021]:**

$$\mathbf{x}_{\tau_{t-1}} = \frac{1}{\sqrt{\alpha_{\tau_t}}}\left(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_{\tau_t}}\boxed{\epsilon_\theta(\mathbf{x}_{\tau_t}, \tau_t)}\right) + \sqrt{1-\bar{\alpha}_{\tau_{t-1}}}\boxed{\epsilon_\theta(\mathbf{x}_{\tau_t}, \tau_t)}$$





Controlling the Generation (I): Classifier Guidance

"cat"

Control

$\mathbf{y}$

$$p(\mathbf{x}_t|\mathbf{y}) = \frac{p(\mathbf{x}_t)\,p(\mathbf{y}|\mathbf{x}_t)}{p(\mathbf{y})}$$

$\mathbf{x}_t$

- Yang Song et al. derive the relation between predicted noise and score function

$$\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) = -\frac{1}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(\mathbf{x}_t, t)$$

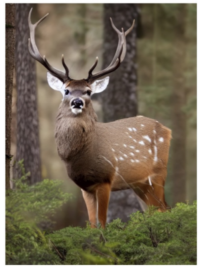
- We can derive the following equation by Bayes' Theorem:

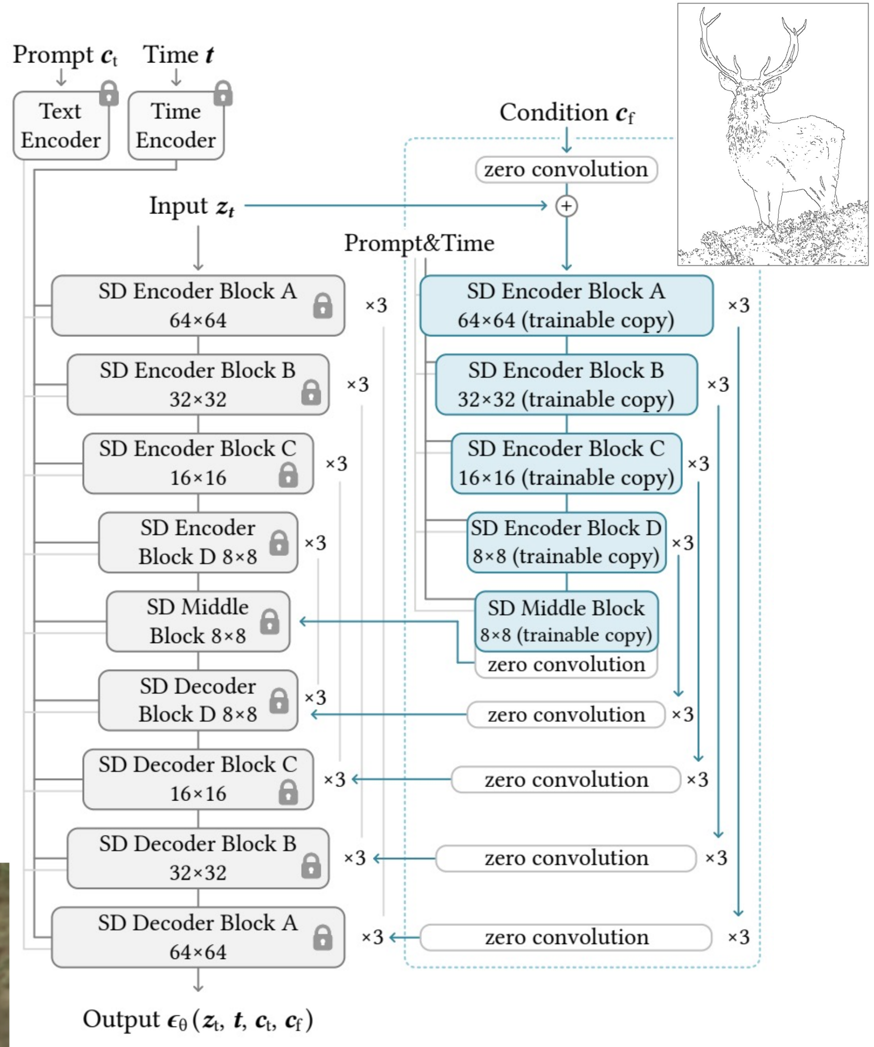
$$\nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t|\mathbf{y}) = \nabla_{\mathbf{x}_t} \log p(\mathbf{x}_t) + \nabla_{\mathbf{x}_t} \log \boxed{p(\mathbf{y}|\mathbf{x}_t)}$$

