Deep Learning for Computer Vision

113-1/Fall 2024

https://cool.ntu.edu.tw/courses/41702 (NTU COOL)

http://vllab.ee.ntu.edu.tw/dlcv.html (Public website)

Yu-Chiang Frank Wang 王鈺強, Professor

Dept. Electrical Engineering, National Taiwan University

2024/09/24

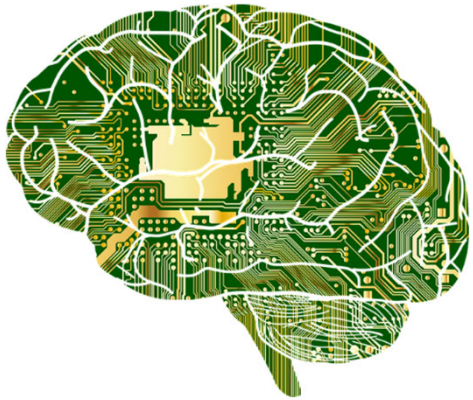
Slightly updated syllabus

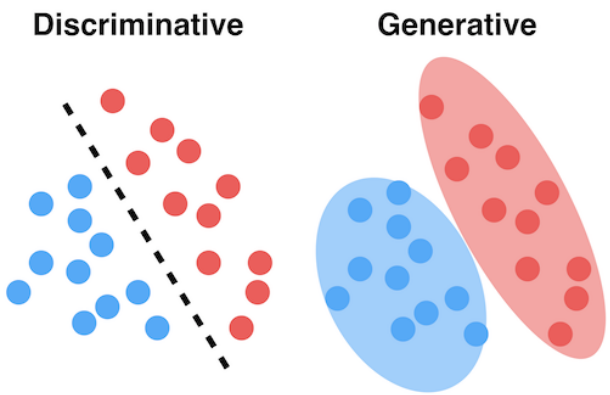
Week	Date	Topic	Course Materials	Remarks
1	09/03	Course Logistics & Registration; Intro to Neural Nets	W1-1 W1-2	
2	09/10	Convolutional Neural Networks & Image Segmentation	W2	**HW #1 out**
3	09/17	**No class**		Mid-Autumn Festival
4	09/24	Generative Models (I) - AE, VAE, & Diffusion Model (I)		**HW #1 due**
5	10/01	**Guest Lecture:** Dr. Jun-Cheng Chen, Academia Sinica		ECCV week
6	10/8	Generative Models (II) - Diffusion Model (II), GAN		**HW # 2 out**
7	10/15	Recurrent Neural Networks & Transformer		
8	10/22	Transformer; Vision & Language Models		
9	10/29	Vision & Language Models; Multi-Modal Learning		**HW #2 due; HW #3 out**
10	11/05	Parameter-Efficient Finetuning; Unlearning, Debiasing, and Interoperability		
11	11/12	**Guest Lecture:** Linda Huang, Senior Dir., GeValyn Associates		
12	11/19	3D Vision		**HW #3 due; HW #4 out**
13	11/26	Object Detection		**Final Project Announcement**
14	12/03	**Guest Lecture:** Prof. Ming-Ching Chang, SUNY, Albany; Federated Learning and advanced topics in DLCV		**HW #4 due**
15	12/10	**Progress Check for Final Projects**		NeurIPS week
17	12/25 Wed	**Final Project Presentation**		

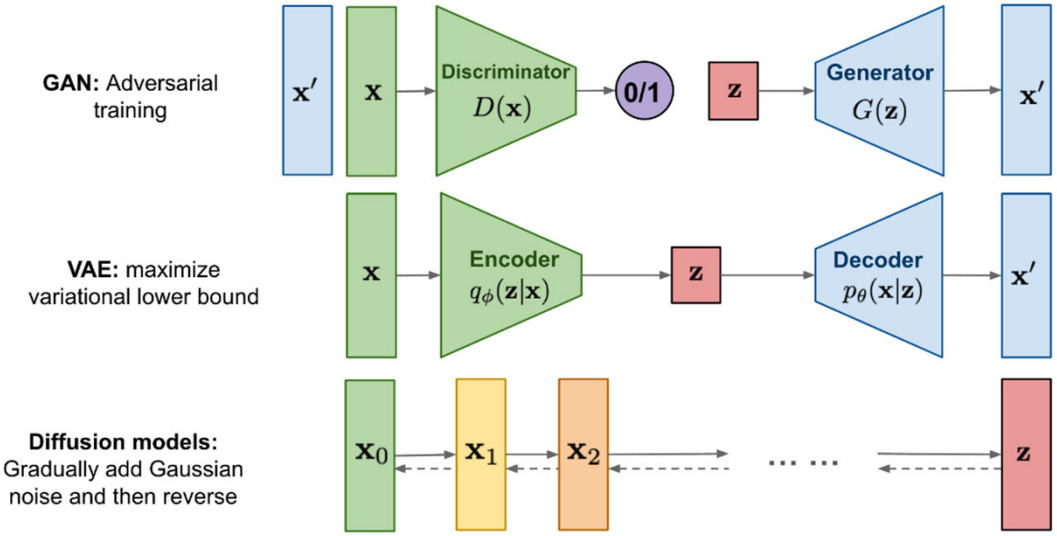
Will finalize in early Oct.

What's to Be Covered Today…

- Generative Models
  - Autoencoder
  - Variational Autoencoder
  - Diffusion Model
  - Generative Adversarial Network (next lecture)
- Oct. 1st, Tue.
  - Guest lecture, Dr. Jung-Cheng Chen, Academia Sinica

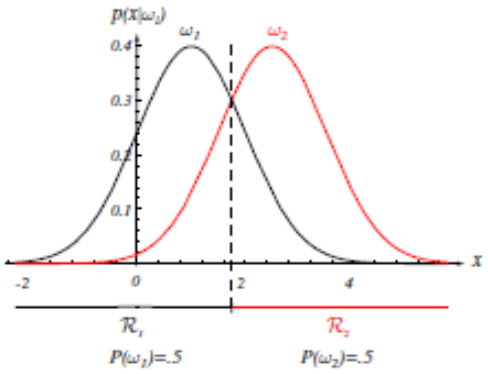




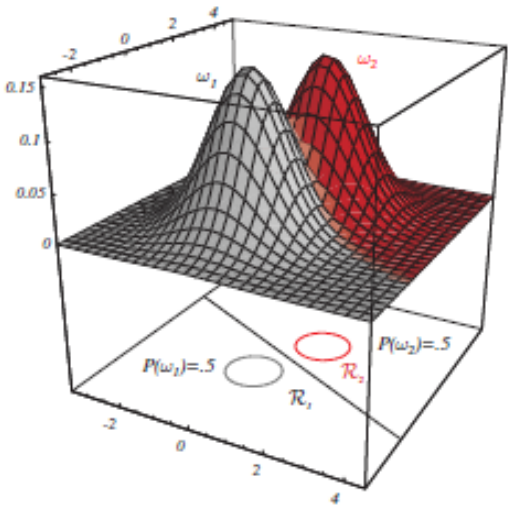


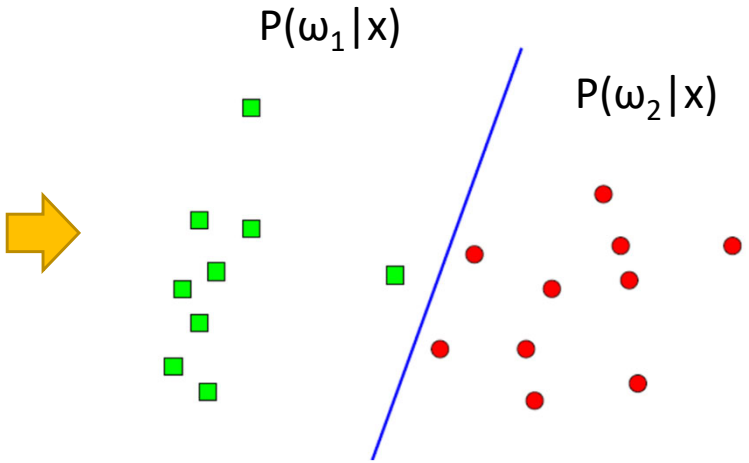
Discriminative vs. Generative Models

- Discriminative Models
  - Model posteriors P(ω|x) from likelihoods P(x|ω)
    where x is the input data, and ω indicates the class of interest
  - Example (posterior)



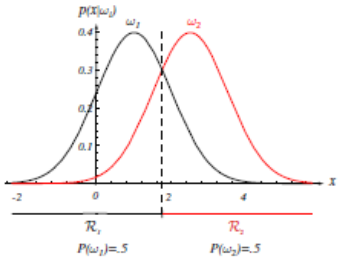
or

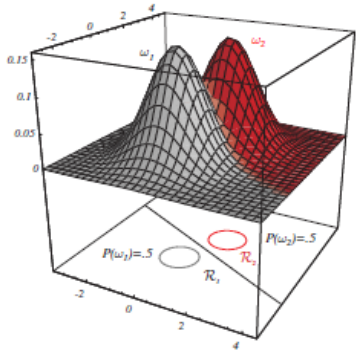




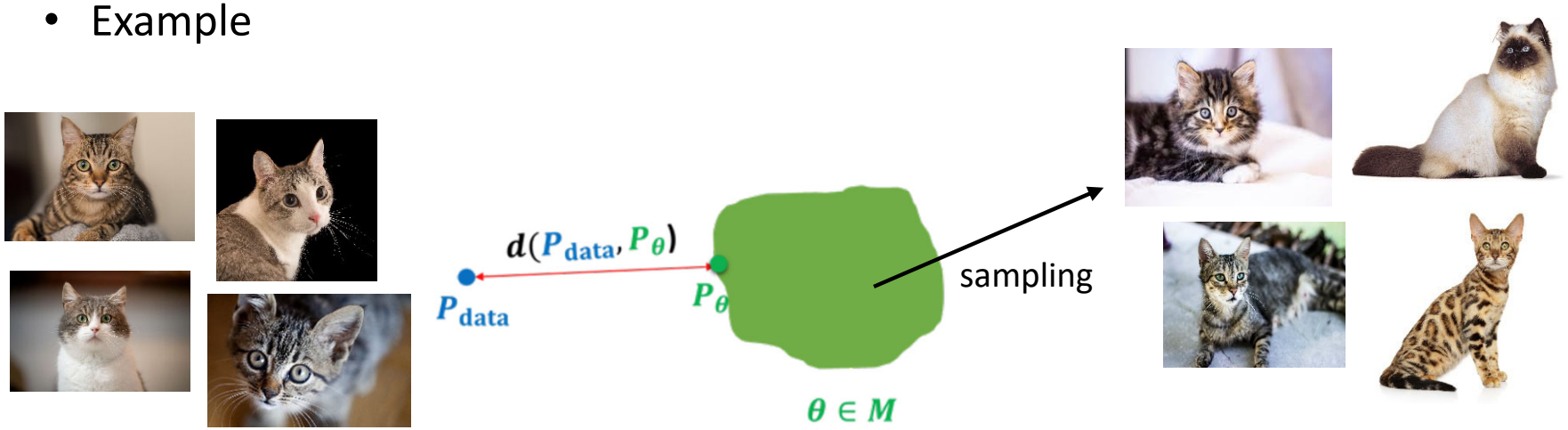
Discriminative vs. Generative Models (cont'd)

- Generative Models
  - Model likelihoods P(x|ω) with priors P(ω) (i.e., modeling P(x|ω) P(ω)) where x is the input data, and ω indicates the class of interest





  - Example



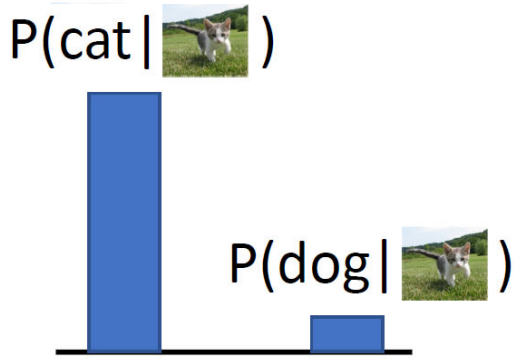
$x_i \sim p_{data}$

$\hat{x}_i \sim p_\theta$

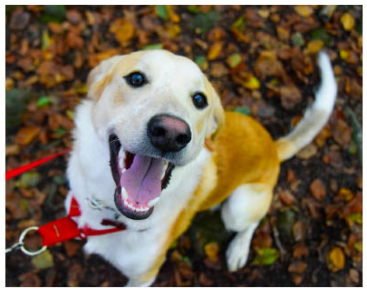
Image credit: CS886 at UWaterloo

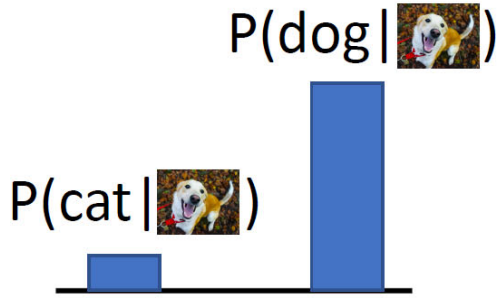
Discriminative vs. Generative Models (cont'd)

Discriminative Model: Learn a probability distribution p(y|x)



Generative Model: Learn a probability distribution p(x)





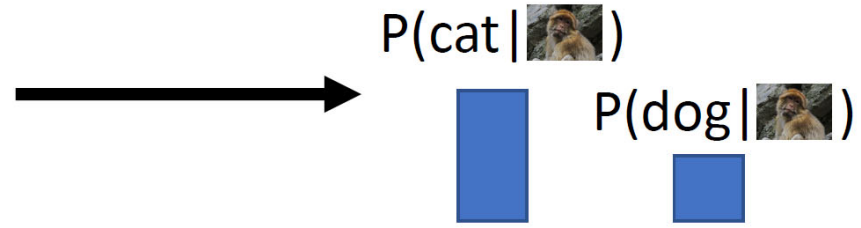
Conditional Generative Model: Learn p(x|y)

Discriminative model: the possible labels for each input "compete" for probability mass. But no competition between **images**

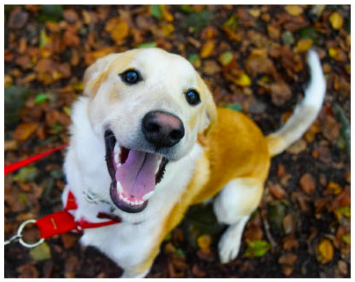
Slide credit: UMich EECS 498-007

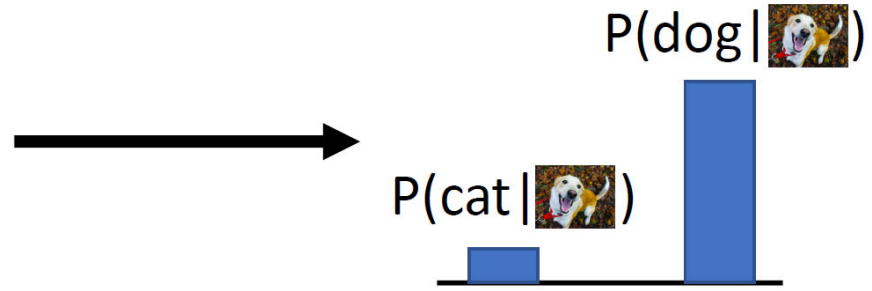
Discriminative vs. Generative Models (cont'd)

Discriminative Model: Learn a probability distribution p(y|x)



Generative Model: Learn a probability distribution p(x)





Conditional Generative Model: Learn p(x|y)

Discriminative model: No way for the model to handle unreasonable inputs; it must give label distributions for all images

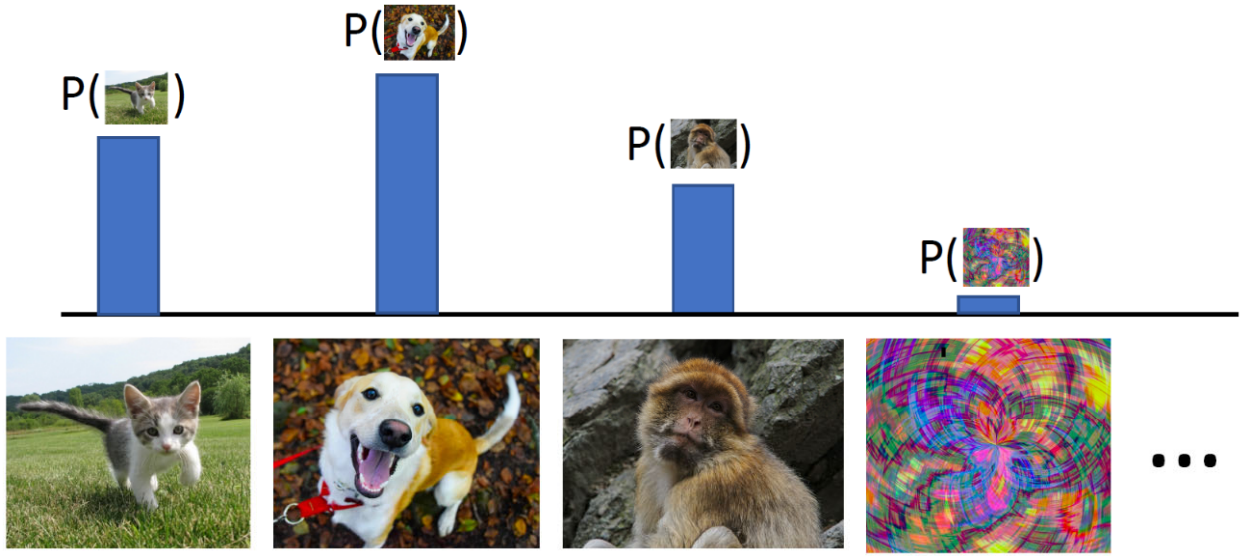
Slide credit: UMich EECS 498-007

Discriminative vs. Generative Models (cont'd)

Discriminative Model: Learn a probability distribution p(y|x)

Generative Model: Learn a probability distribution p(x)

Conditional Generative Model: Learn p(x|y)



Generative model: All possible images compete with each other for probability mass

Model can "reject" unreasonable inputs by assigning them small values

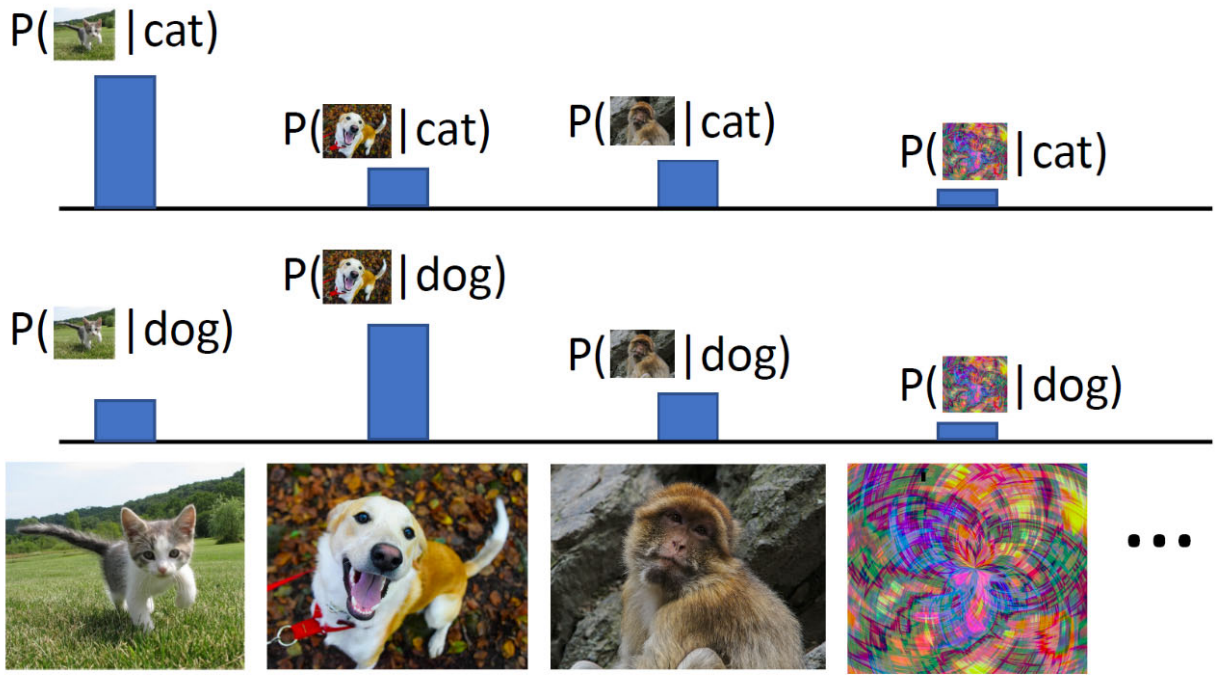
Slide credit: UMich EECS 498-007

Discriminative vs. Generative Models (cont'd)

Discriminative Model: Learn a probability distribution p(y|x)

Generative Model: Learn a probability distribution p(x)

Conditional Generative Model: Learn p(x|y)



Conditional Generative Model: Each possible label induces a competition among all images

Discriminative vs. Generative Models (cont'd)

Discriminative Model: Learn a probability distribution p(y|x)

Generative Model: Learn a probability distribution p(x)

Conditional Generative Model: Learn p(x|y)

Recall **Bayes' Rule:**

$$P(x \mid y) = \frac{P(y \mid x)}{P(y)} P(x)$$

Conditional Generative Model

Discriminative Model

Prior over labels

(Unconditional) Generative Model

We can build a conditional generative model from other components!

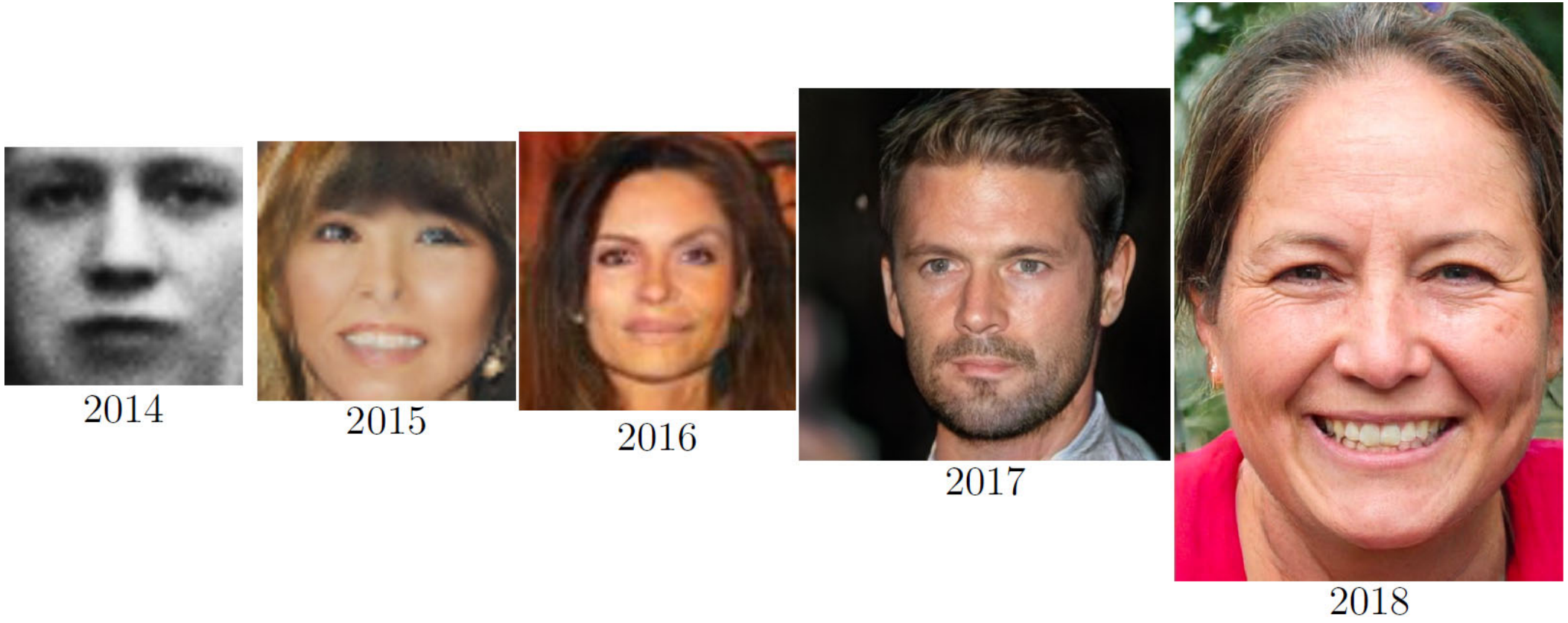
Additional Remarks

- Discriminative Models
  - Goal: Learn a (posterior) probability distribution p(y|x)
  - Task: Assign labels to each instance x (e.g., classification, regression, etc.)
  - Supervised learning

- Generative Models
  - Goal: Learn a probability distribution p(x)
  - Task: Data representation, generation, detect outliers, etc.
  - (Mostly) unsupervised learning

What Have Been Done Using Deep Generative Models?