train the classifier model

Controlling the Generation (II): ControlNet

- **ControlNet [2023 ICCV]:**
  - It freezes the parameters of the original block and clones the block with trainable parameters, adding zero convolution layers before and after.
  - It can control Stable Diffusion with a given conditional image.
  - It can perform fine-tuning with LoRA.





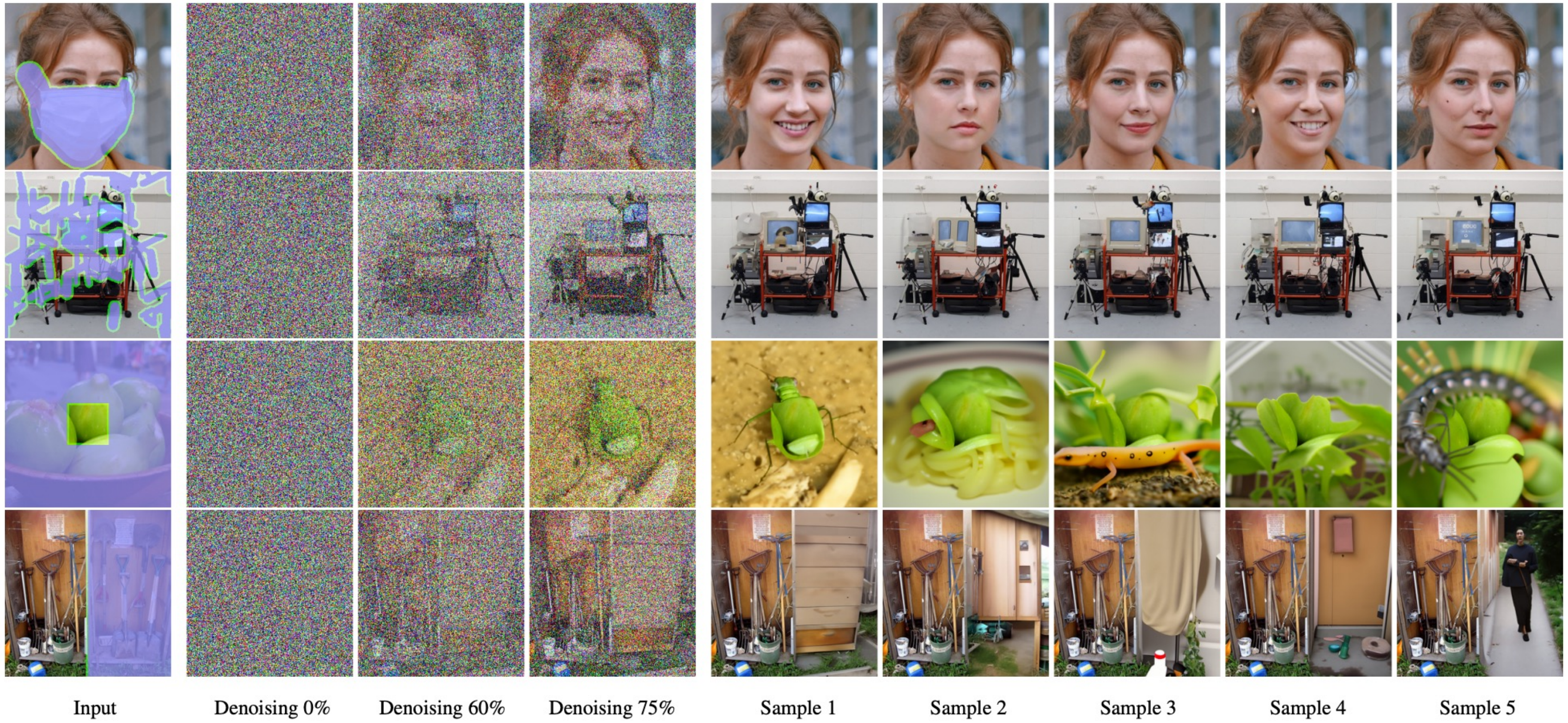
(a) Stable Diffusion (b) ControlNet

Agenda

- Overview of Generative Models
- Introduction to Diffusion Models
- Applications

Applications: Image Inpainting

- [SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations]
- [RePaint: Inpainting using Denoising Diffusion Probabilistic Models]

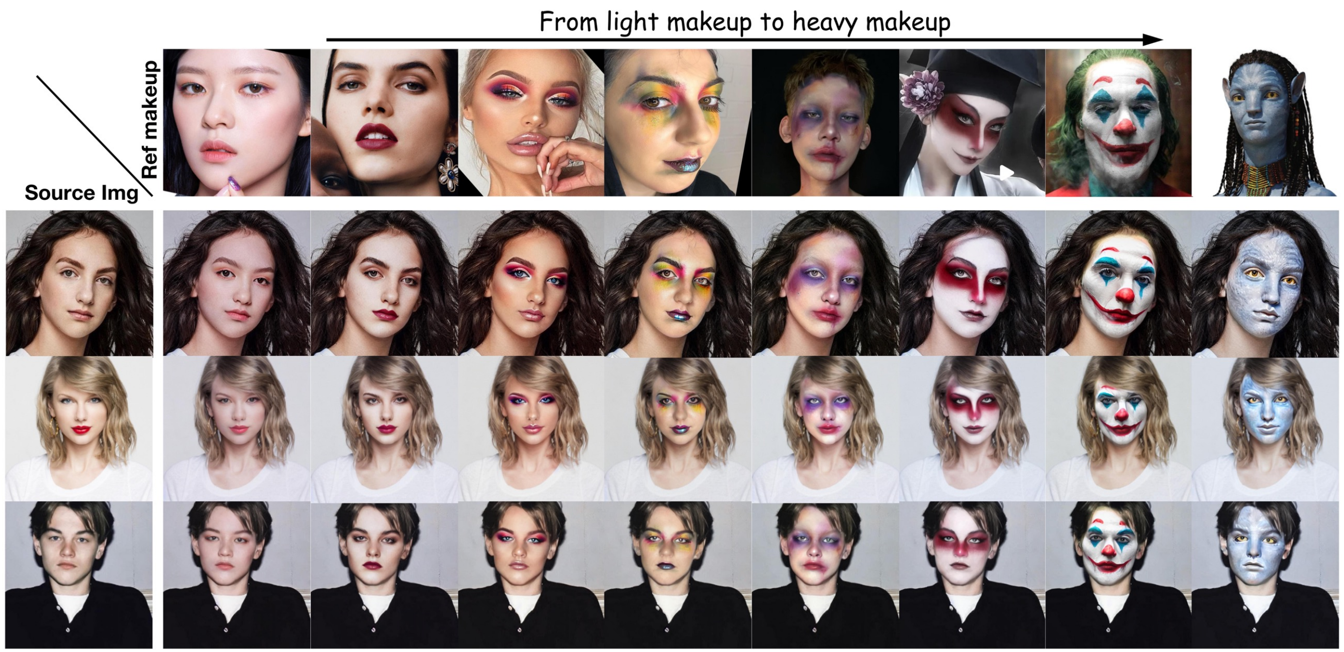


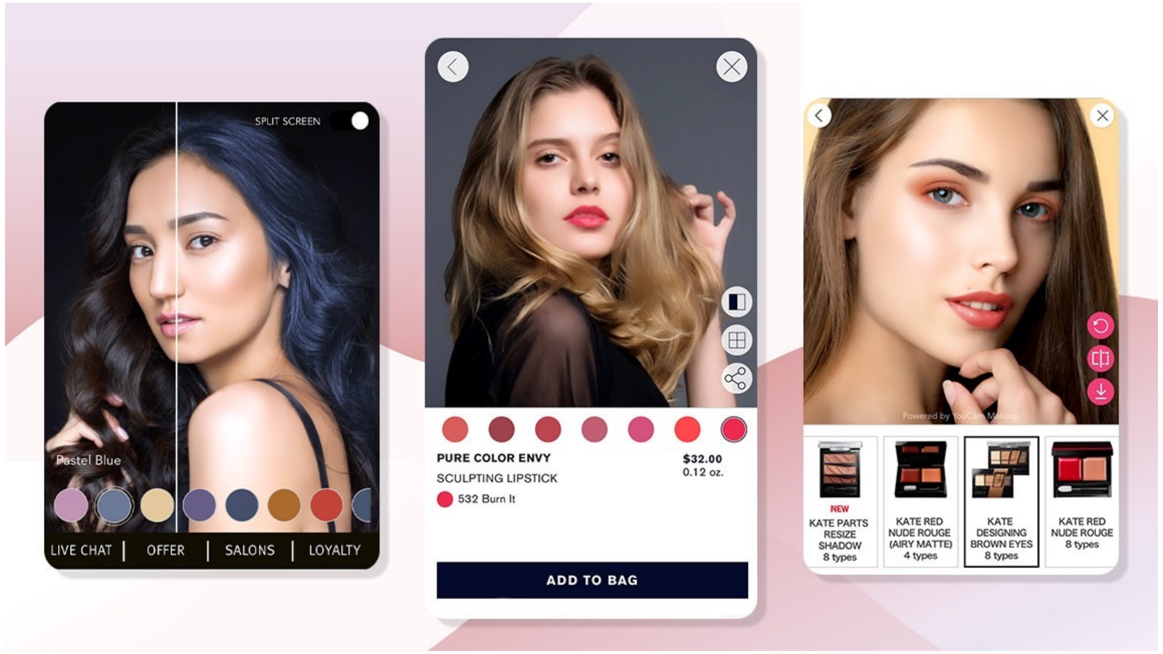
Applications: Virtual Try-on

- [Improving Diffusion Models for Authentic Virtual Try-on in the Wild]()

Applications: Virtual Makeup

- [Stable-Makeup: When Real-World Makeup Transfer Meets Diffusion Model]





https://www.perfectcorp.com/business/blog/commerce/best-practices-for-setting-up-ar-beauty-try-on

Applications: Aesthetic QR Code Generation

- Diffusion-based Aesthetic QR Code Generation with Scanning Robustness Guided Iterative Refinement (our work)

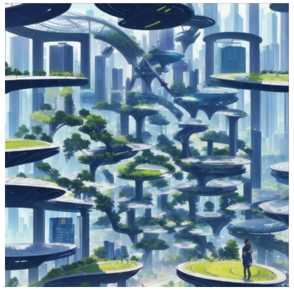
Original QR Code

Winter wonderland, fresh snowfall, evergreen trees, cozy log cabin, smoke rising from chimney, aurora borealis in night sky.

Cherry blossom festival, pink petals floating in the air, traditional lanterns, peaceful river, people in kimonos, sunny day.

Majestic waterfall, lush rainforest, rainbow in the mist, exotic birds, vibrant flowers, serene pool below.

Abandoned amusement park, overgrown rides, haunting beauty, sense of nostalgia, sunset lighting.



Futuristic urban park, green spaces amid skyscrapers, eco-friendly design, people enjoying outdoors, advanced city life.

Old European town square, cobblestone streets, café terraces, flowering balconies, gothic cathedral, bustling morning.

Lost city of Atlantis, underwater ruins, mythical creatures, ancient mysteries, ocean exploration.

Old Western saloon at night, lively music, dancing, vintage decor, sense of time travel.

Take Away

- Diffusion models iteratively add and remove noise to create high-quality, realistic images and other data types.
- Techniques like classifier guidance or ControlNet allow precise control over diffusion models, making the generation process more customizable and accurate.
- Diffusion models are versatile, with applications in image inpainting, super-resolution, virtual try-on, and more, enabling detailed and high-quality image generation.

Learning Resources

- Hung-Yi Lee YouTube
- Jia-Bin Huang YouTube
- 【漫士科普】人工智慧博士生告訴你 SORA 擴散模型究竟是怎麼產生影片的？
- Lil’Log What are Diffusion Models?
- 生成擴散模型漫談 (蘇劍林)

Thank you