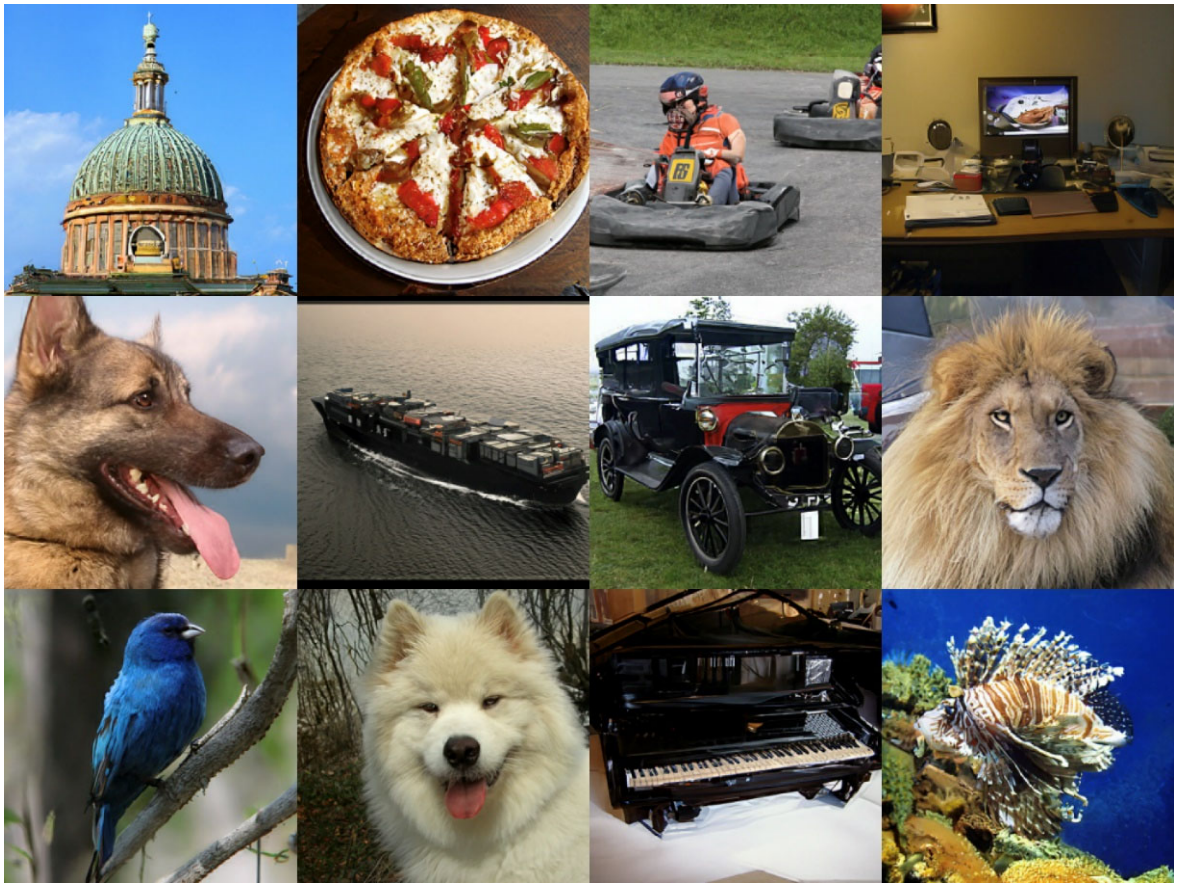
- 5+ years of progress on synthesizing face images



Slide credit: I. Goodfellow

What Have Been Done Using Deep Generative Models?

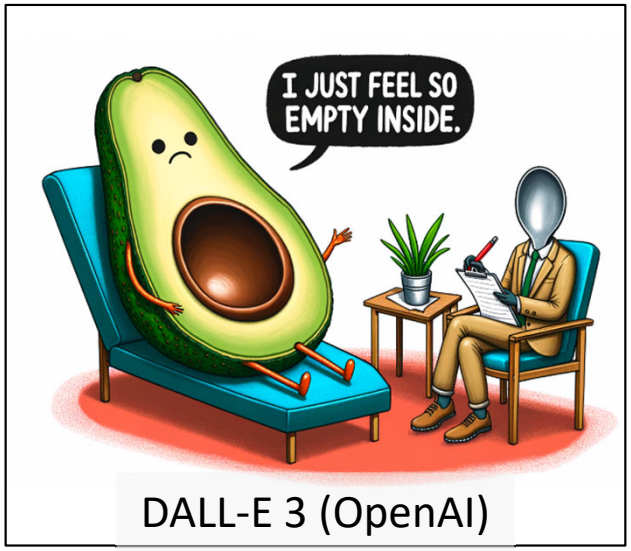
- Progress on synthesizing images (ImageNet)



Super-Resolution via Repeated Refinements (SR3) by
Class Diffusion Models (Google, 2021)

Slide credit: I. Goodfellow & SR3

What Have Been Done Using Deep Generative Models?



DALL-E 3 (OpenAI)


Firefly (Adobe)


Midjourney


SDXL (Stability.ai)

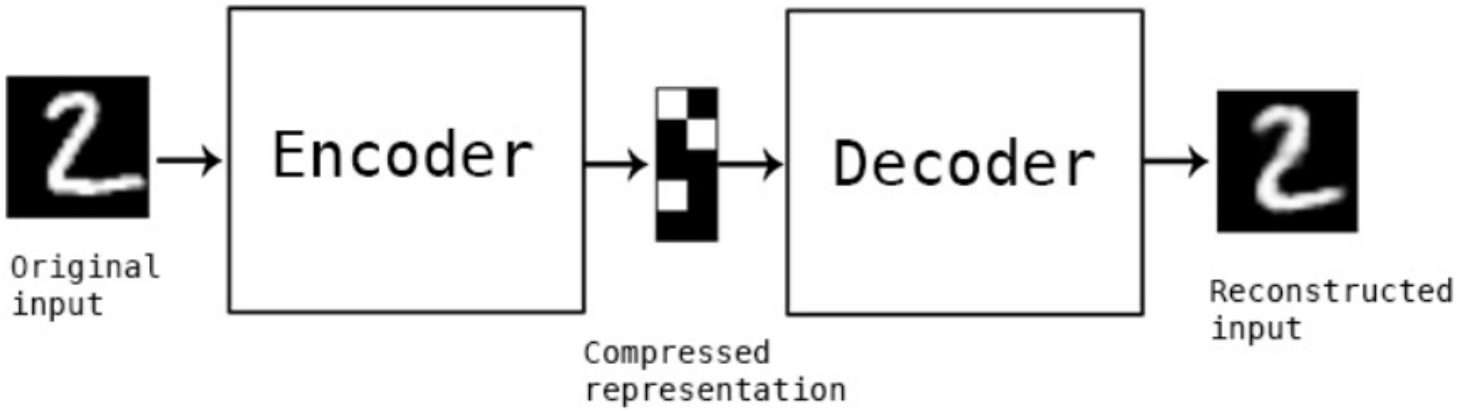
Slide credit: J.-B. Huang

ChatGPT can read, see, hear, and speak…

https://openai.com/blog/chatgpt-can-now-see-hear-and-speak

How GenAI Works?
Let's Start from Autoencoder…

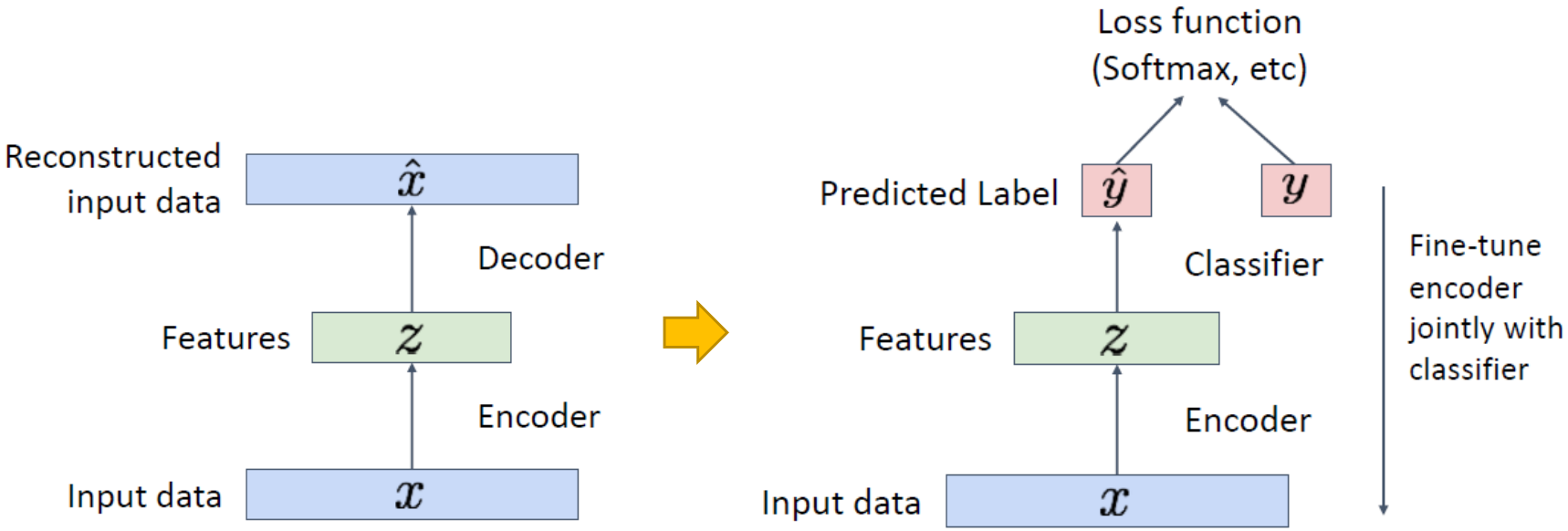
- Autoencoder (AE)
  - Autoencoding = encoding itself with recovery purposes
  - In other words, encode/decode data with reconstruction guarantees
  - Latent variables/features as deep representations
  - Example objective/loss function at output:
    - L2 norm between input and output, i.e.,



Slide credit: W. Chiu

AE for Learning Latent Variables/Representations

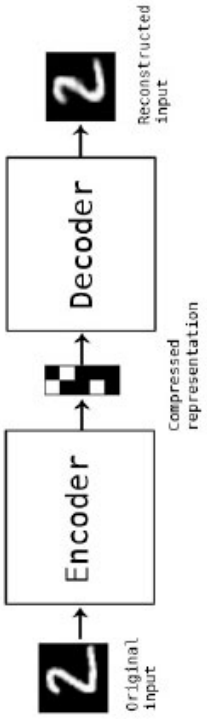
- Why/when AE may be favorable?
  i.e., unsupervised learning for latent representation…
  Any application comes to your mind??
- Train autoencoder (AE) for downstream tasks
  - Train AE with reconstruction guarantees
  - Keep encoder (and the derived features) for downstream tasks (e.g., classification)
  - Thus, a trained encoder can be applied to initialize a supervised model

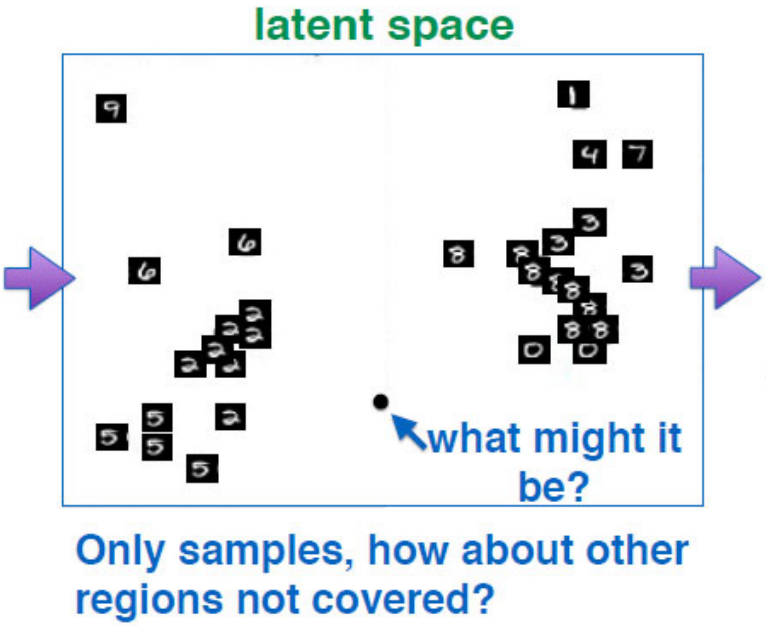


Slide credit: UMich EECS 498-007

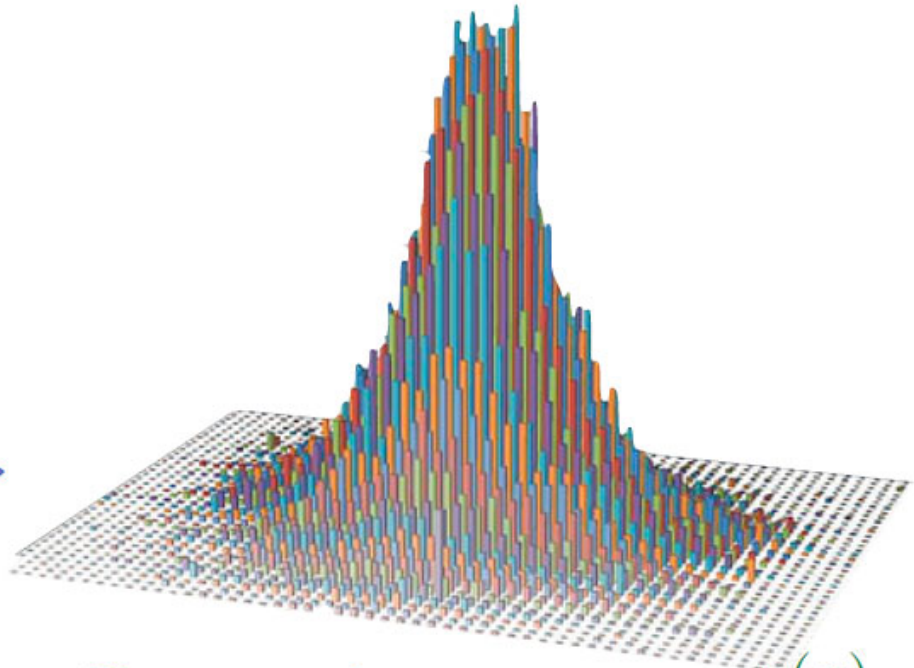
Take a Deep Look to Discover Latent Variables/Representations (cont'd)

- Limitation/concern of autoencoder?





Only samples, how about other regions not covered?

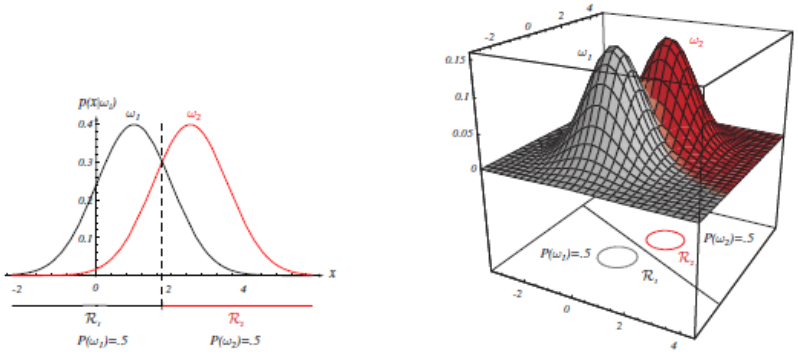


We wanna to have *"distribution"* $p(z)$ where we can sample from any location

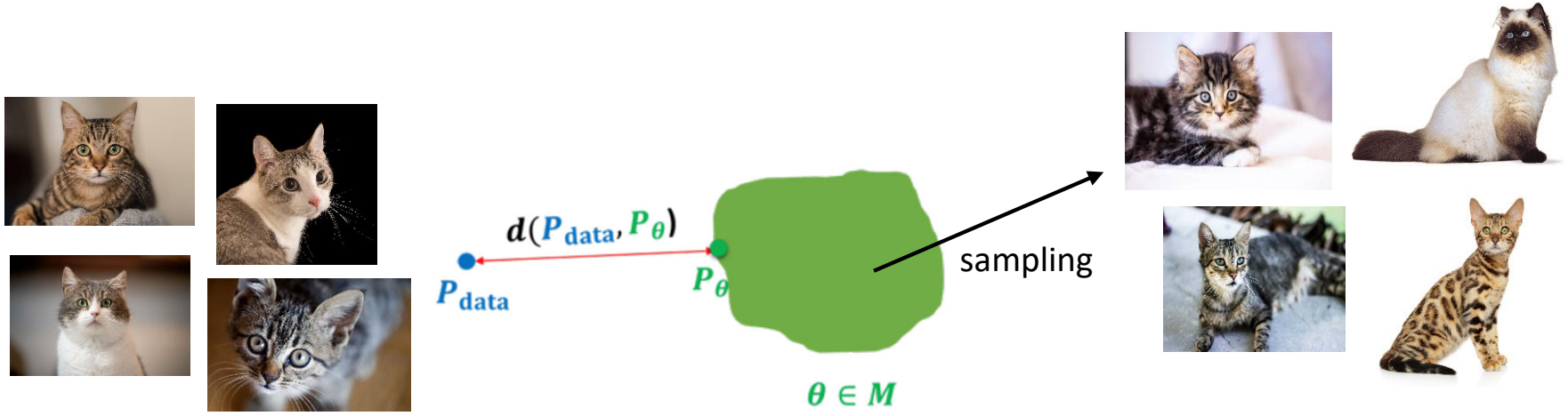
Slide credit: W. Chiu

Revisit of Generative Models

- Generative Models
  - Model likelihoods P(x|ω) with priors P(ω) (i.e., modeling P(x|ω) P(ω)) where x is the input data, and ω indicates the class of interest



  - Take P(x) for an example (i.e., unconditional generation)

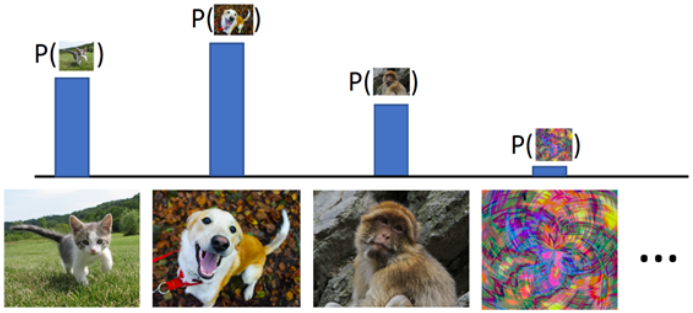


$x_i \sim p_{data}$

$\hat{x}_i \sim p_\theta$

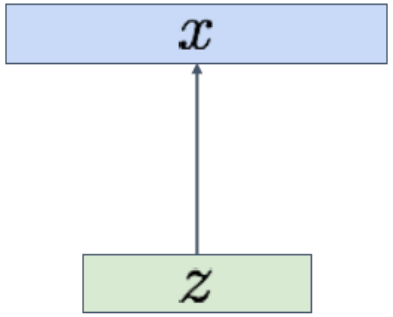
Variational Autoencoder

- Probabilistic Spin on AE
  - Learn latent feature ***z*** from raw input data $x_i \sim p_{data}$
  - Sample from the latent space (via) to generate data
  - For simplicity, assume simple prior p(z) (e.g., Gaussian)

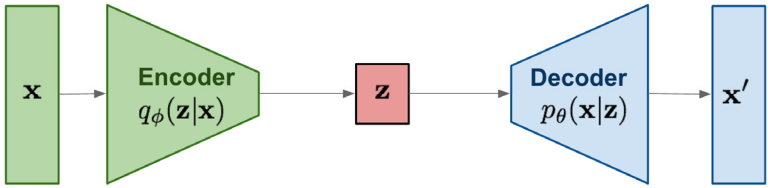


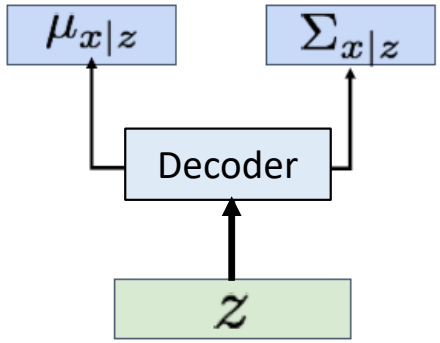
Sample from conditional $p_\theta\ (x \mid z^{(i)})$

Sample z from prior $p_\theta\ (z)$



- Learn p(x|z) via a NN as a (probabilistic) **decoder**



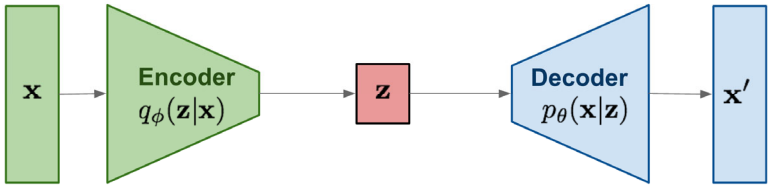


Decoder inputs z, outputs mean $\mu_{x|z}$ and (diagonal) covariance $\sum_{x|z}$

Sample x from Gaussian with mean $\mu_{x|z}$ and (diagonal) covariance $\sum_{x|z}$

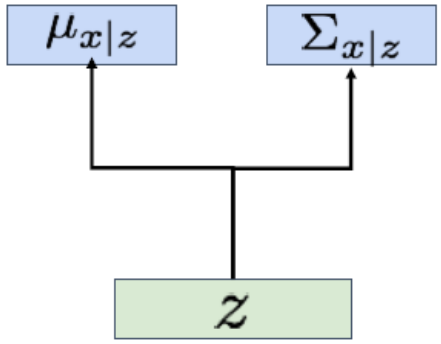
Slide credit: UMich EECS 498-007

Variational Autoencoder (cont'd)



- Remarks
  - Training objective: maximum likelihood of data p(x)
  - Note that we can't possibly observe all latents z & need to marginalize it:

$$p_\theta(x) = \int p_\theta(x,z)dz = \int p_\theta(x|z)p_\theta(z)dz$$



  - We can compute $p_\theta(x|z)$ with the decoder module, and we assume Gaussian prior for $p_\theta(z)$
  - Still, can't integrate over all possible z!
  - What else can we do? Recall that we have Bayes' rule:

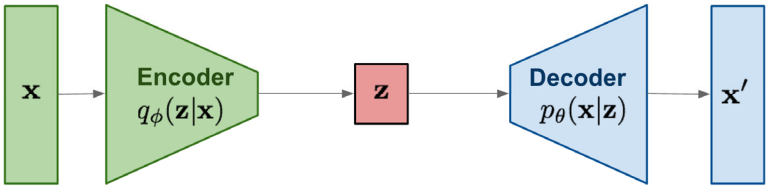
$$p_\theta(x) = \frac{p_\theta(x \mid z)p_\theta(z)}{p_\theta(z \mid x)}$$

We still need $p_\theta(z \mid x)$, which is not explicitly known.
Instead, we train the encoder module to learn

$$q_\phi(z \mid x) \approx p_\theta(z \mid x)$$

Will see why in a minute…

Training Variational Autoencoder



- Recall that, we aim to maximize

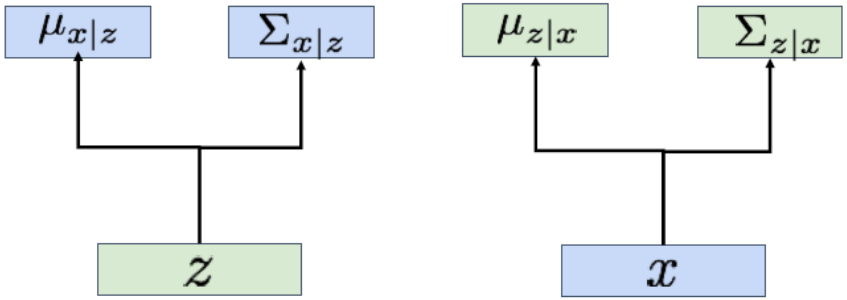
$$p_\theta(x) = \int p_\theta(x,z)dz = \int p_\theta(x|z)p_\theta(z)dz$$

we have…

Decoder network inputs latent code z, gives distribution over data x

Encoder network inputs data x, gives distribution over latent codes z

$$p_\theta(x \mid z) = N(\mu_{x|z}, \Sigma_{x|z}) \qquad q_\phi(z \mid x) = N(\mu_{z|x}, \Sigma_{z|x})$$

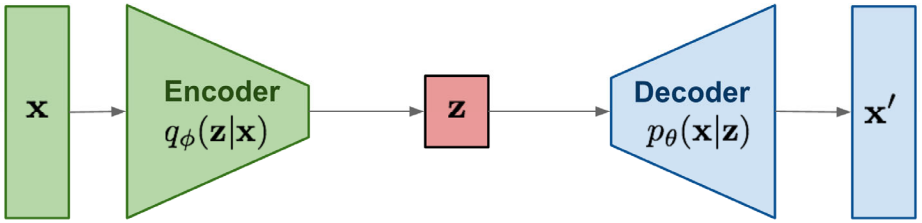


- Via training, if we ensure $q_\phi(z \mid x) \approx p_\theta(z \mid x)$

then we have $p_\theta(x) = \dfrac{p_\theta(x \mid z)p_\theta(z)}{p_\theta(z \mid x)} \approx \dfrac{p_\theta(x \mid z)p(z)}{q_\phi(z \mid x)}$

Slide credit: UMich EECS 498-007

Training VAE



$$\log p_\theta(x) = \log \frac{p_\theta(x \mid z)p(z)}{p_\theta(z \mid x)} = \log \frac{p_\theta(x|z)p(z)q_\phi(z|x)}{p_\theta(z|x)q_\phi(z|x)}$$

$$= E_z[\log p_\theta(x|z)] - E_z\left[\log \frac{q_\phi(z|x)}{p(z)}\right] + E_z\left[\log \frac{q_\phi(z|x)}{p_\theta(z|x)}\right]$$

$$= E_{z\sim q_\phi(z|x)}[\log p_\theta(x|z)] - D_{KL}\left(q_\phi(z|x), p(z)\right) + D_{KL}(q_\phi(z|x), p_\theta(z|x))$$

Data reconstruction

KL divergence between sample distribution from the encoder and the prior

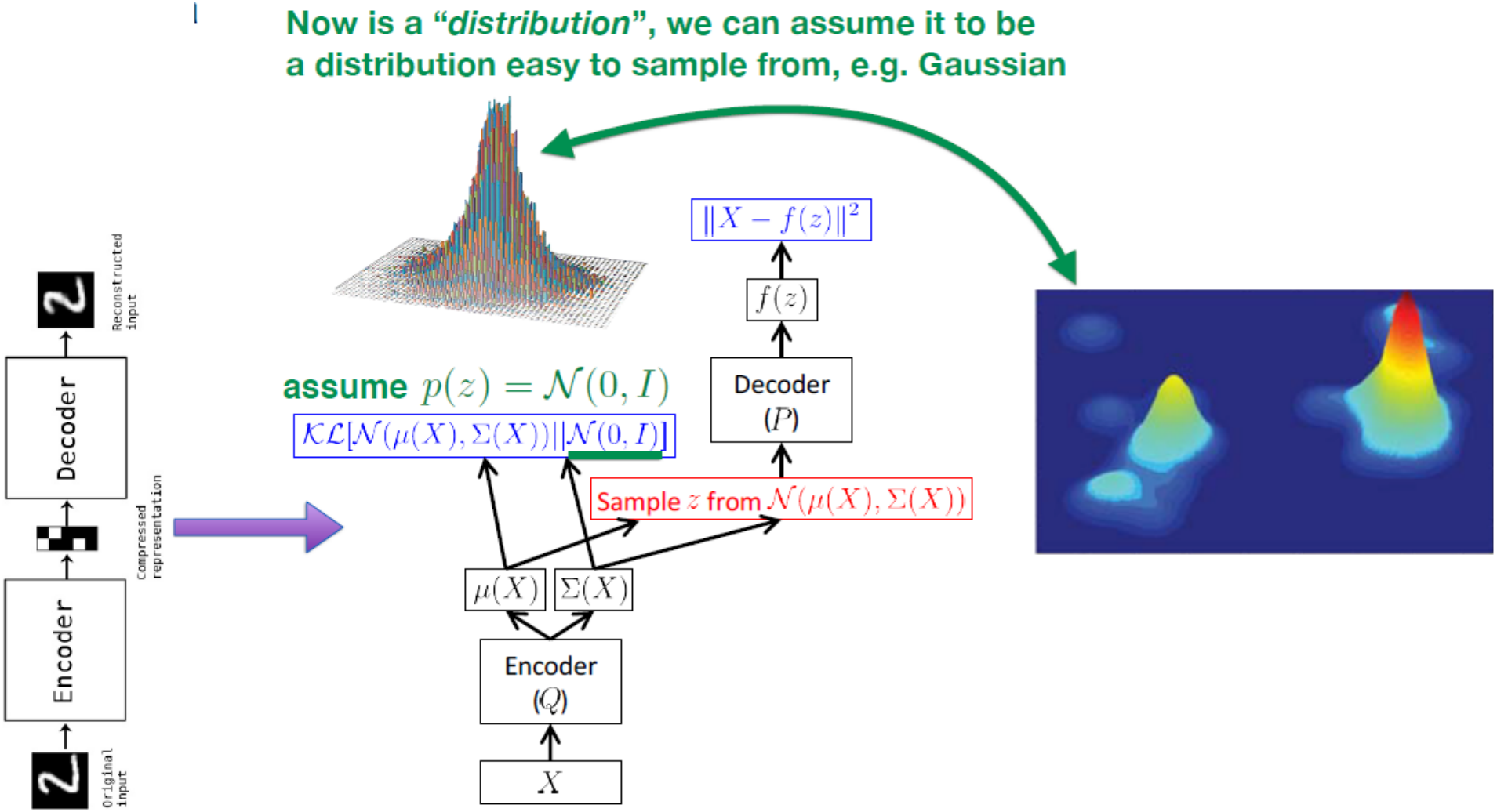
KL divergence between sample distribution from the encoder and the posterior of data

$$\log p_\theta(x) \geq E_{z\sim q_\phi(z|x)}[\log p_\theta(x|z)] - D_{KL}\left(q_\phi(z|x), p(z)\right)$$

which is the variational lower bound, aka evidence lower bound (ELBO), on the data likelihood $p_\theta(x)$

Slide credit: UMich EECS 498-007

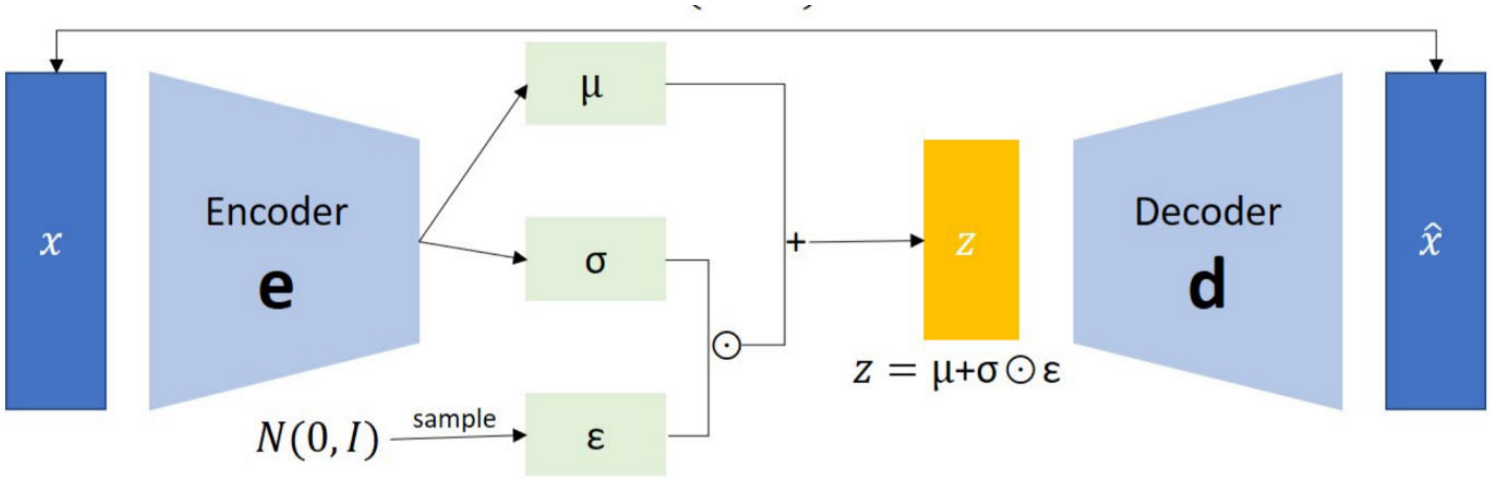
Overview of VAE Training Objectives



Slide credit: W.-C. Chiu

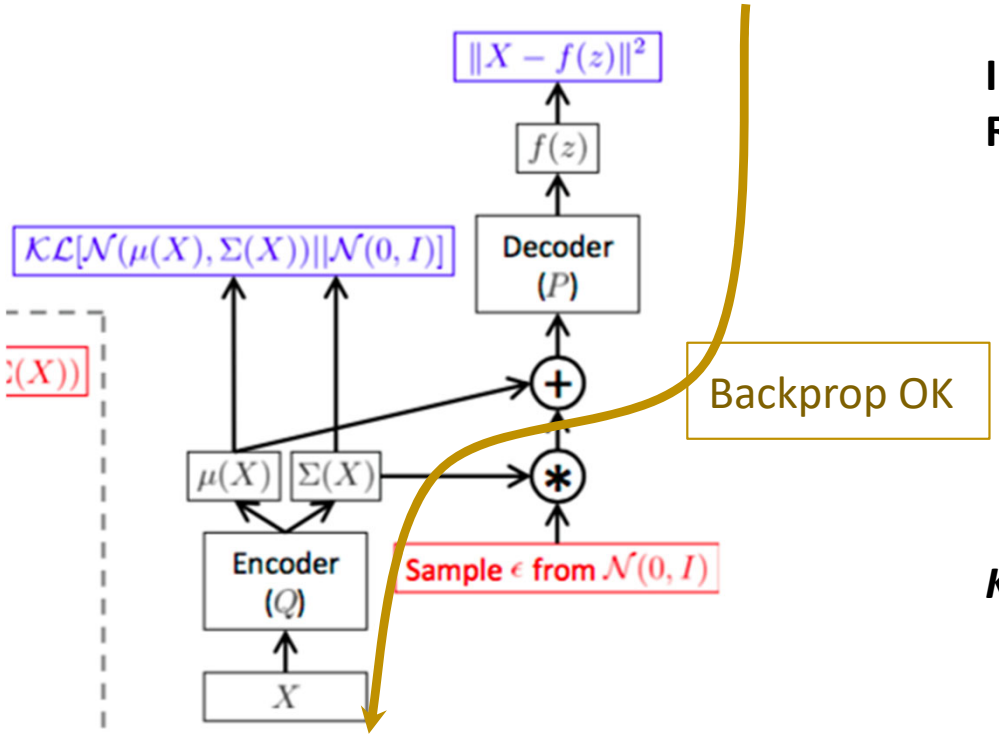
Reparameterization Trick in VAE

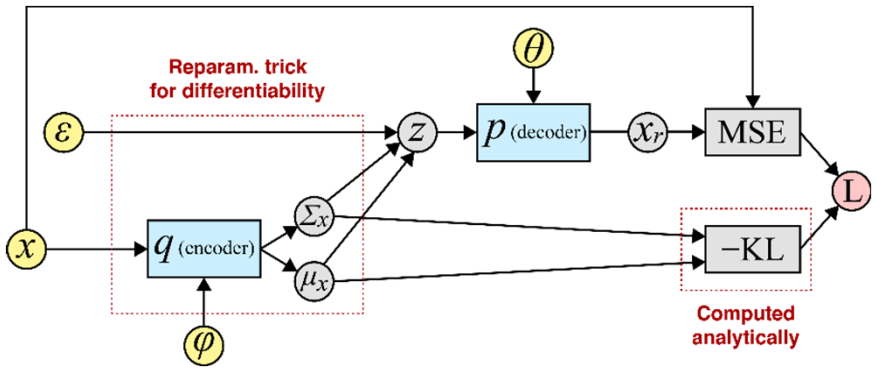
- Remarks
  - Given x, sample z from latent distribution (described by output parameters of encoder)
  - However, this creates a bottleneck since backpropagation (BP) cannot flow through
  - Alternatively, we apply $z = \mu + \sigma \odot \varepsilon$ (ε simply generated by **Normal distribution**).
  - This enables BP gradients in encoder through **μ** and **σ**,
    while maintaining stochasticity via **ε** (for generative synthesis purposes).



https://medium.com/geekculture/variational-autoencoder-vae-9b8ce5475f68

Implementation of VAE Training

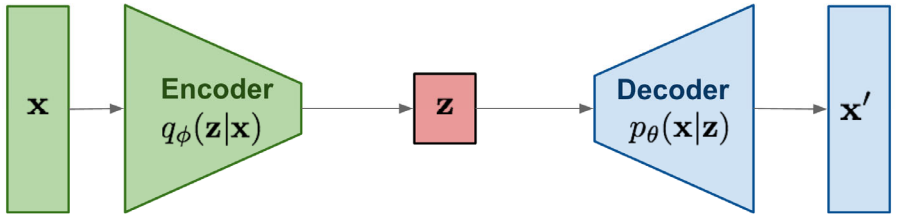




Initialize parameters of encoder and decoder
Repeat:
Get mini-batch of **X**
mu_X, **var_X** = encoder(**X**)
ϵ = sampling from Normal(**0**, **I**)
z = **mu_X** + ***ϵ******var_X**
X' = decoder(**z**)
recon_loss = ***MSE***(**X**,**X'**)
latent_loss = ***KLD***(Normal(**mu_X**,**var_X**)||Normal(**0**,**I**))
all_loss = recon_loss + latent_loss
all_loss.backward()
Until: parameters of encoder & decoder converge
Return parameters of encoder and decoder

First sample noise ***ϵ*** from Normal(0,I),
then reparameterize z by mu_X + ***ϵ****var_X,
(*equivalently sampled Normal(mu_X, var_X)*).
The model is now differentiable!

Before We Move On…



$$\log p_\theta(x) = \log \frac{p_\theta(x \mid z)p(z)}{p_\theta(z \mid x)} = \log \frac{p_\theta(x|z)p(z)q_\phi(z|x)}{p_\theta(z|x)q_\phi(z|x)}$$

$$= E_z[\log p_\theta(x|z)] - E_z\left[\log \frac{q_\phi(z|x)}{p(z)}\right] + E_z\left[\log \frac{q_\phi(z|x)}{p_\theta(z|x)}\right]$$

$$= E_{z\sim q_\phi(z|x)}[\log p_\theta(x|z)] - D_{KL}\left(q_\phi(z|x), p(z)\right) + D_{KL}(q_\phi(z|x), p_\theta(z|x))$$

Data reconstruction

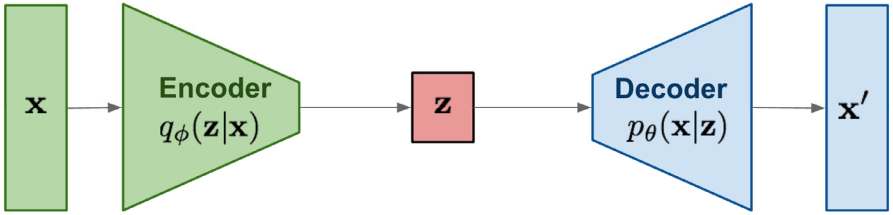
KL divergence between sample distribution from the encoder and the prior

KL divergence between sample distribution from the encoder and the posterior of data

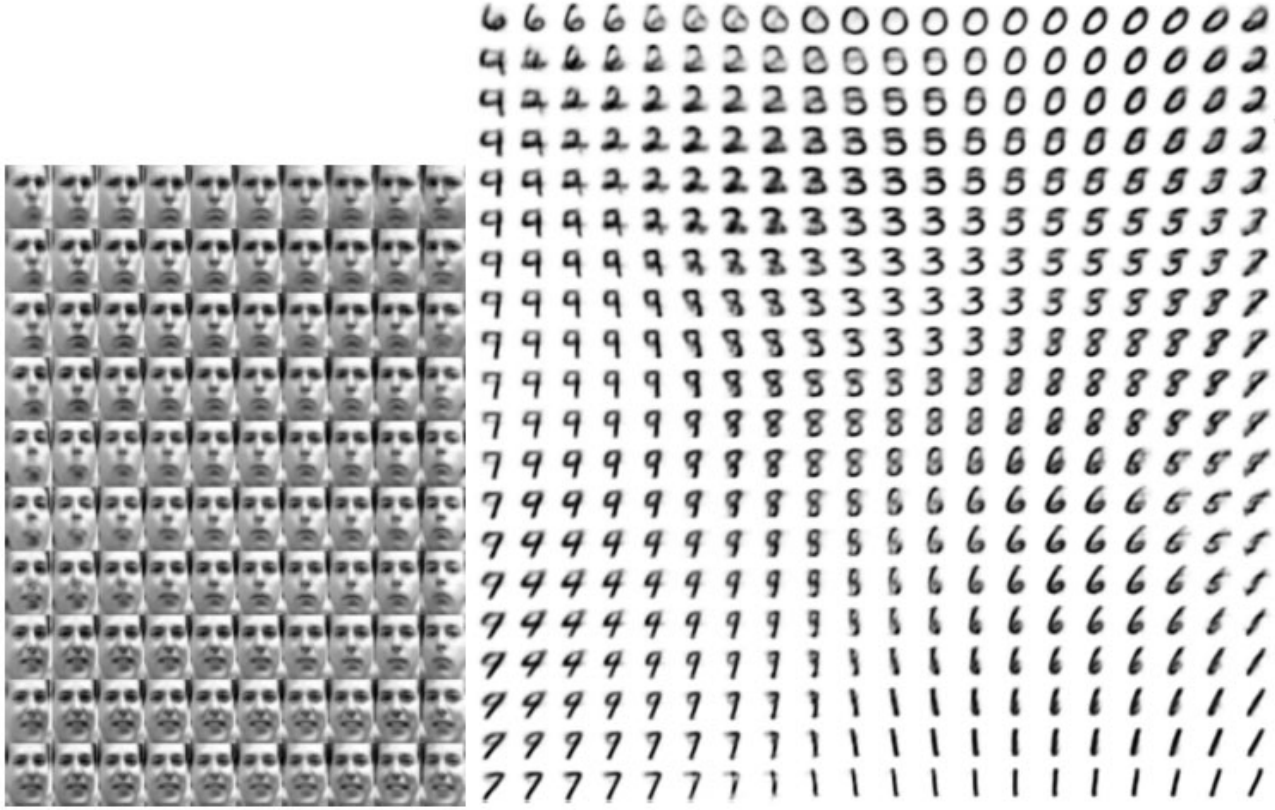
$$\log p_\theta(x) \geq E_{z\sim q_\phi(z|x)}[\log p_\theta(x|z)] - D_{KL}\left(q_\phi(z|x), p(z)\right)$$

i.e., variational lower bound, aka evidence lower bound (ELBO),
on the data likelihood $p_\Theta(x)$

From Autoencoder to VAE (cont'd)

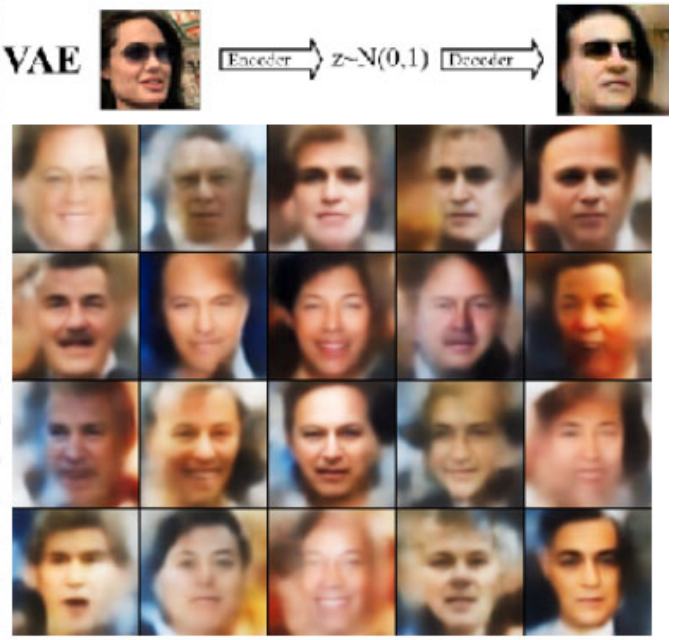


- Some empirical results



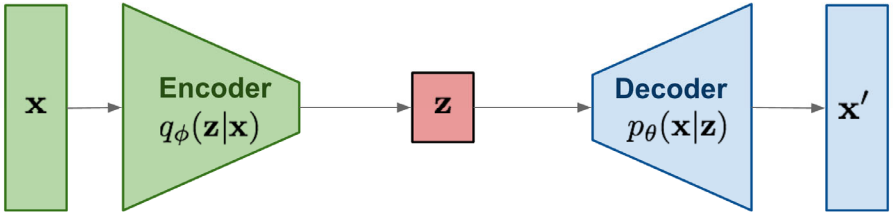
(a) Learned Frey Face manifold

(b) Learned MNIST manifold

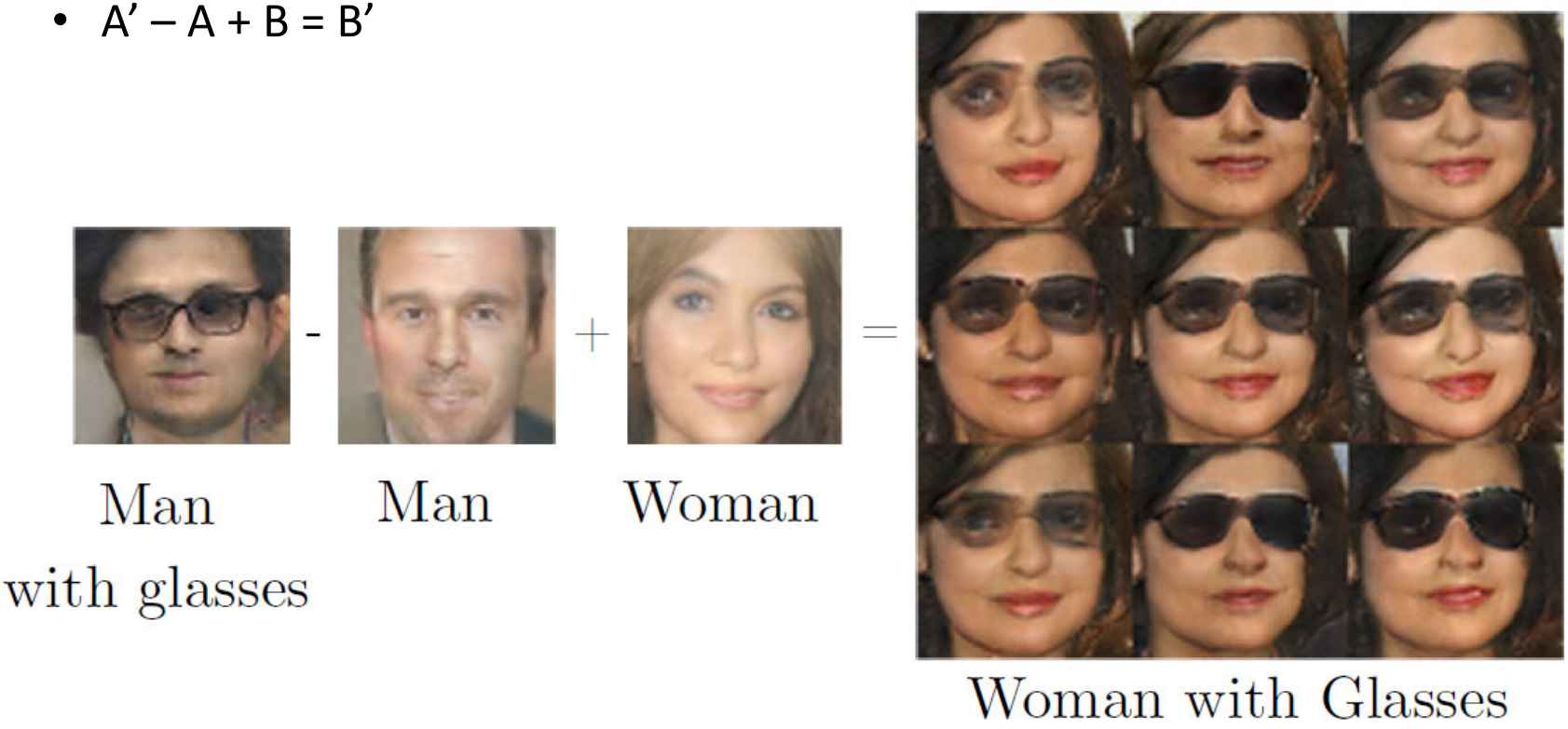


Kingma et al., 2013

From Autoencoder to VAE (cont'd)



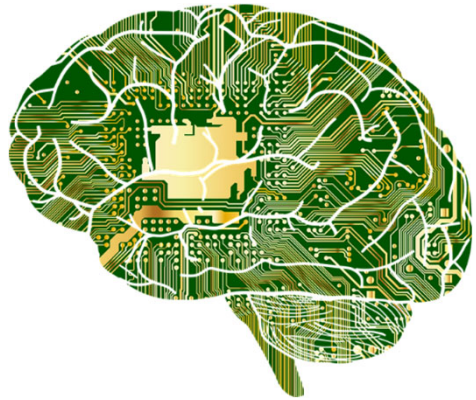
- Additional observations…
  - A' – A + B = B'

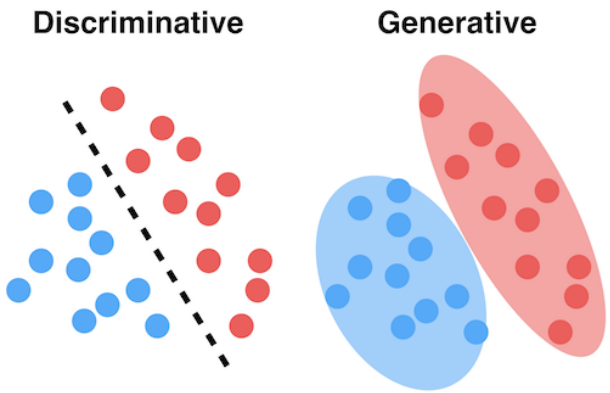


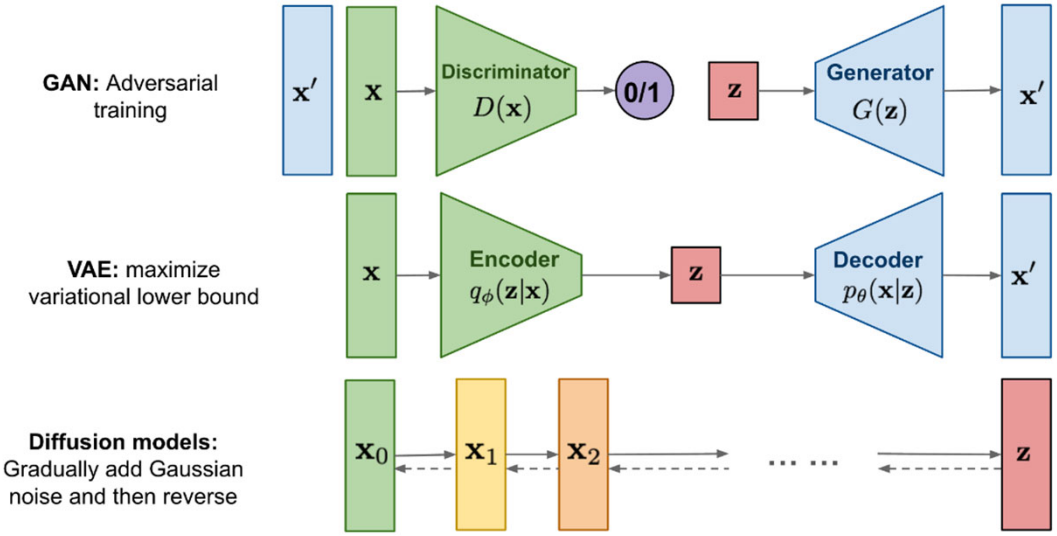
Radford et al., 2015

What's to Be Covered Today…

- Generative Models
  - Autoencoder
  - Variational Autoencoder
  - Diffusion Model
    - Denoising Diffusion Probabilistic Model (DDPM)
    - Latent Diffusion Model (LDM)
    - Denoising Diffusion Implicit Model (DDIM)
  - Generative Adversarial Network







Lecture/slide credit: Prof. Jia-Bin Huang @ U Maryland
https://youtu.be/i2qSxMVeVLI?si=ZIwgo9T-6lEg3kJM

Denoising Diffusion Models

- Emerging as powerful **visual** generative models
  - Unconditional image synthesis
  - Conditional image synthesis

Diffusion Models Beat GANs on Image Synthesis, Dhariwai & Nochol, OpenAI, 2021

Cascaded Diffusion Models for High Fidelity Image Generation, Ho et al., Google, 2021

Slide credit: Kreis, Gao, & Vahdat

Denoising Diffusion Models

- Emerging as powerful **visual** generative models
  - Unconditional image synthesis
  - Conditional image synthesis

DALL·E 2

"a teddy bear on a skateboard in times square"

Diffusion Models Beat GANs on Image Synthesis, Dhariwai & Nochol, OpenAI, 2021

Imagen

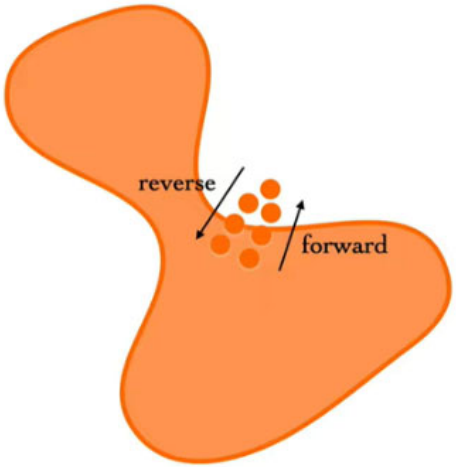
A group of teddy bears in suit in a corporate office celebrating the birthday of their friend. There is a pizza cake on the desk.

Cascaded Diffusion Models for High Fidelity Image Generation, Ho et al., Google, 2021

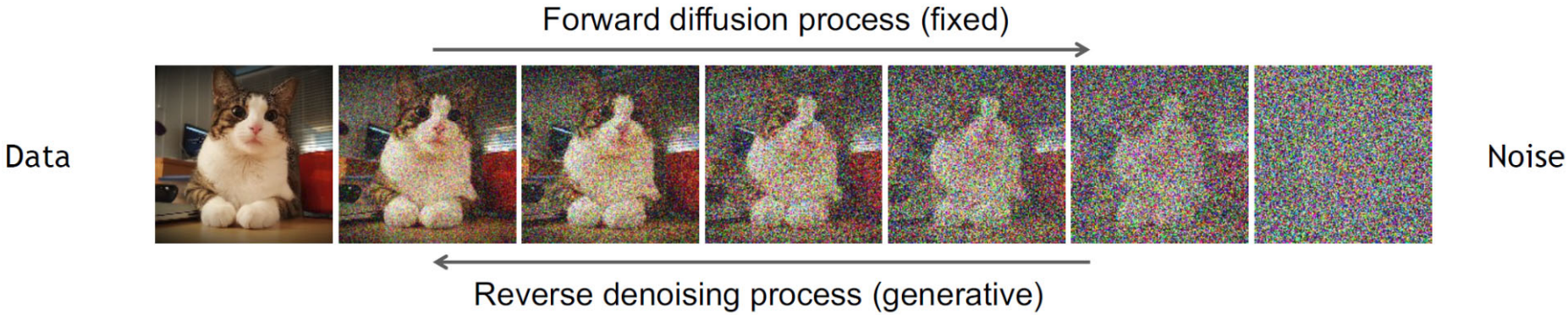
Slide credit: Kreis, Gao, & Vahdat

Denoising Diffusion Probabilistic Models (DDPM)

Learning to generate by denoising



- 2 processes required for training:
  - **Forward diffusion process**
    - gradually add noise to input
  - **Reverse diffusion process**
    - learns to generate/restore data by denoising
    - typically implemented via a **conditional U-net**

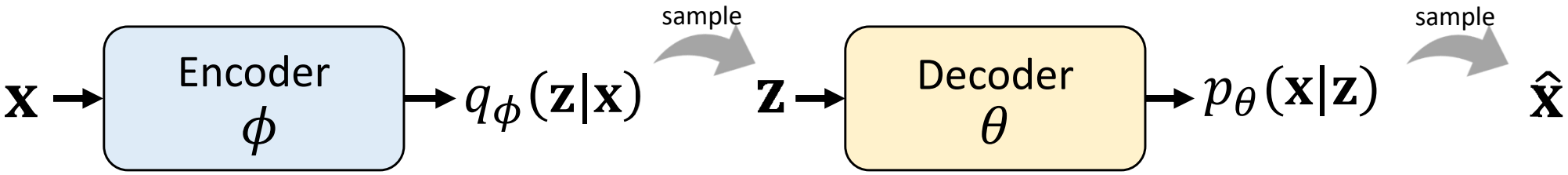


Ho et al., Denoising Diffusion Probabilistic Models, NeurIPS 2020
Song et al., Score-Based Generative Modeling through Stochastic Differential Equations, ICLR 2021

Slide credit: Kreis, Gao, & Vahdat

VAE vs. DDPM

Variational Autoencoder

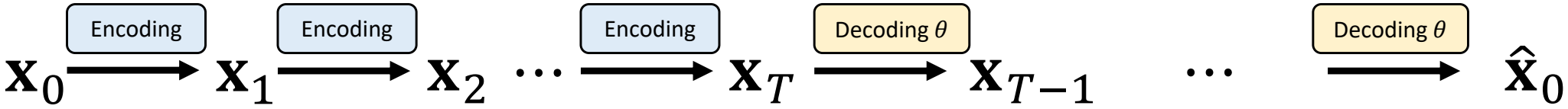


Observed $\mathbf{x}$

Latent $\mathbf{z}$

Maximize $\mathbb{E}_{q(\mathbf{z}|\mathbf{x})}\left[\log \frac{p(\mathbf{x},\mathbf{z})}{q(\mathbf{z}|\mathbf{x})}\right]$

Diffusion model



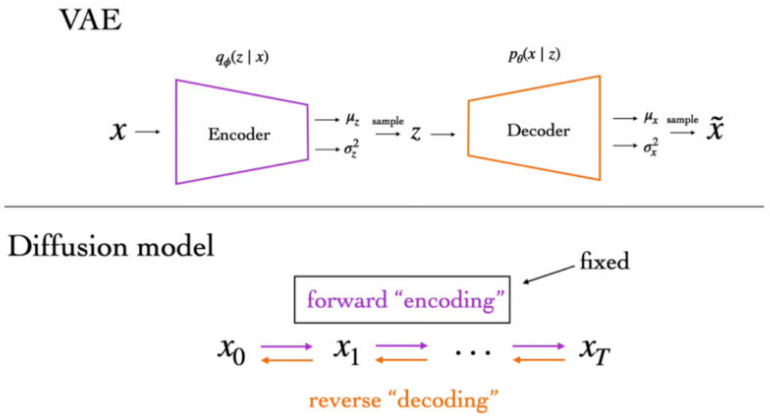
Observed $\mathbf{x}_0$

Latent $\mathbf{x}_1, \ldots, \mathbf{x}_T$

Maximize $\mathbb{E}_{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\left[\log \frac{p(\mathbf{x}_0:\mathbf{x}_T)}{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\right]$

Slide credit: J.-B. Huang

Learning of Diffusion Models



- $\log p_\theta(x) \geq$ variational lower bound
  - Variational bound for optimizing VAE models

Variational Autoencoder $\log p_\theta(x) \geq E_{z\sim q_\phi(z|x)}[\log p_\theta(x|z)] - D_{KL}\left(q_\phi(z|x), p(z)\right)$

vs. **Diffusion model** $\mathbb{E}_{q(\mathbf{x}_0)}\left[-\log p_\theta(\mathbf{x}_0)\right] \leq \mathbb{E}_{q(\mathbf{x}_0)q(\mathbf{x}_{1:T}|\mathbf{x}_0)}\left[-\log \frac{p_\theta(\mathbf{x}_{0:T})}{q(\mathbf{x}_{1:T}|\mathbf{x}_0)}\right] =: L$

- According to Ho *et al.* [NeurIPS'20], it is shown that

$$L = \mathbb{E}_q\left[\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0)||p(\mathbf{x}_T))}_{L_T} + \sum_{t>1}\underbrace{D_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))}_{L_{t-1}} \underbrace{-\log p_\theta(\mathbf{x}_0|\mathbf{x}_1))}_{L_0}\right]$$

$\mathcal{N}(x_{t-1}; \mu_\theta(x_t, t), \beta I)$

Prior matching | **Denoising matching** | **Reconstruction**

$\mathcal{N}(x_{t-1}; \tilde{\mu}_t(x_t, x_0), \tilde{\beta}_t I)$

$$\tilde{\mu}_t(x_t, x_0) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}x_t + \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1-\bar{\alpha}_t}x_0 \quad \text{fixed}$$

$$= \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\varepsilon)$$

Ref: https://lilianweng.github.io/posts/2021-07-11-diffusion-models/
https://youtu.be/fbLgFrlTnGU; https://youtu.be/HoKDTa5jHvg

Let's Take a Look at the Encoding Part…

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \cdots \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

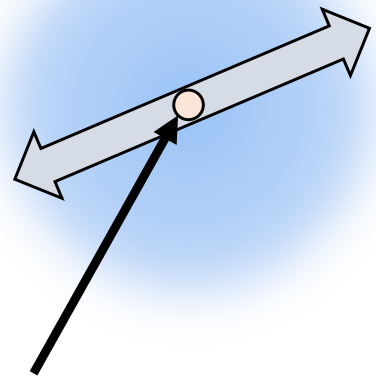
Observed $\mathbf{x}_0$

Latent $\mathbf{x}_1, \ldots, \mathbf{x}_T$

Maximize $\mathbb{E}_{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)} \left[\log \frac{p(\mathbf{x}_0:\mathbf{x}_T)}{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)} \right]$

$$q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0) = \prod_{t=1}^{T} q(\mathbf{x}_t|\mathbf{x}_{t-1})$$

Encoding



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

Encoding

Mean

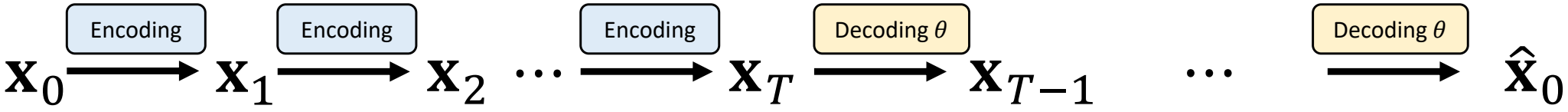
Variance

Coeff < 1…why?

Will get back to this encoding part later.

Slide credit: J.-B. Huang

Now Let's Focus on the Training Objective



Maximize $\mathbb{E}_{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\left[\log \frac{p(\mathbf{x}_0:\mathbf{x}_T)}{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\right]$

1. $-\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0)||p(\mathbf{x}_T))$ **Prior matching**

2. $\mathbb{E}_{q(\mathbf{x}_1|\mathbf{x}_0)}[\log p_\theta(\mathbf{x}_0|\mathbf{x}_1)]$ **Reconstruction**

3. $-\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$ **Denoising matching**

Slide credit: J.-B. Huang

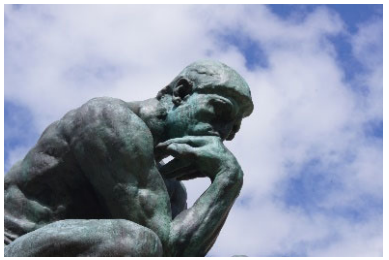
$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\mathbb{E}_{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\left[\log \frac{p(\mathbf{x}_0:\mathbf{x}_T)}{q(\mathbf{x}_1:\mathbf{x}_T|\mathbf{x}_0)}\right]$$

① $-\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_T|\mathbf{x}_0)||p(\mathbf{x}_T))$ **Prior matching** ✗

② $\mathbb{E}_{q(\mathbf{x}_1|\mathbf{x}_0)}[\log p_\theta(\mathbf{x}_0|\mathbf{x}_1)]$ **Reconstruction** ✗

③ $-\sum_{t=2}^{\mathrm{T}} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$ **Denoising matching**



Reparameterization trick

$$\mathbf{x} \sim \mathcal{N}\left(\mathbf{x}; \boldsymbol{\mu}, \sigma^2\right)$$

$$\mathbf{x} = \boldsymbol{\mu} + \sigma\boldsymbol{\epsilon}, \text{ where } \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \mathbf{I})$$

$\mathcal{N}(\boldsymbol{\epsilon}; 0, \mathbf{I})$ $\mathcal{N}\left(\mathbf{x}; \boldsymbol{\mu}, \sigma^2\right)$

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \cdots \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$$

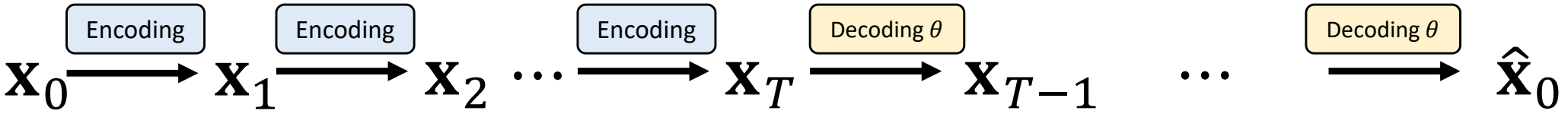
$q(\mathbf{x}_t|\mathbf{x}_0) = \ ?$

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

Encoding — Mean — Variance

$$\mathbf{x}_1 = \sqrt{\alpha_1}\mathbf{x}_0 + \sqrt{1-\alpha_1}\boldsymbol{\epsilon}_0 \qquad \alpha_t = 1-\beta_t, \bar{\alpha}_t = \prod_{i=1}^{t} \alpha_i \qquad \boldsymbol{\epsilon}_0 \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \mathbf{I})$$

$$\mathbf{x}_2 = \sqrt{\alpha_2}\ \mathbf{x}_1 \ + \sqrt{1-\alpha_2}\boldsymbol{\epsilon}_1 \qquad \boldsymbol{\epsilon}_1 \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \mathbf{I})$$



$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \cdots \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$$

$q(\mathbf{x}_t|\mathbf{x}_0) = \ ?$

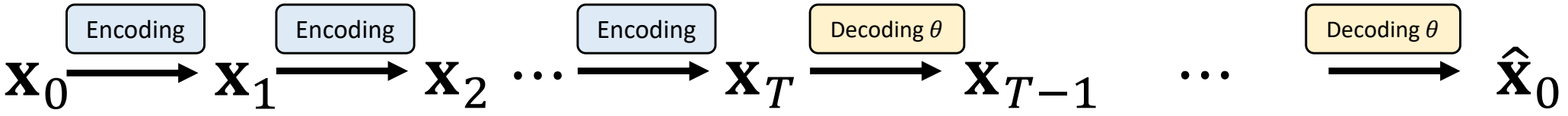
$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_{t-1}, (1-\alpha_t)\mathbf{I})$$

Encoding — Mean — Variance

$$\mathbf{x}_1 = \sqrt{\alpha_1}\mathbf{x}_0 + \sqrt{1-\alpha_1}\boldsymbol{\epsilon}_0 \qquad \boldsymbol{\epsilon}_0 \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \mathbf{I})$$

$$\mathbf{x}_2 = \sqrt{\alpha_2}\ (\sqrt{\alpha_1}\mathbf{x}_0 + \sqrt{1-\alpha_1}\boldsymbol{\epsilon}_0) \quad + \sqrt{1-\alpha_2}\boldsymbol{\epsilon}_1 \qquad \boldsymbol{\epsilon}_1 \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \mathbf{I})$$

$$= \sqrt{\alpha_1\alpha_2}\mathbf{x}_0 + \sqrt{\alpha_2(1-\alpha_1)}\boldsymbol{\epsilon}_0 + \sqrt{1-\alpha_2}\boldsymbol{\epsilon}_1$$



$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$$

$q(\mathbf{x}_t|\mathbf{x}_0) = \ ?$

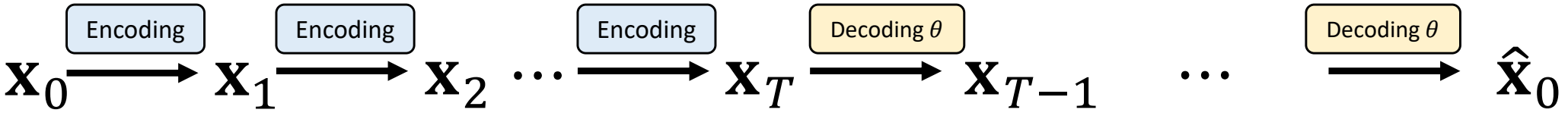
$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{\alpha_t}\mathbf{x}_{t-1}, (1-\alpha_t)\mathbf{I})$$

Encoding — Mean — Variance

$$\mathbf{x}_1 = \sqrt{\alpha_1}\mathbf{x}_0 + \sqrt{1-\alpha_1}\boldsymbol{\epsilon}_0 \qquad \boldsymbol{\epsilon}_0 \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \mathbf{I})$$

$$\mathbf{x}_2 = \sqrt{\alpha_2}\ (\sqrt{\alpha_1}\mathbf{x}_0 + \sqrt{1-\alpha_1}\boldsymbol{\epsilon}_0) \ + \sqrt{1-\alpha_2}\boldsymbol{\epsilon}_1 \qquad \boldsymbol{\epsilon}_1 \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \mathbf{I})$$

$$= \sqrt{\alpha_1\alpha_2}\mathbf{x}_0 + \sqrt{1-\alpha_1\alpha_2}\boldsymbol{\epsilon} \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \mathbf{I})$$



$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$$

$$q(\mathbf{x}_t|\mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_t; \sqrt{\overline{\alpha_t}}\mathbf{x}_0, (1-\overline{\alpha_t})\mathbf{I}\right)$$

Mean Variance

$$\mathbf{x}_t = \sqrt{\overline{\alpha_t}}\mathbf{x}_0 + \sqrt{1-\overline{\alpha_t}}\boldsymbol{\epsilon} \qquad \overline{\alpha_t} = \prod_{i=1}^{t} \alpha_i \qquad \boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \mathbf{I})$$

What does this equation imply?

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \cdots \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$$

$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$

$\mathbf{x}_0$ $\mathbf{x}_{t-1}$ $\mathbf{x}_t$

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \cdots \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$$

$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$

$\mathbf{x}_0$

$\mathbf{x}_{t-1}$

$\mathbf{x}_t$

$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$

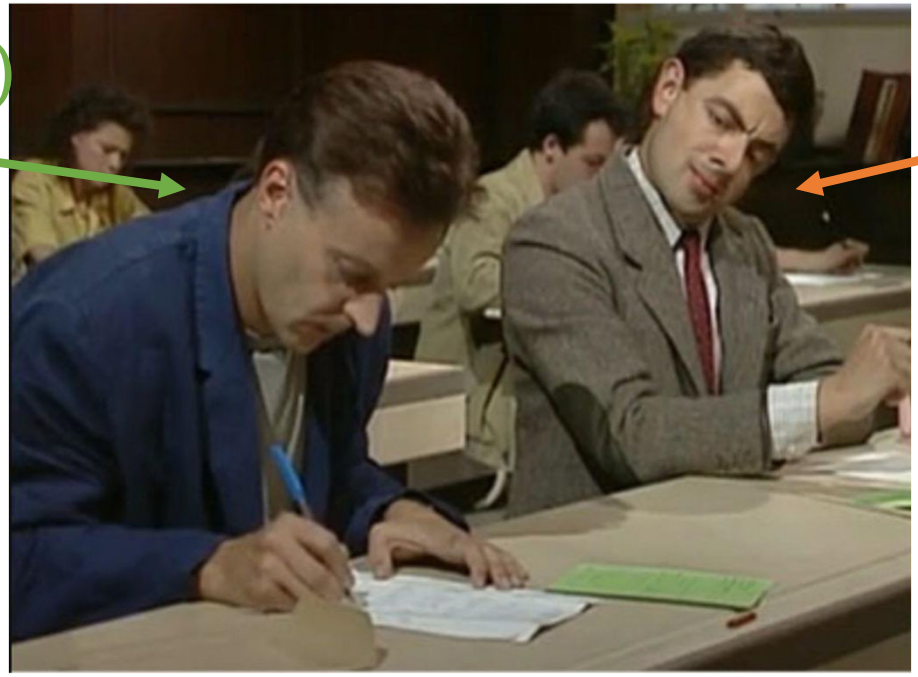
$\mathbf{x}_{t-1}$

$\mathbf{x}_t$

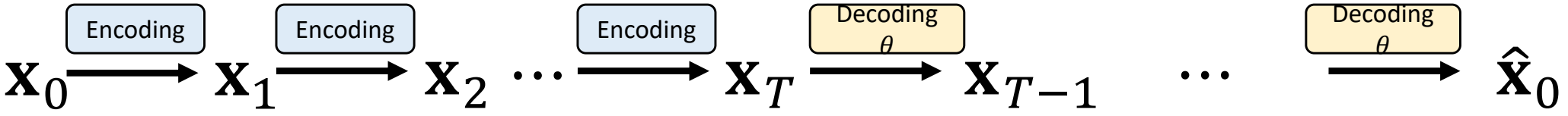
$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \cdots \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$$

$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)$



$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$



$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow[\theta]{\text{Decoding}} \mathbf{x}_{T-1} \cdots \xrightarrow[\theta]{\text{Decoding}} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$$

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1})q(\mathbf{x}_{t-1}|\mathbf{x}_0)q(\mathbf{x}_0)}{q(\mathbf{x}_t|\mathbf{x}_0)\,q(\mathbf{x}_0)}$$

$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow[\theta]{\text{Decoding}} \mathbf{x}_{T-1} \quad \cdots \quad \xrightarrow[\theta]{\text{Decoding}} \hat{\mathbf{x}}_0$$

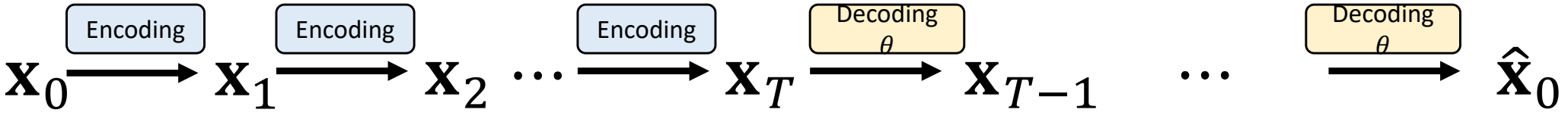
$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$$

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \frac{q(\mathbf{x}_t|\mathbf{x}_{t-1})q(\mathbf{x}_{t-1}|\mathbf{x}_0)}{q(\mathbf{x}_t|\mathbf{x}_0)}$$

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

$$q(\mathbf{x}_{t-1}|\mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_{t-1}; \sqrt{\overline{\alpha_{t-1}}}\mathbf{x}_0, (1-\overline{\alpha_{t-1}})\mathbf{I}\right)$$

$$q(\mathbf{x}_t|\mathbf{x}_0) \quad = \mathcal{N}\left(\mathbf{x}_t; \sqrt{\overline{\alpha_t}}\mathbf{x}_0, (1-\overline{\alpha_t})\mathbf{I}\right)$$



$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$$

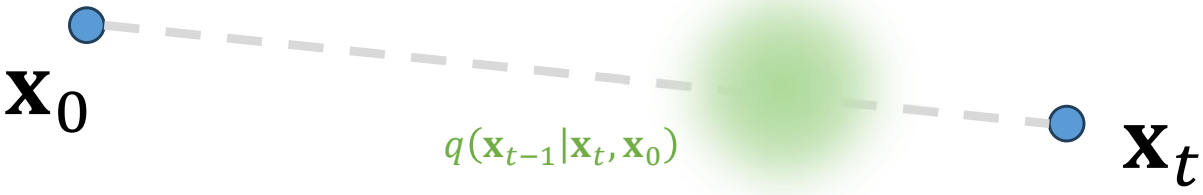
$$q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_{t-1};\mu_q(\mathbf{x}_t,\mathbf{x}_0),\Sigma_q(t)\right)$$

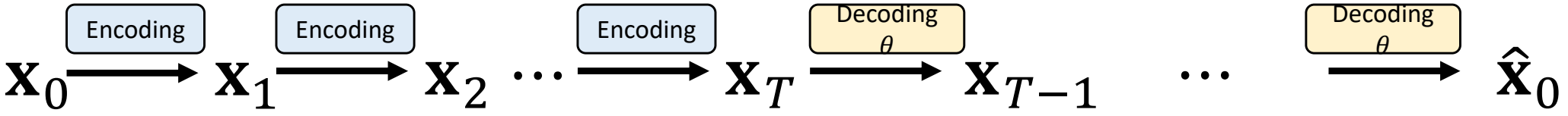
$$\mathbf{x}_0 \xrightarrow{\text{Encoding}} \mathbf{x}_1 \xrightarrow{\text{Encoding}} \mathbf{x}_2 \cdots \xrightarrow{\text{Encoding}} \mathbf{x}_T \xrightarrow{\text{Decoding } \theta} \mathbf{x}_{T-1} \cdots \xrightarrow{\text{Decoding } \theta} \hat{\mathbf{x}}_0$$

$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$$

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_{t-1}; \mu_q(\mathbf{x}_t, \mathbf{x}_0), \Sigma_q(t)\right)$$

$$\underbrace{\mu_q(\mathbf{x}_t, \mathbf{x}_0)}_{\text{Mean}} = \frac{\sqrt{\alpha_t}(1-\overline{\alpha_{t-1}})}{1-\overline{\alpha_t}}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha_{t-1}}}(1-\alpha_t)}{1-\overline{\alpha_t}}\mathbf{x}_0$$





$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$$

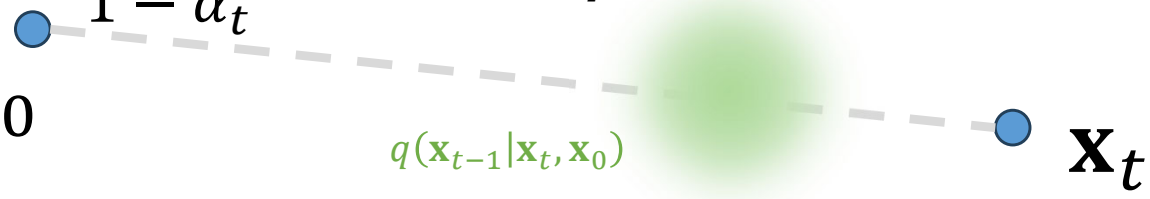
$$q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_{t-1}; \mu_q(\mathbf{x}_t,\mathbf{x}_0), \Sigma_q(t)\right)$$

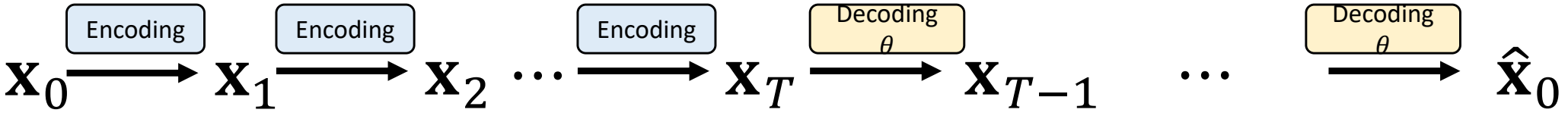
$$\mu_q(\mathbf{x}_t,\mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha_{t-1}})}{1-\overline{\alpha_t}}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha_{t-1}}}(1-\alpha_t)}{1-\overline{\alpha_t}}\mathbf{x}_0$$

Mean

$$\Sigma_q(t) = \frac{(1-\alpha_t)(1-\overline{\alpha_{t-1}})}{1-\overline{\alpha_t}}\boldsymbol{I} = \boldsymbol{\sigma}_q^2(t)\boldsymbol{I}$$

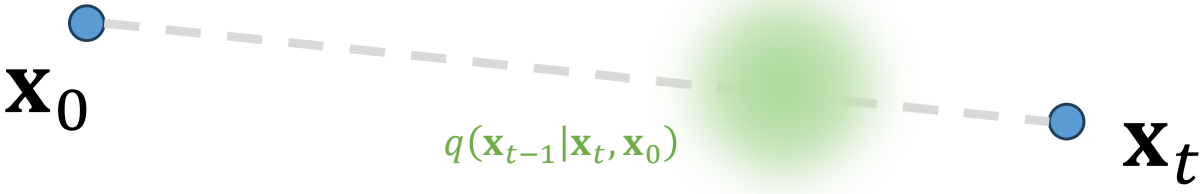
Variance

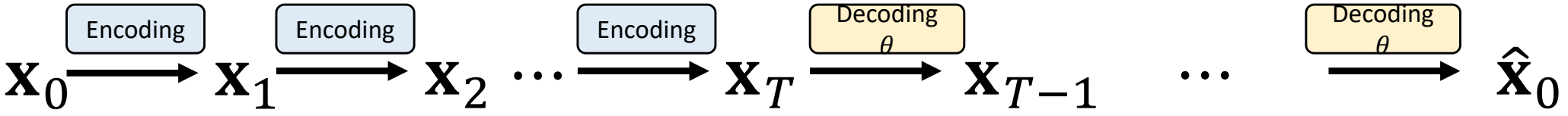




$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$$

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_{t-1};\mu_q(\mathbf{x}_t,\mathbf{x}_0),\Sigma_q(t)\right)$$

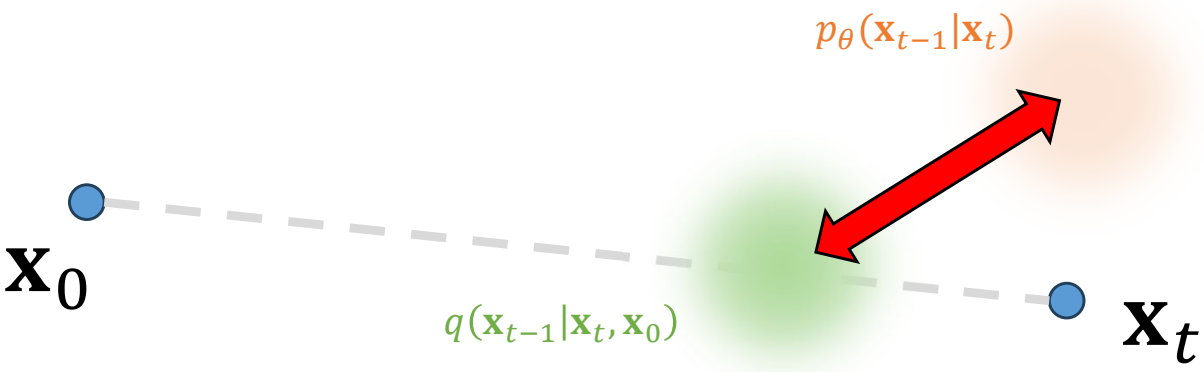


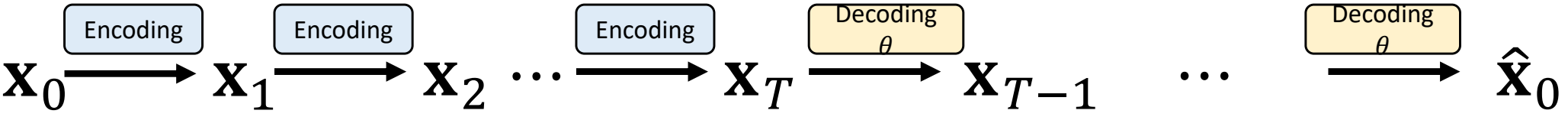


$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$$

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_{t-1};\mu_q(\mathbf{x}_t,\mathbf{x}_0),\Sigma_q(t)\right)$$

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}\left(\mathbf{x}_{t-1};\mu_\theta(\mathbf{x}_t,t),\Sigma_q(t)\right)$$

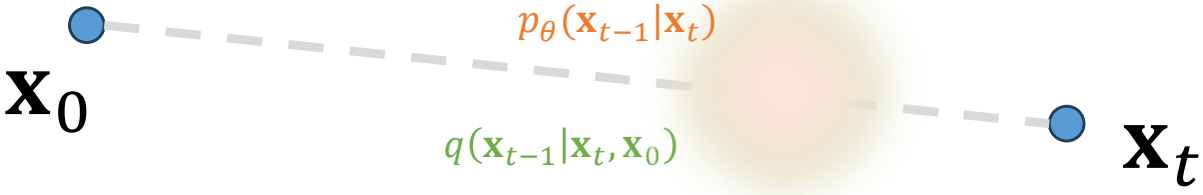


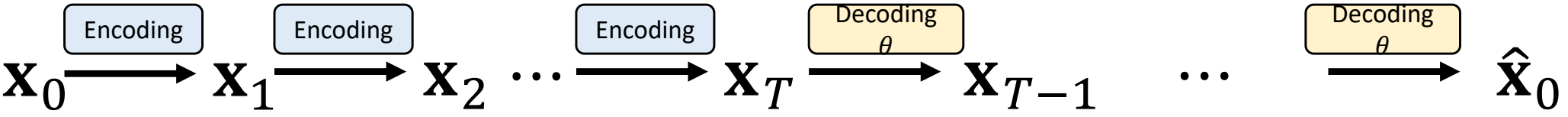


$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$$

$$q(\mathbf{x}_{t-1}|\mathbf{x}_t,\mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_{t-1};\mu_q(\mathbf{x}_t,\mathbf{x}_0),\Sigma_q(t)\right)$$

$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}\left(\mathbf{x}_{t-1};\mu_\theta(\mathbf{x}_t,t),\Sigma_q(t)\right)$$

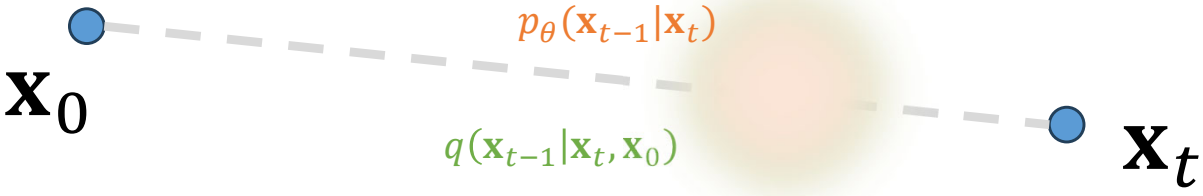


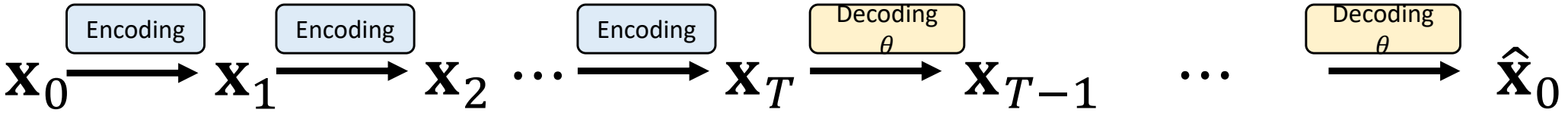


$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}[\mathrm{D}_{\mathrm{KL}}(q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0)||p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t))]$$

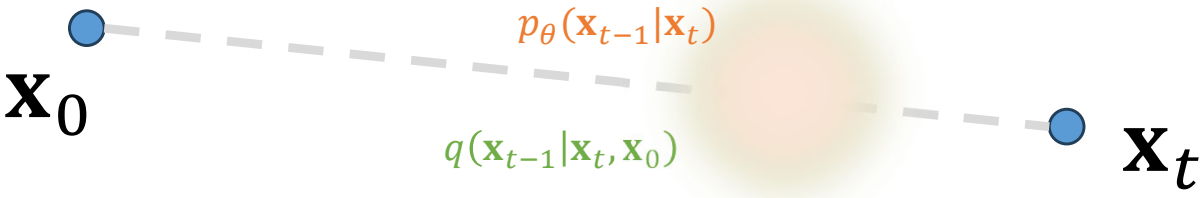
$$q(\mathbf{x}_{t-1}|\mathbf{x}_t, \mathbf{x}_0) = \mathcal{N}\left(\mathbf{x}_{t-1}; \boldsymbol{\mu_q}(\mathbf{x_t}, \mathbf{x_0}), \Sigma_q(t)\right)$$

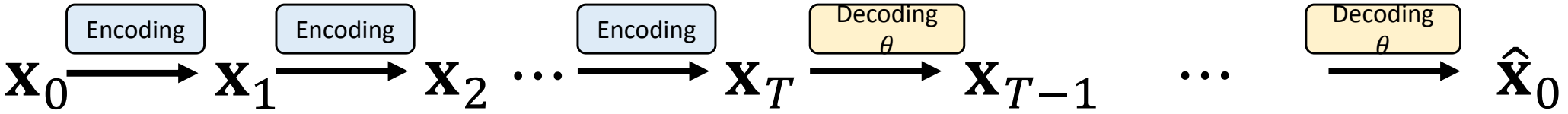
$$p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}\left(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \Sigma_q(t)\right)$$



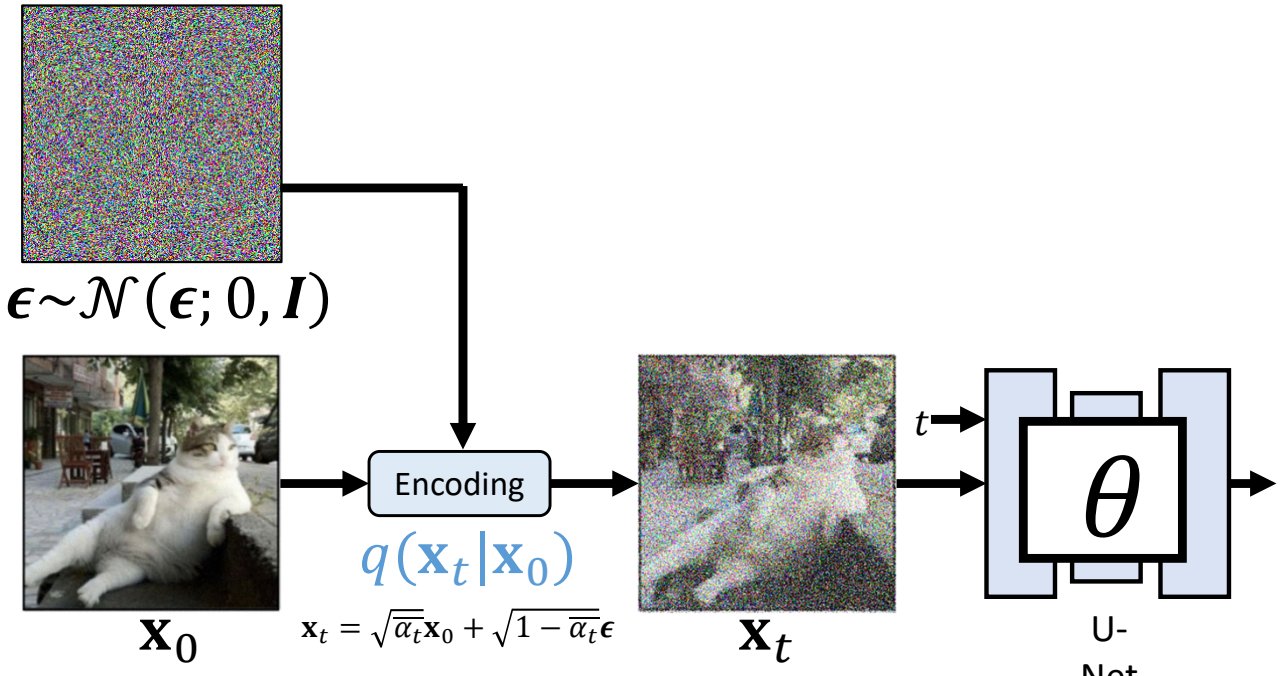


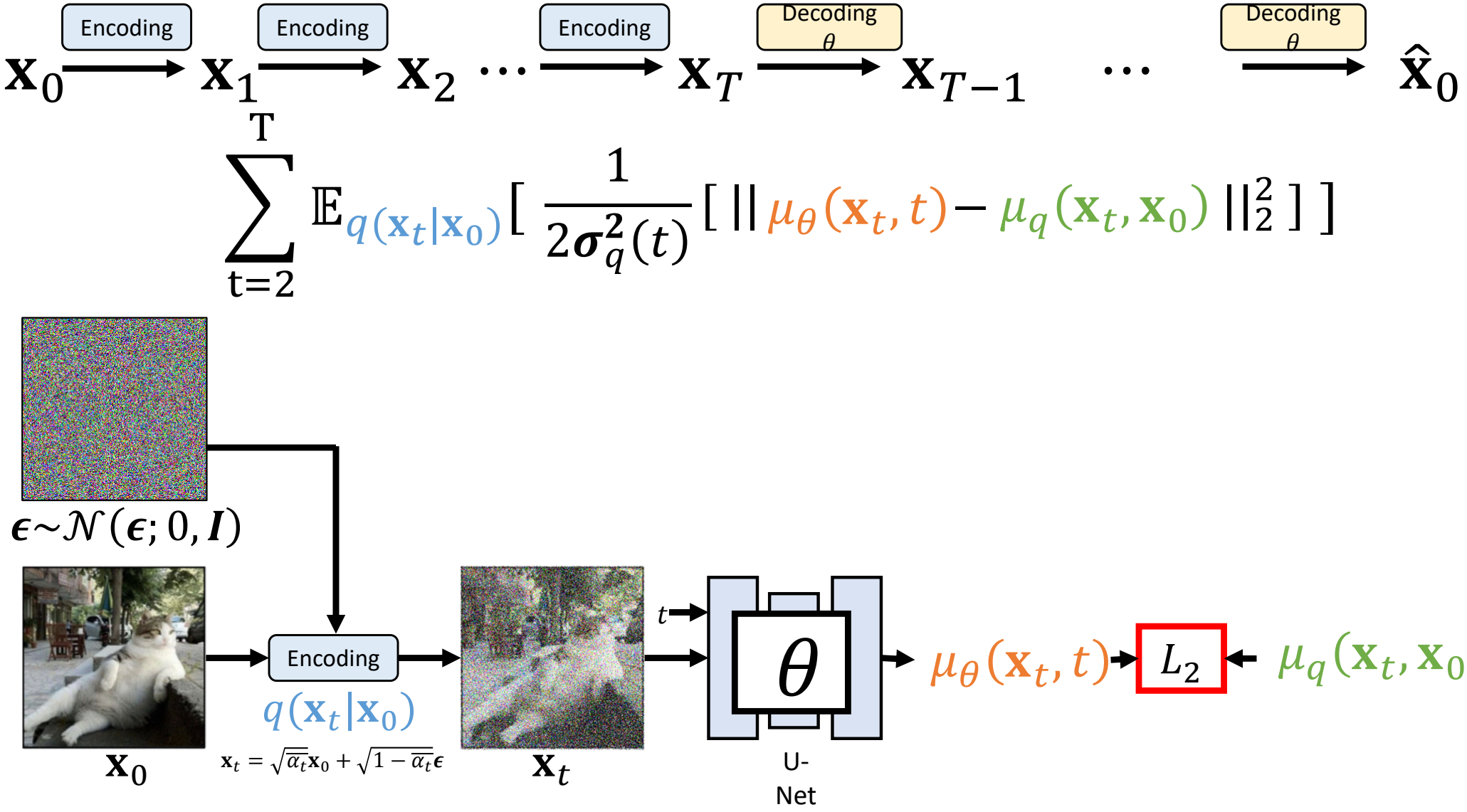
$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\frac{1}{2\boldsymbol{\sigma}_q^2(t)} \left[\| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) \|_2^2 \right] \right]$$



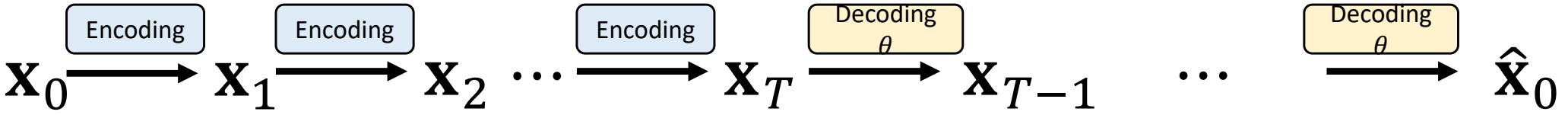


$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\frac{1}{2\boldsymbol{\sigma}_q^2(t)} \left[\| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) \|_2^2 \right] \right]$$




Encoding
Encoding
Encoding
Decoding
θ
Decoding
θ
$\mathbf{x}_0$ $\mathbf{x}_1$ $\mathbf{x}_2$ ⋯ $\mathbf{x}_T$ $\mathbf{x}_{T-1}$ ⋯ $\hat{\mathbf{x}}_0$
$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\frac{1}{2\sigma_q^2(t)}\left[\|\mu_\theta(\mathbf{x}_t,t)-\mu_q(\mathbf{x}_t,\mathbf{x}_0)\|_2^2\right]\right]$
$\epsilon \sim \mathcal{N}(\epsilon; 0, I)$
Encoding
$q(\mathbf{x}_t|\mathbf{x}_0)$
$\mathbf{x}_t = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\epsilon$
$\mathbf{x}_0$
$\mathbf{x}_t$
t
θ
U-Net
$\mu_\theta(\mathbf{x}_t,t)$
L_2
$\mu_q(\mathbf{x}_t,\mathbf{x}_0)$

Observation #1



$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\frac{1}{2\sigma_q^2(t)}\left[\,\|\mu_\theta(\mathbf{x}_t,t)-\mu_q(\mathbf{x}_t,\mathbf{x}_0)\|_2^2\,\right]\right]$$

$$\mu_q(\mathbf{x}_t,\mathbf{x}_0)=\frac{\sqrt{\alpha_t}(1-\overline{\alpha_{t-1}})}{1-\overline{\alpha_t}}\mathbf{x}_t+\frac{\sqrt{\overline{\alpha_{t-1}}}(1-\alpha_t)}{1-\overline{\alpha_t}}\mathbf{x}_0$$

$$\mu_\theta(\mathbf{x}_t,t)=\frac{\sqrt{\alpha_t}(1-\overline{\alpha_{t-1}})}{1-\overline{\alpha_t}}\mathbf{x}_t+\frac{\sqrt{\overline{\alpha_{t-1}}}(1-\alpha_t)}{1-\overline{\alpha_t}}\hat{\mathbf{x}}_\theta(\mathbf{x}_t,t)$$

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

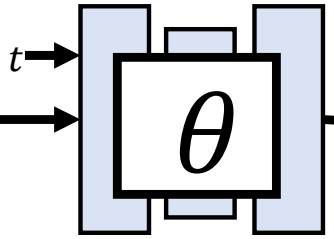
$\mathbf{x}_0$

Encoding

$q(\mathbf{x}_t|\mathbf{x}_0)$

$\mathbf{x}_t = \sqrt{\overline{\alpha_t}}\mathbf{x}_0 + \sqrt{1-\overline{\alpha_t}}\epsilon$

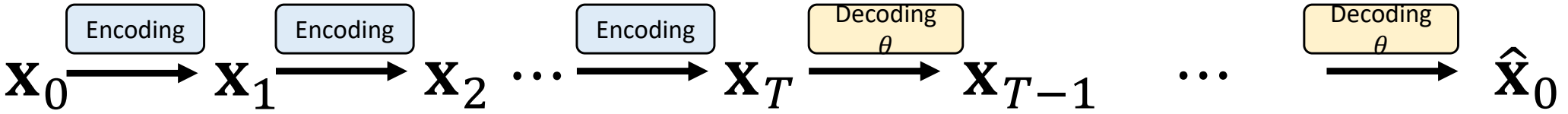
$\mathbf{x}_t$



U-Net

$\mu_\theta(\mathbf{x}_t,t)$ L_2 $\mu_q(\mathbf{x}_t,\mathbf{x}_0)$

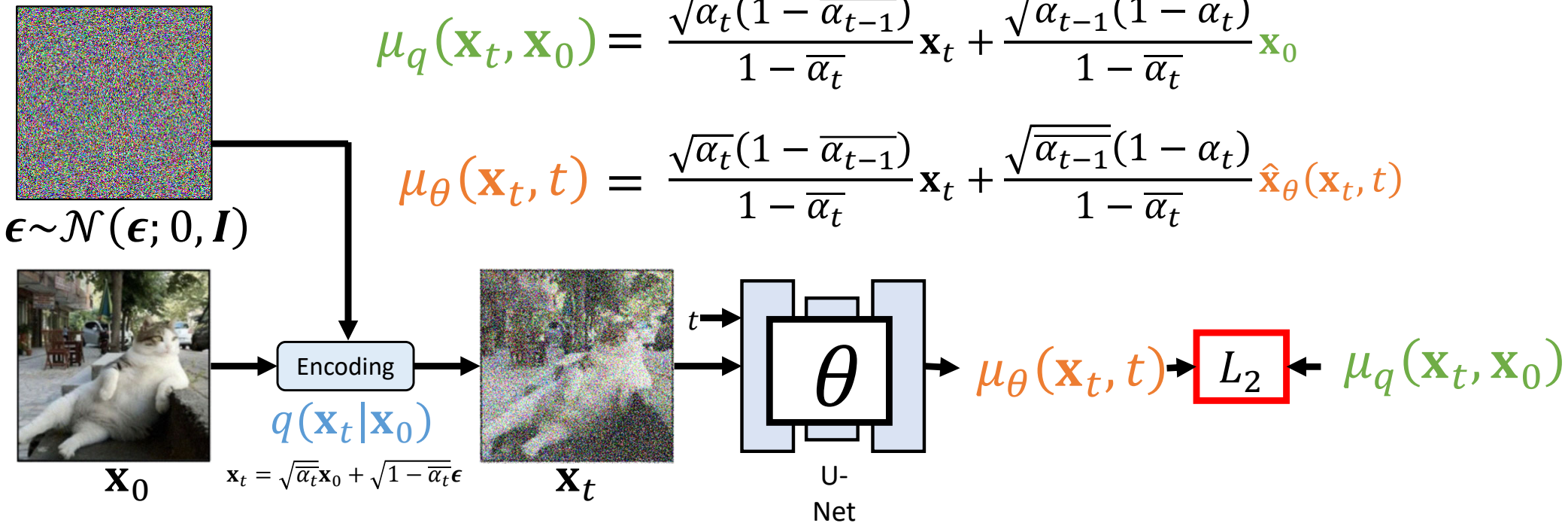
Observation #1



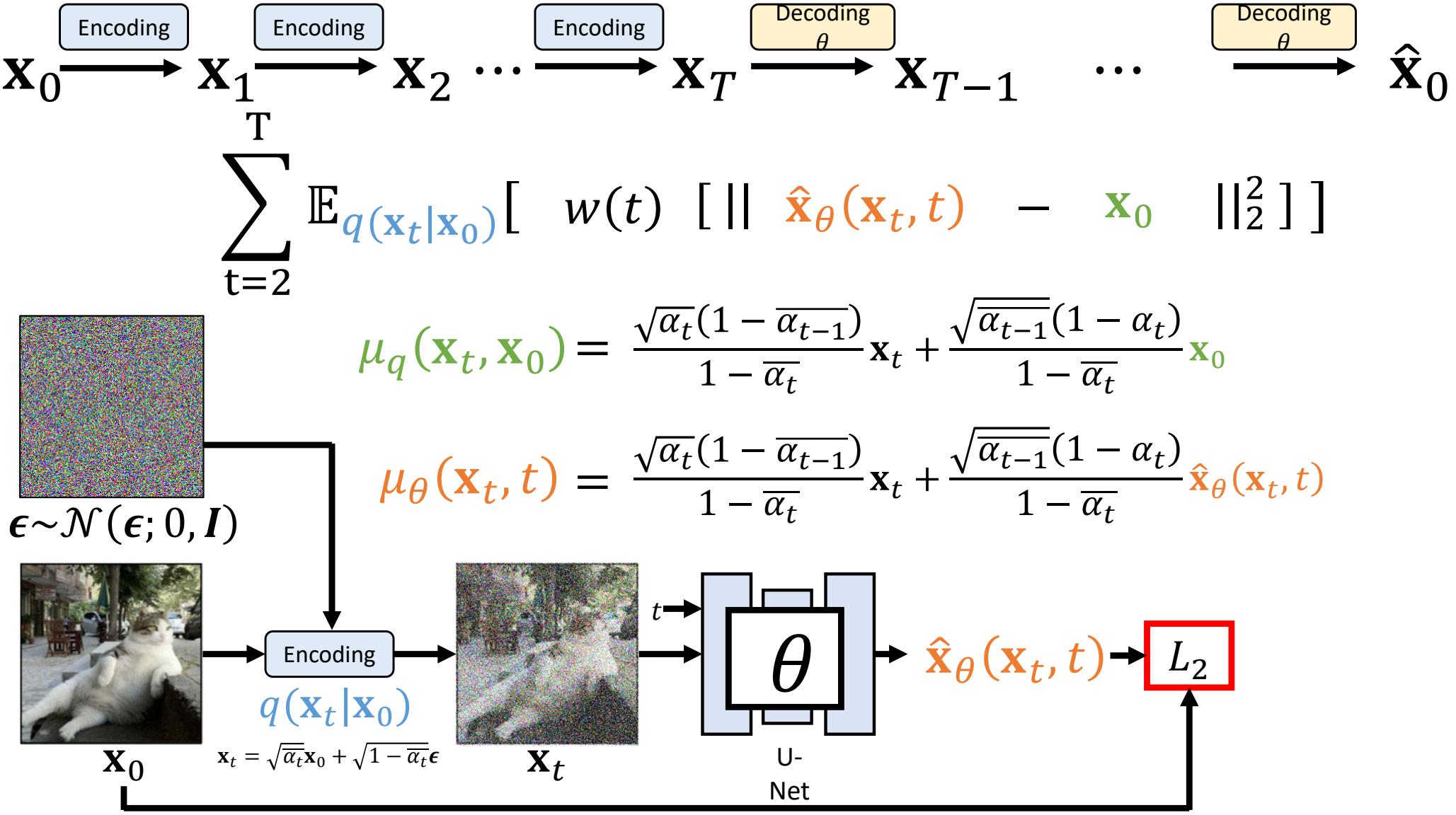
$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\; w(t) \; \left[\| \; \hat{\mathbf{x}}_\theta(\mathbf{x}_t, t) \; - \; \mathbf{x}_0 \; \|_2^2 \; \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha_{t-1}})}{1-\overline{\alpha_t}}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha_{t-1}}}(1-\alpha_t)}{1-\overline{\alpha_t}}\mathbf{x}_0$$

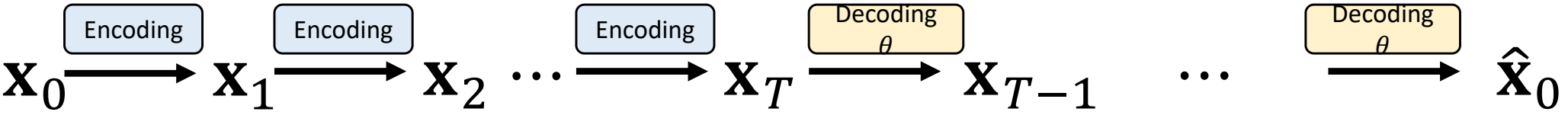
$$\mu_\theta(\mathbf{x}_t, t) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha_{t-1}})}{1-\overline{\alpha_t}}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha_{t-1}}}(1-\alpha_t)}{1-\overline{\alpha_t}}\hat{\mathbf{x}}_\theta(\mathbf{x}_t, t)$$



Observation #1

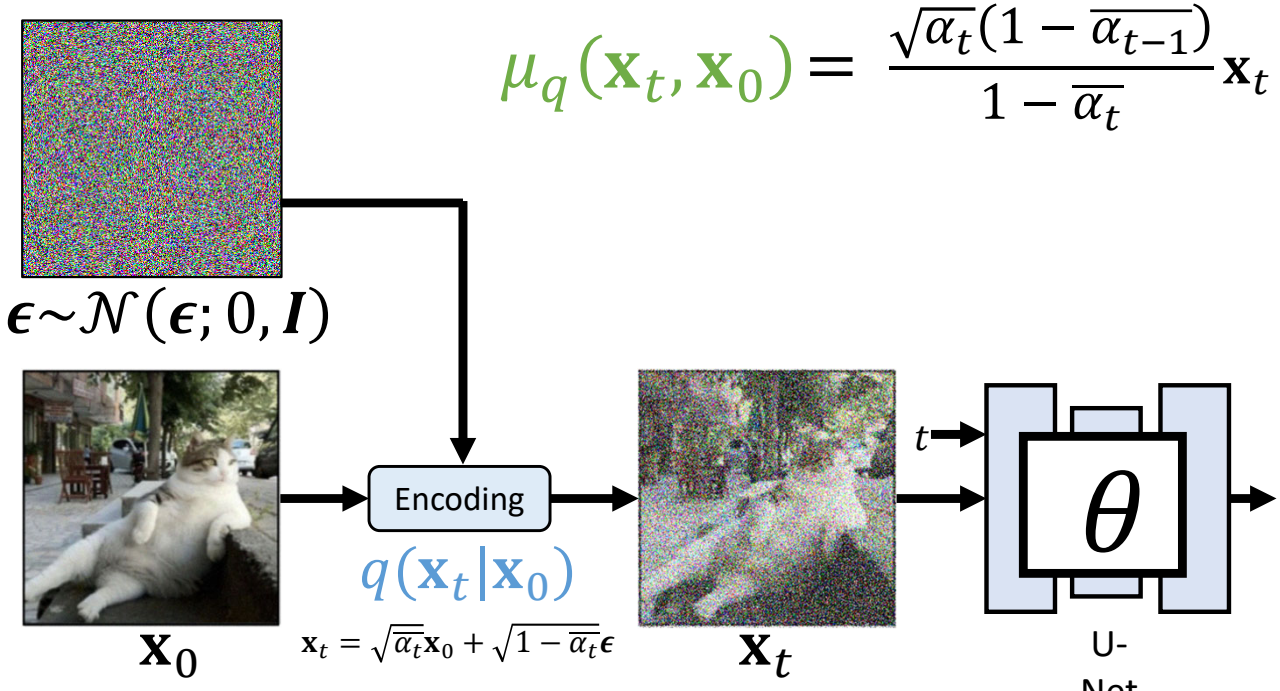


Observation #1

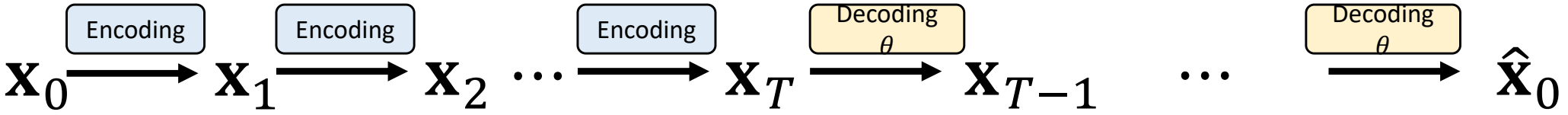


$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\; w(t) \; \left[\left|\left| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) \right|\right|_2^2 \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha_{t-1}})}{1-\overline{\alpha_t}}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha_{t-1}}}(1-\alpha_t)}{1-\overline{\alpha_t}}\mathbf{x}_0$$



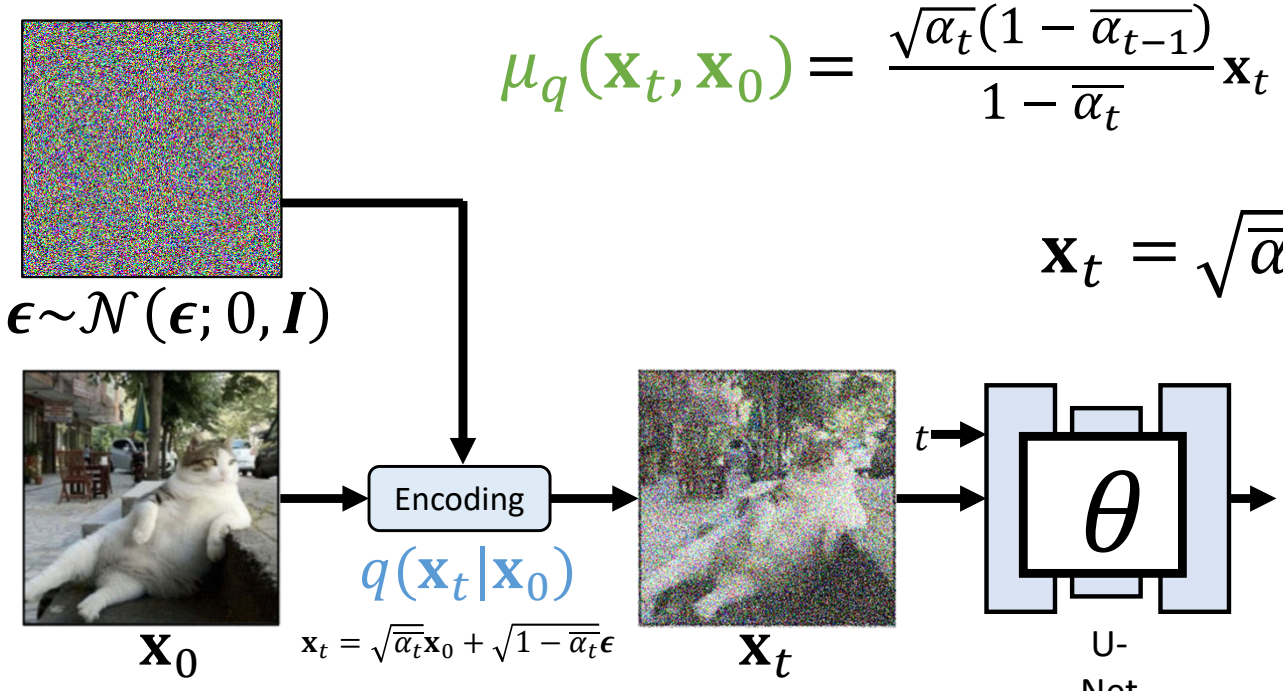
Observation #2



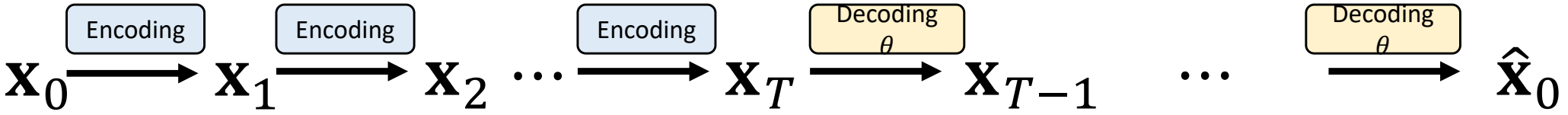
$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\; w(t) \; \left[\| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) \|_2^2 \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha_{t-1}})}{1-\overline{\alpha_t}}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha_{t-1}}}(1-\alpha_t)}{1-\overline{\alpha_t}}\mathbf{x}_0$$

$$\mathbf{x}_t = \sqrt{\overline{\alpha_t}}\mathbf{x}_0 + \sqrt{1-\overline{\alpha_t}}\boldsymbol{\epsilon}$$



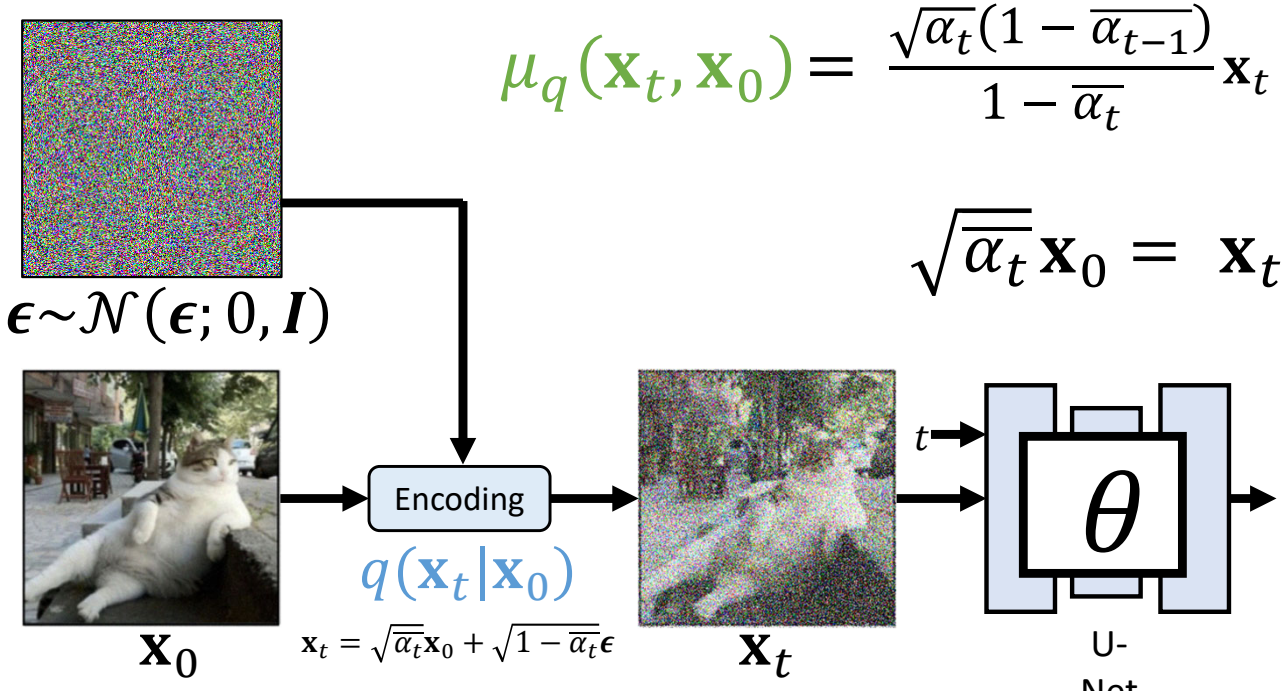
Observation #2



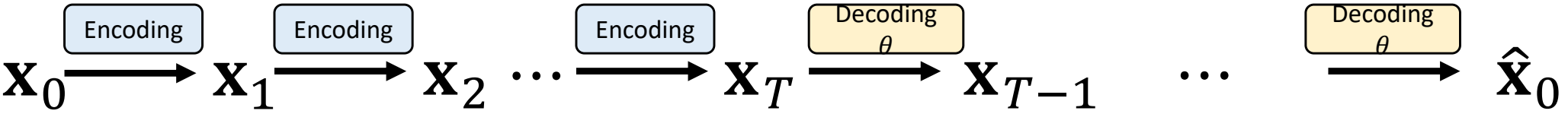
$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\; w(t) \; \left[\| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) \|_2^2 \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha_{t-1}})}{1-\overline{\alpha_t}}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha_{t-1}}}(1-\alpha_t)}{1-\overline{\alpha_t}}\mathbf{x}_0$$

$$\sqrt{\overline{\alpha_t}}\mathbf{x}_0 = \mathbf{x}_t - \sqrt{1-\overline{\alpha_t}}\boldsymbol{\epsilon}$$



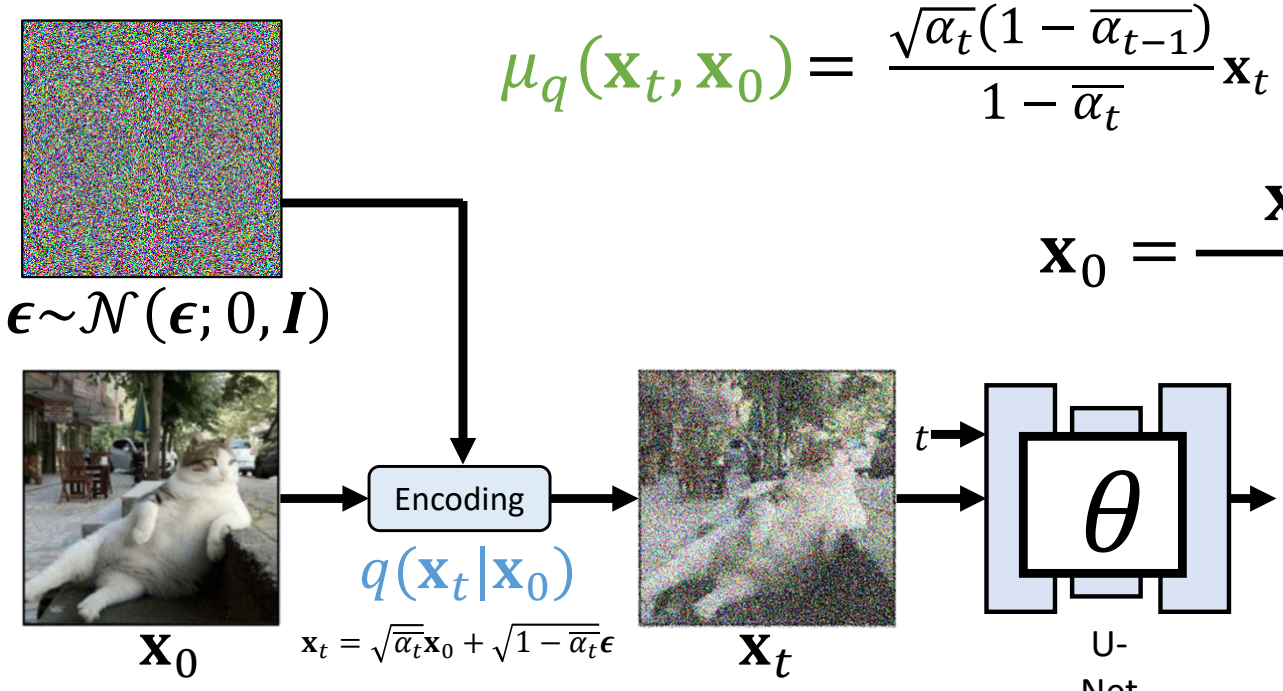
Observation #2



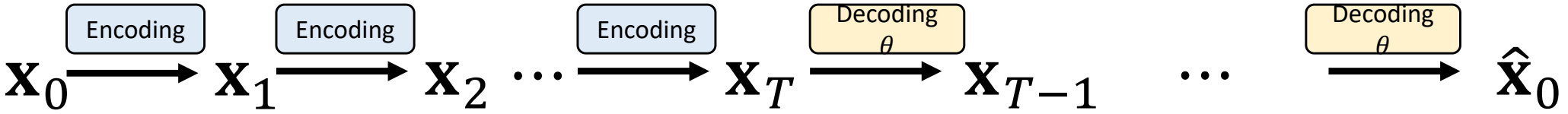
$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\; w(t) \; \left[\|\mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0)\|_2^2 \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha_{t-1}})}{1-\overline{\alpha_t}}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha_{t-1}}}(1-\alpha_t)}{1-\overline{\alpha_t}}\mathbf{x}_0$$

$$\mathbf{x}_0 = \frac{\mathbf{x}_t - \sqrt{1-\overline{\alpha_t}}\boldsymbol{\epsilon}}{\sqrt{\overline{\alpha_t}}}$$

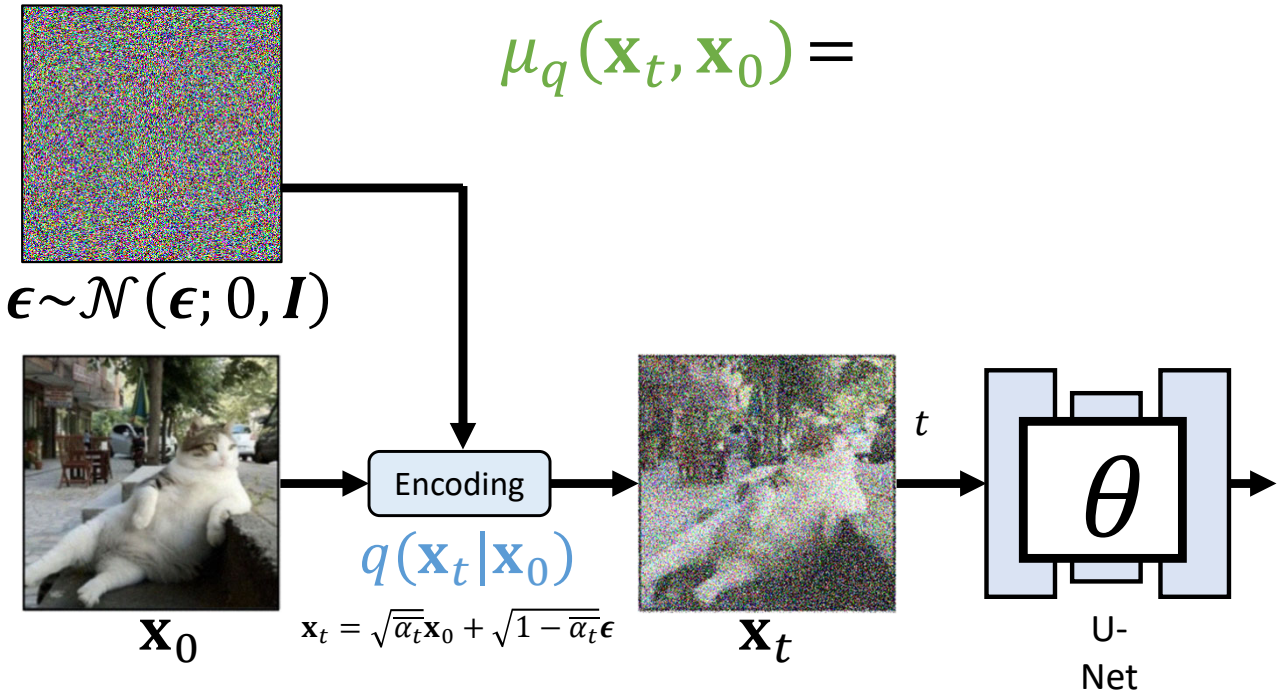


Observation #2

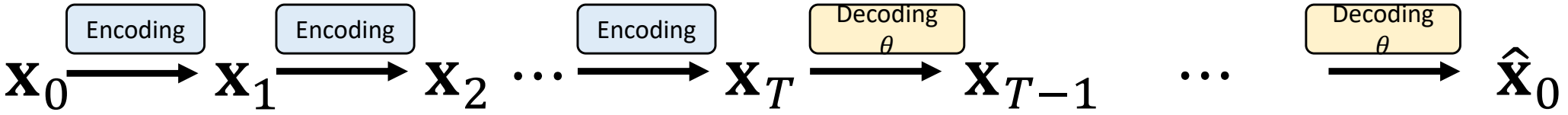


$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\; w(t) \; \left[\| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) \|_2^2 \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) =$$



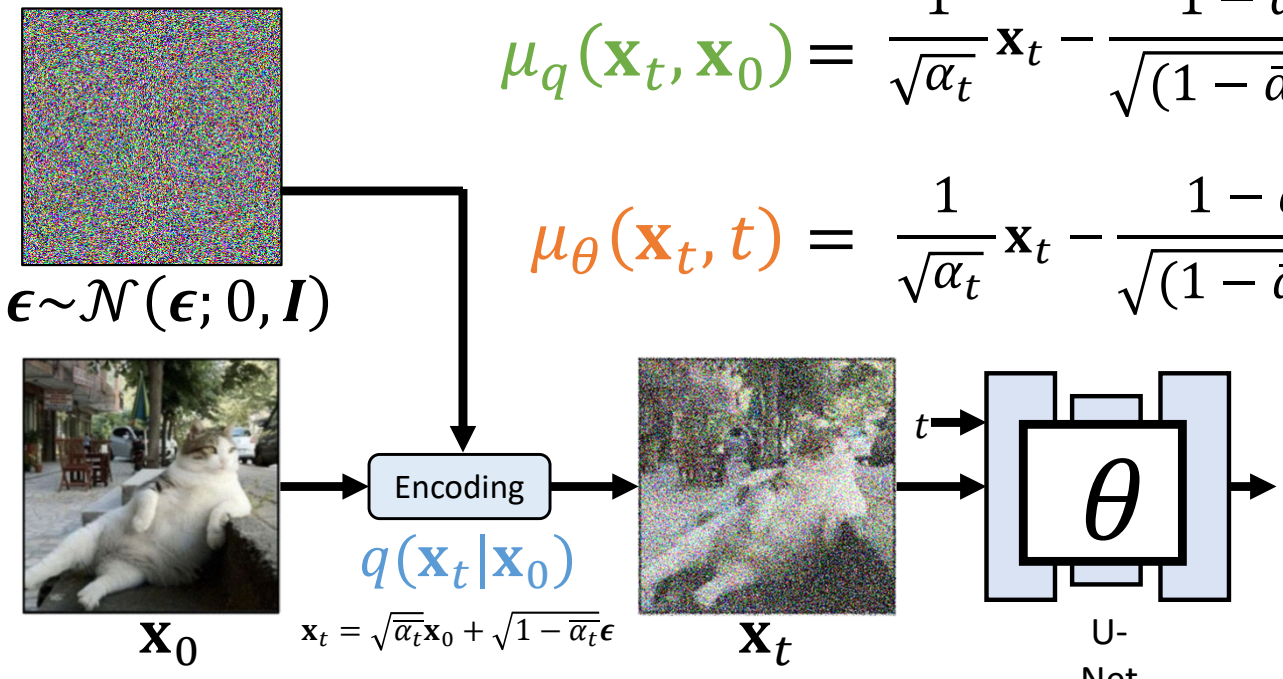
Observation #2



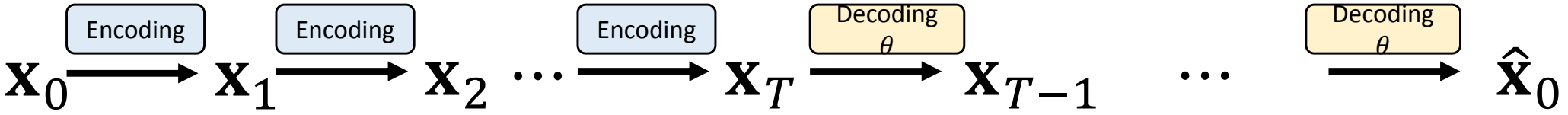
$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\; w(t) \; \left[\left|\left| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) \right|\right|_2^2 \right] \right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\epsilon$$

$$\mu_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\hat{\epsilon}_\theta(\mathbf{x}_t, t)$$



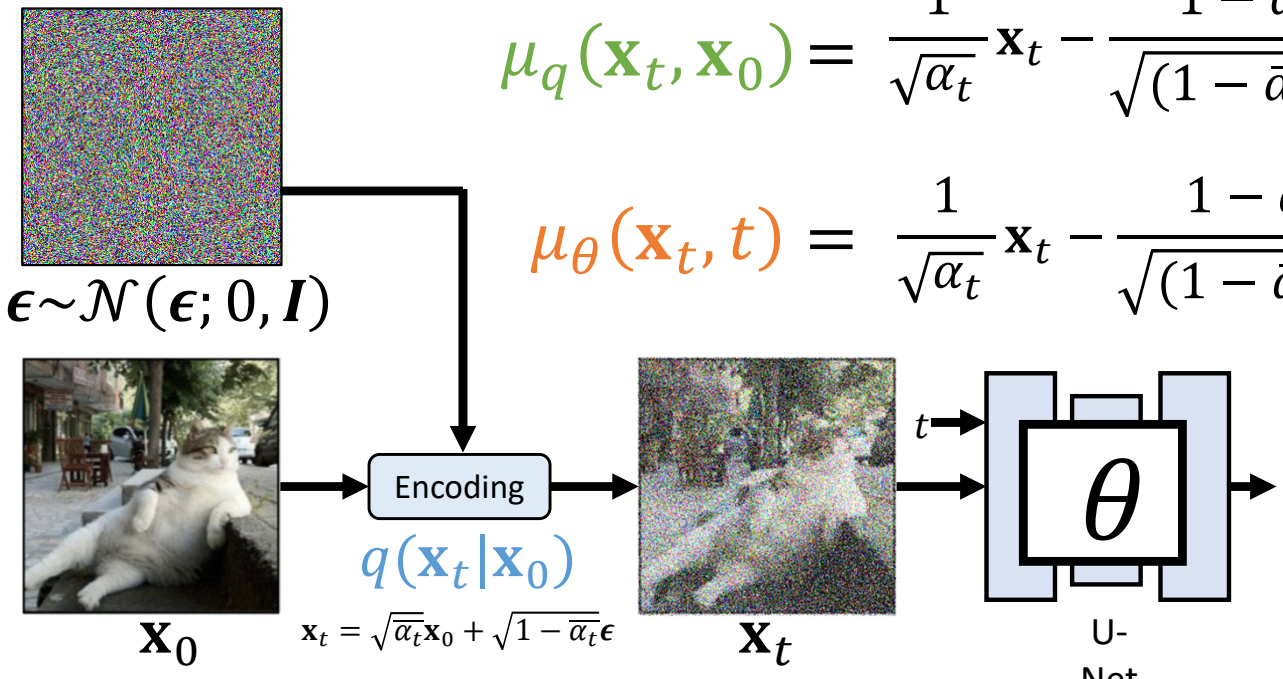
Observation #2



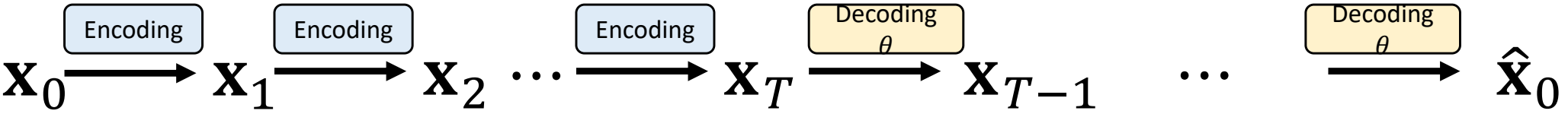
$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\; w(t) \; \left[\, \| \mu_\theta(\mathbf{x}_t, t) - \mu_q(\mathbf{x}_t, \mathbf{x}_0) \|_2^2 \,\right]\right]$$

$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\boldsymbol{\epsilon}$$

$$\mu_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$$



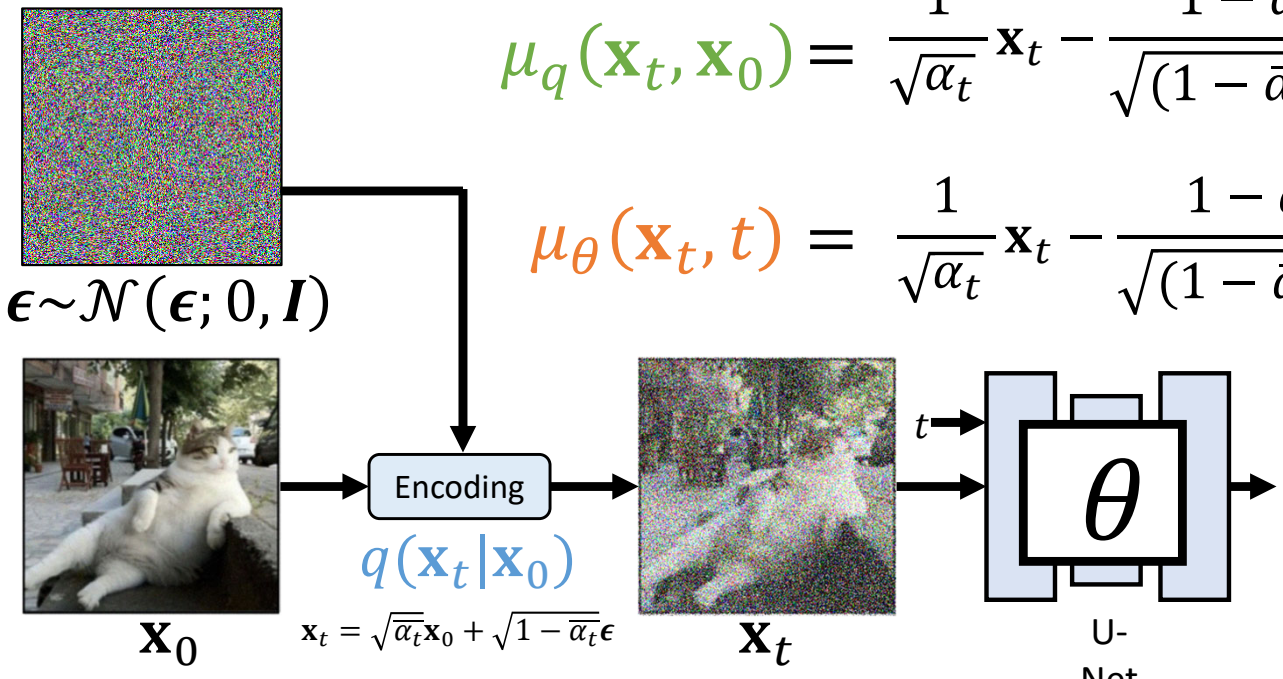
Observation #2



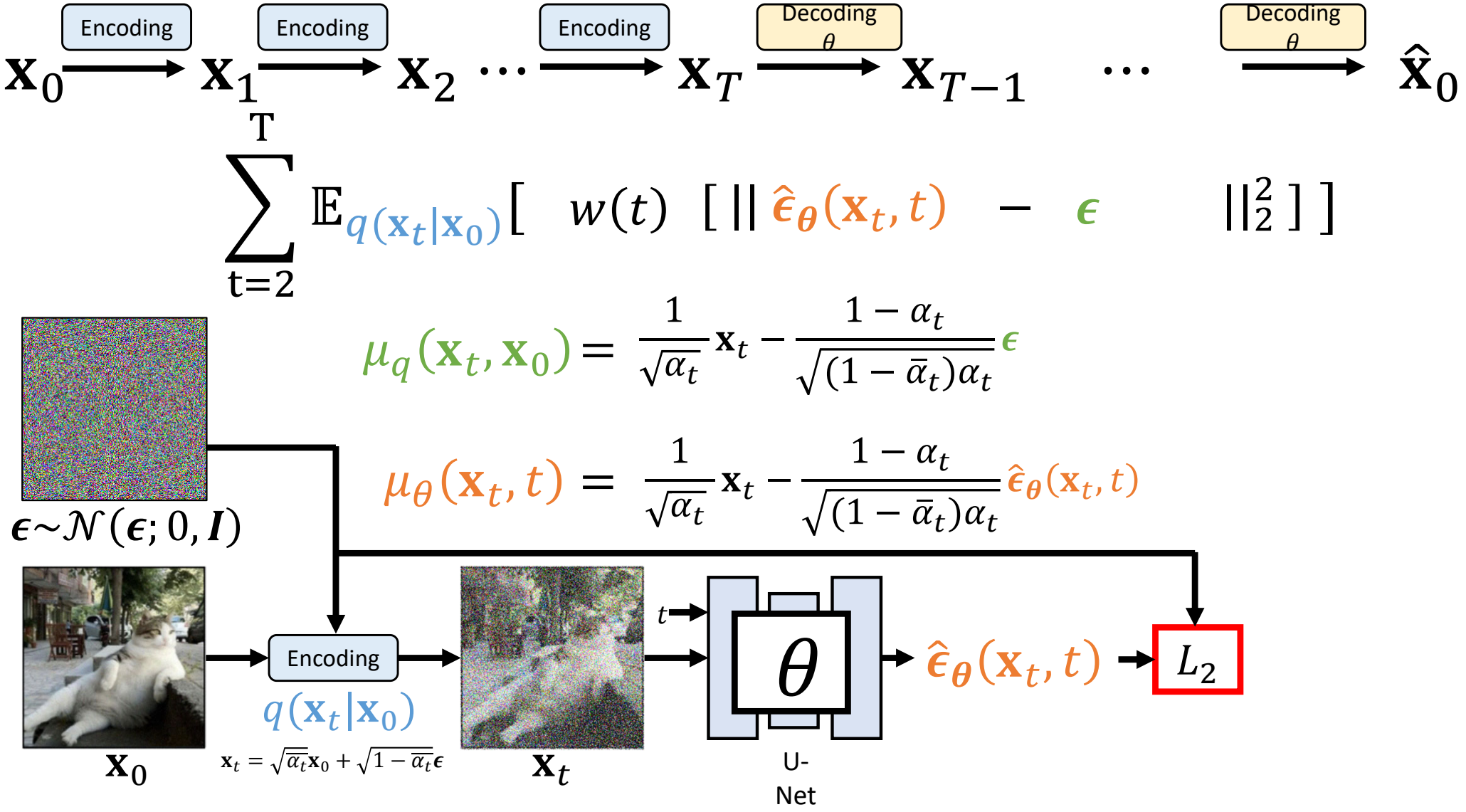
$$\sum_{t=2}^{T} \mathbb{E}_{q(\mathbf{x}_t|\mathbf{x}_0)}\left[\; w(t) \; \left[\left\| \hat{\epsilon}_\theta(\mathbf{x}_t, t) - \epsilon \right\|_2^2 \right] \right]$$

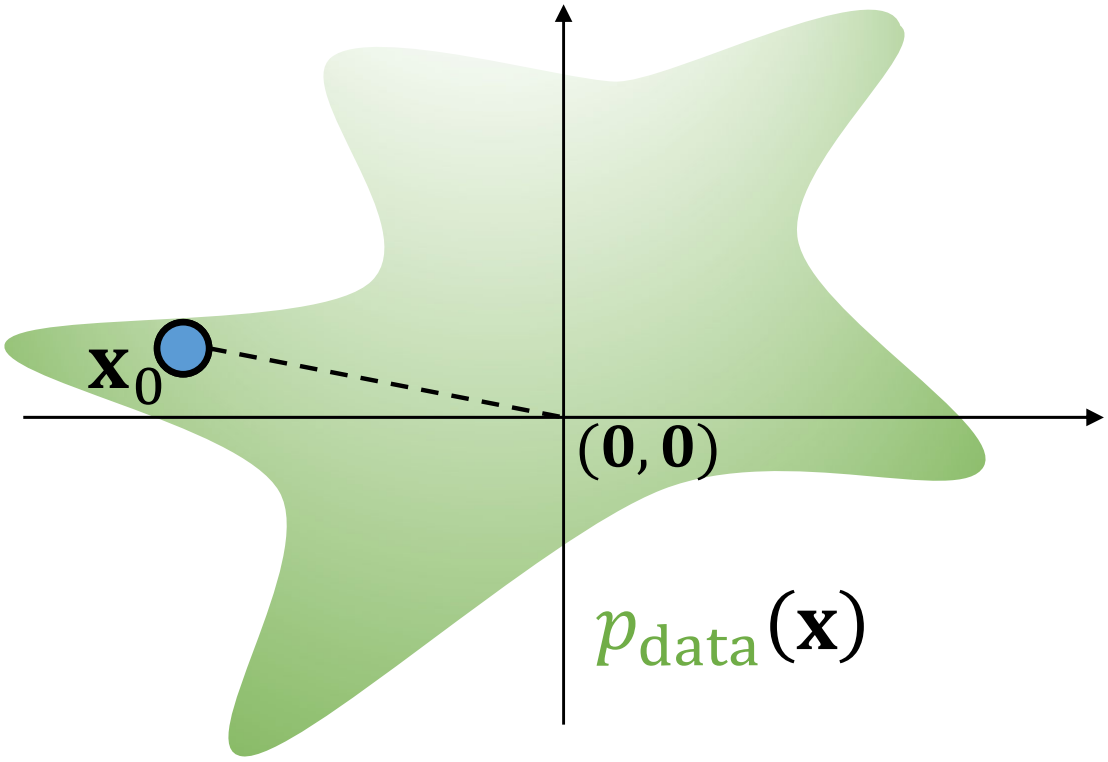
$$\mu_q(\mathbf{x}_t, \mathbf{x}_0) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\epsilon$$

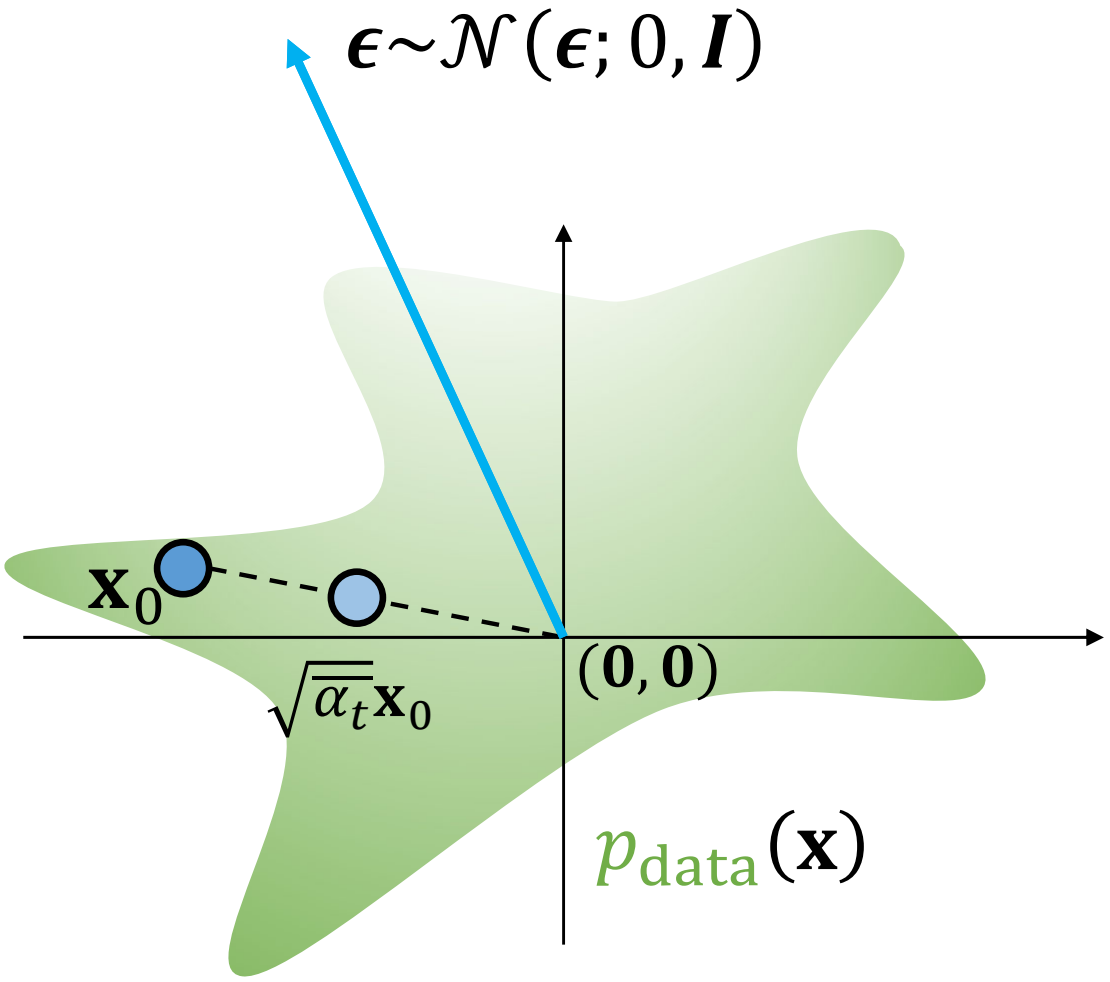
$$\mu_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\hat{\epsilon}_\theta(\mathbf{x}_t, t)$$

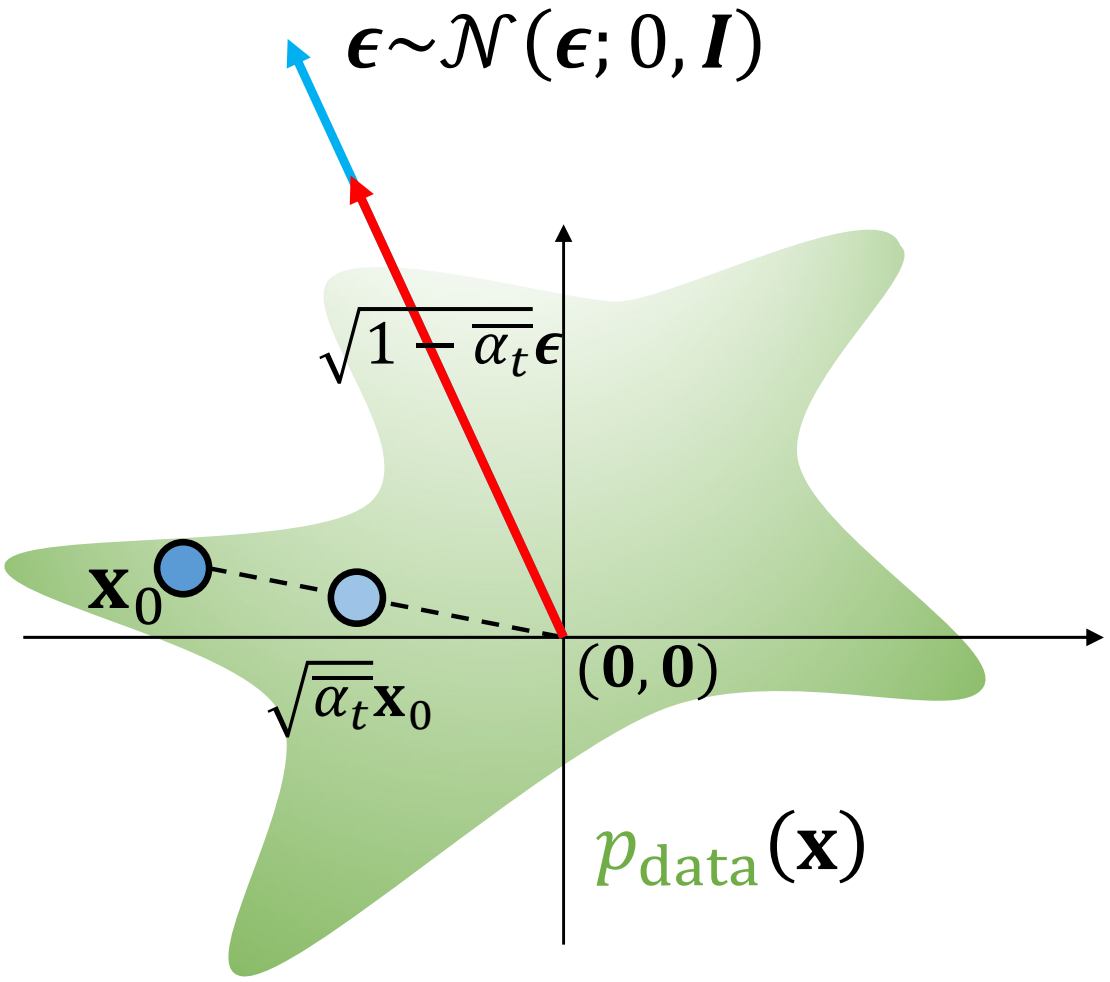


Observation #2

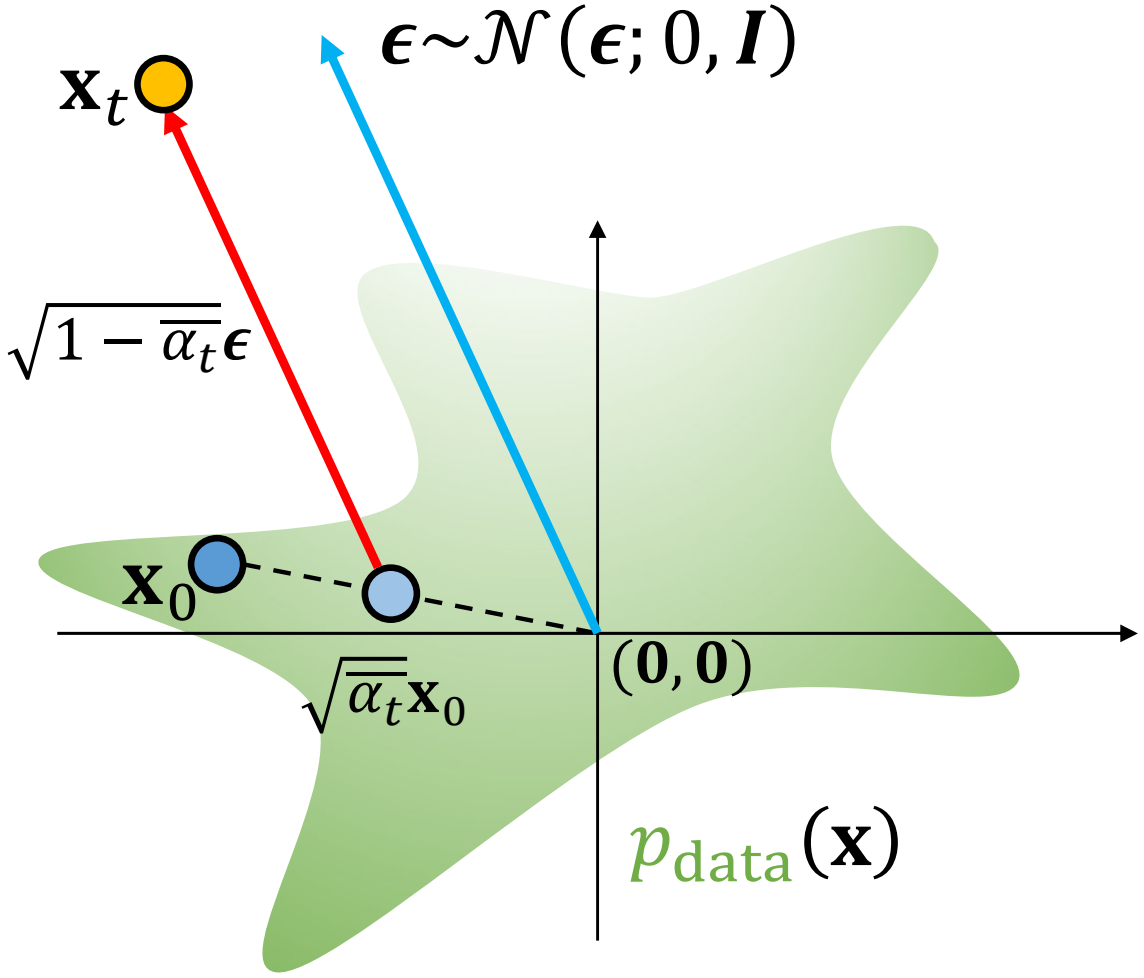



$\mathbf{x}_0$
$(\mathbf{0}, \mathbf{0})$
$p_{\text{data}}(\mathbf{x})$

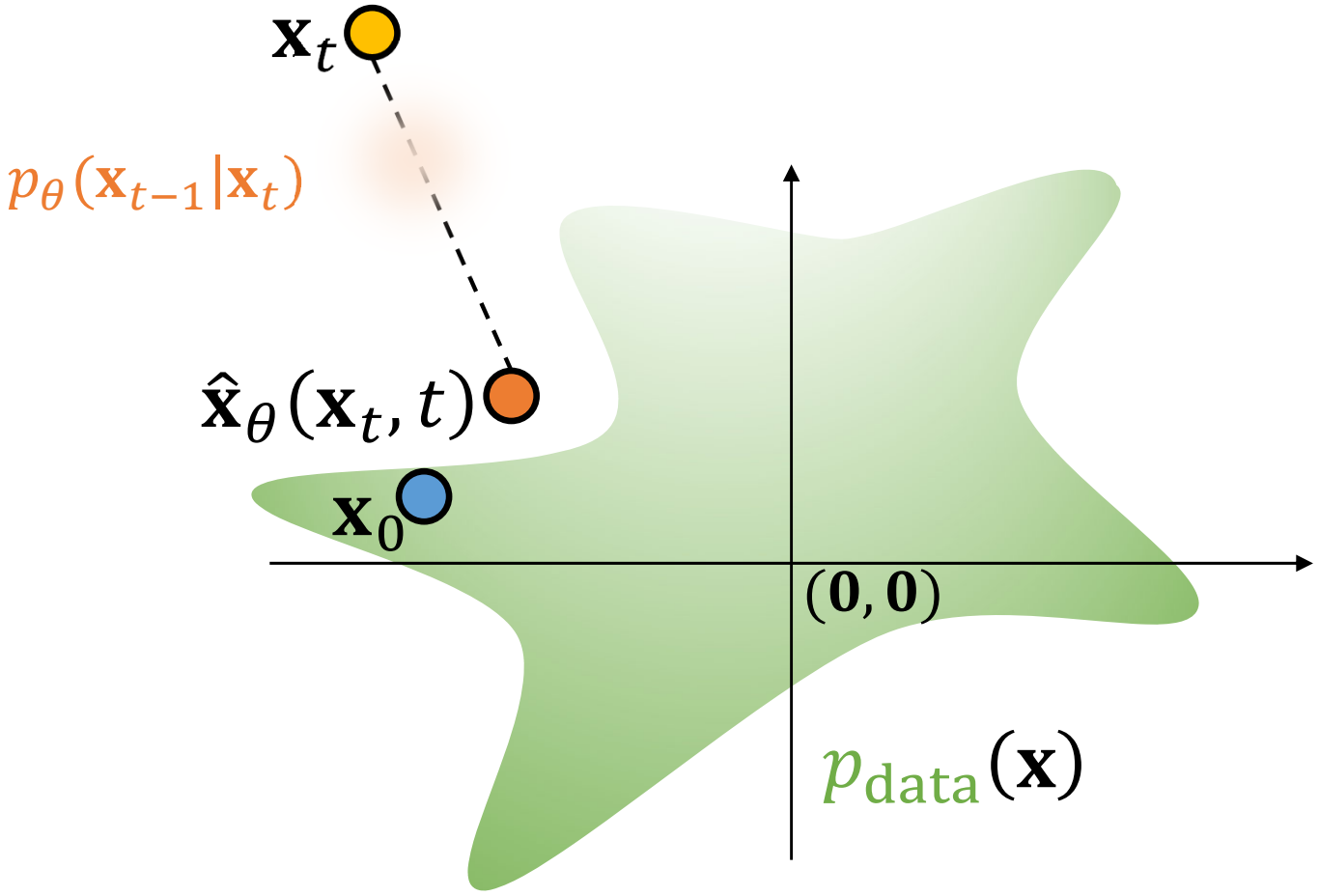

$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \boldsymbol{I})$
$\mathbf{x}_0$
$\sqrt{\overline{\alpha}_t}\mathbf{x}_0$
$(\mathbf{0}, \mathbf{0})$
$p_{\text{data}}(\mathbf{x})$


$\boldsymbol{\epsilon} \sim \mathcal{N}(\boldsymbol{\epsilon}; 0, \boldsymbol{I})$
$\sqrt{1-\overline{\alpha_t}}\boldsymbol{\epsilon}$
$\mathbf{x}_0$
$\sqrt{\overline{\alpha_t}}\mathbf{x}_0$
$(\mathbf{0}, \mathbf{0})$
$p_{\text{data}}(\mathbf{x})$

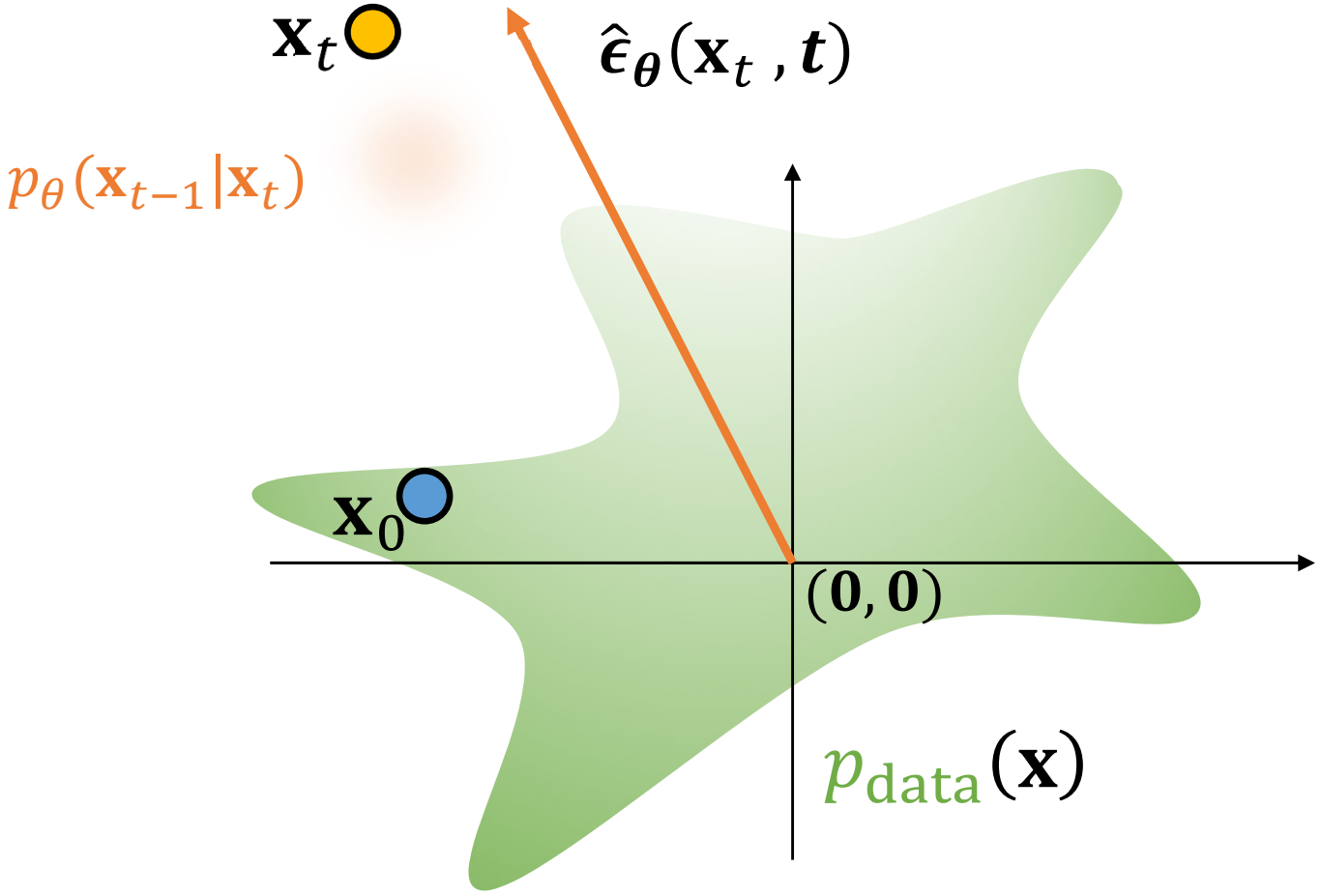
$$\mathbf{x}_t = \sqrt{\overline{\alpha_t}}\mathbf{x}_0 + \sqrt{1-\overline{\alpha_t}}\boldsymbol{\epsilon}$$



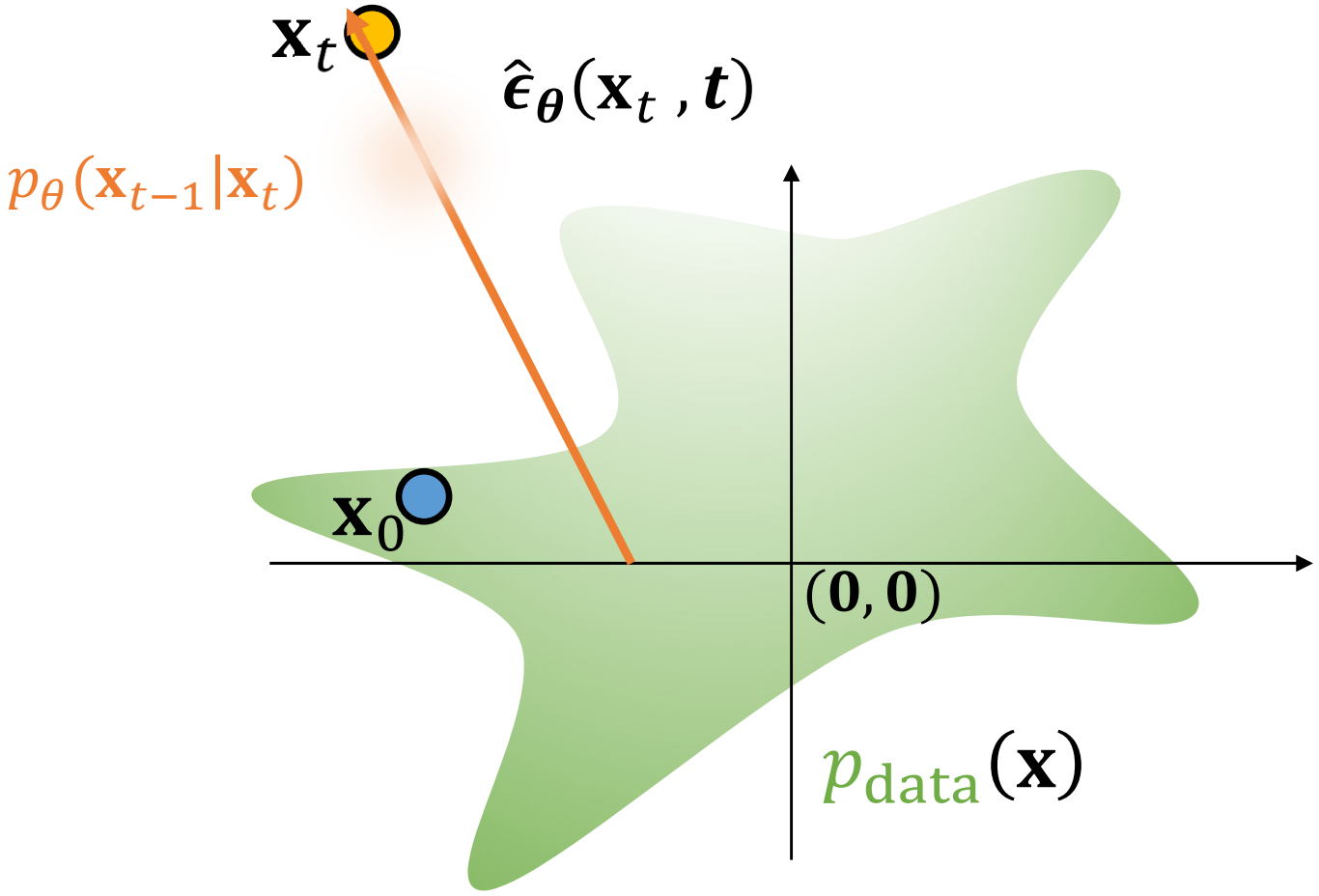
$$\mu_\theta(\mathrm{x}_t, t) = \frac{\sqrt{\alpha_t}(1-\overline{\alpha_{t-1}})}{1-\overline{\alpha_t}}\mathbf{x}_t + \frac{\sqrt{\overline{\alpha_{t-1}}}(1-\alpha_t)}{1-\overline{\alpha_t}}\hat{\mathbf{x}}_\theta(\mathbf{x}_t, t)$$



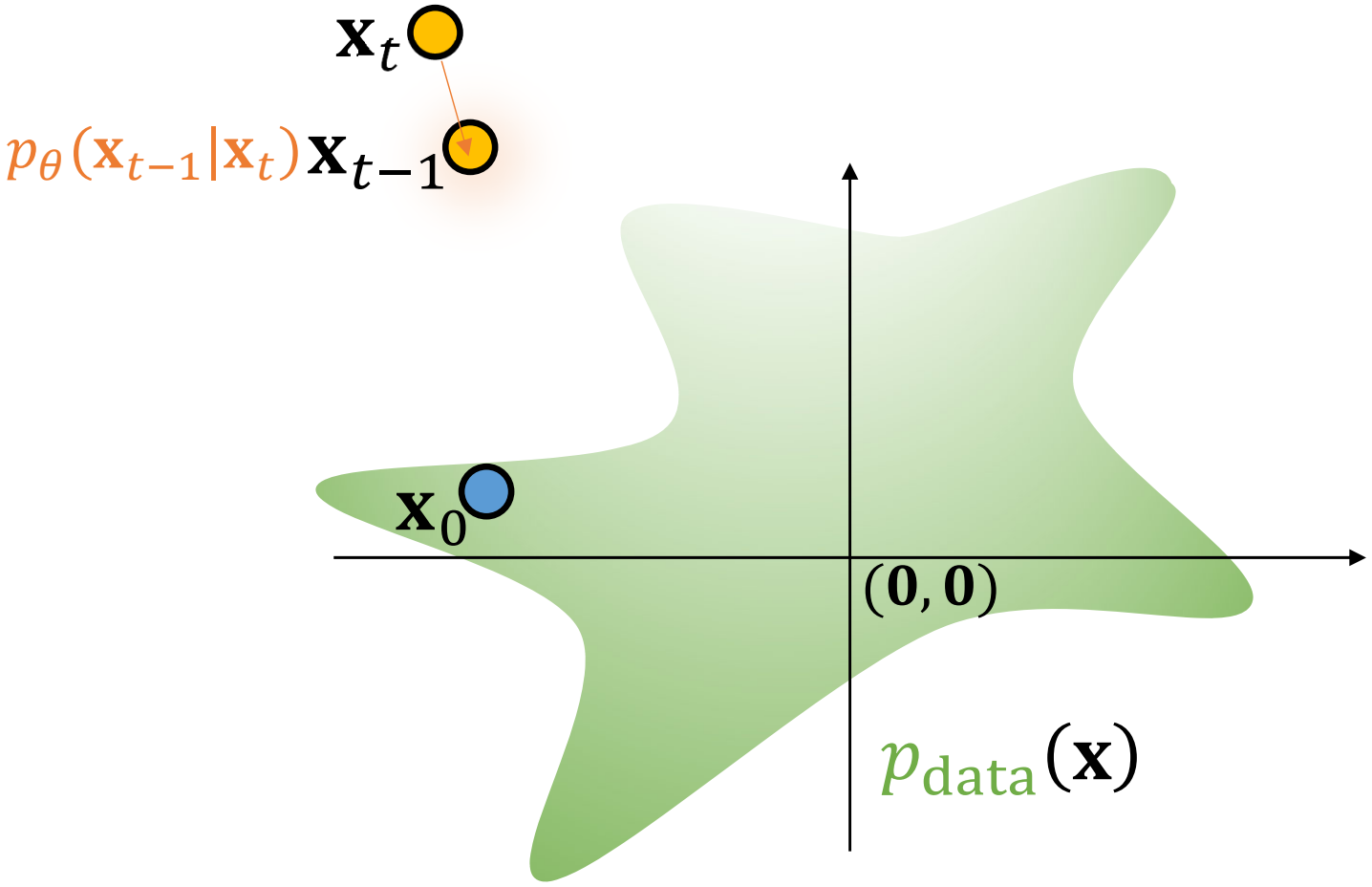
$$\mu_\theta(\mathrm{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$$



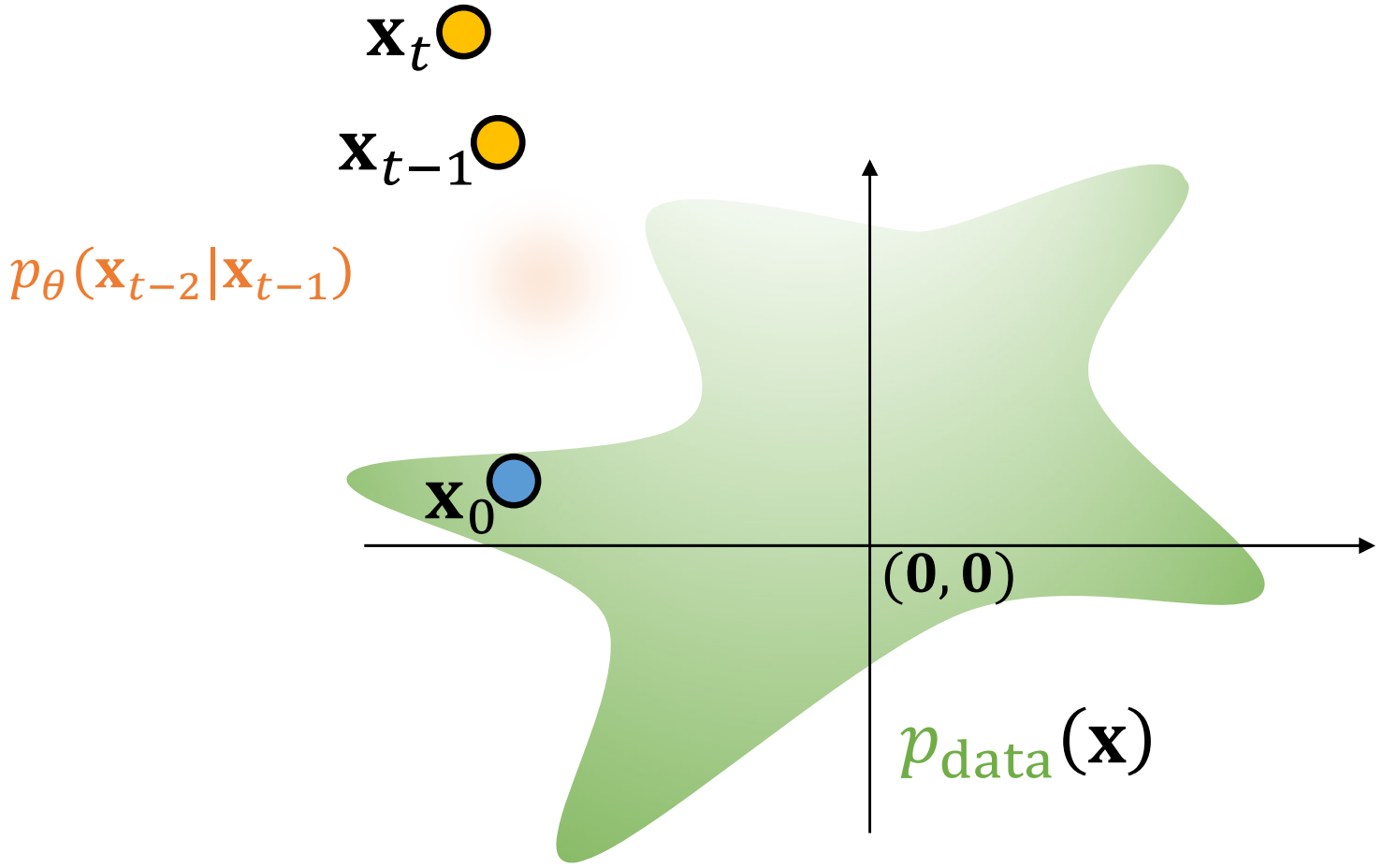
$$\mu_\theta(\mathrm{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$$



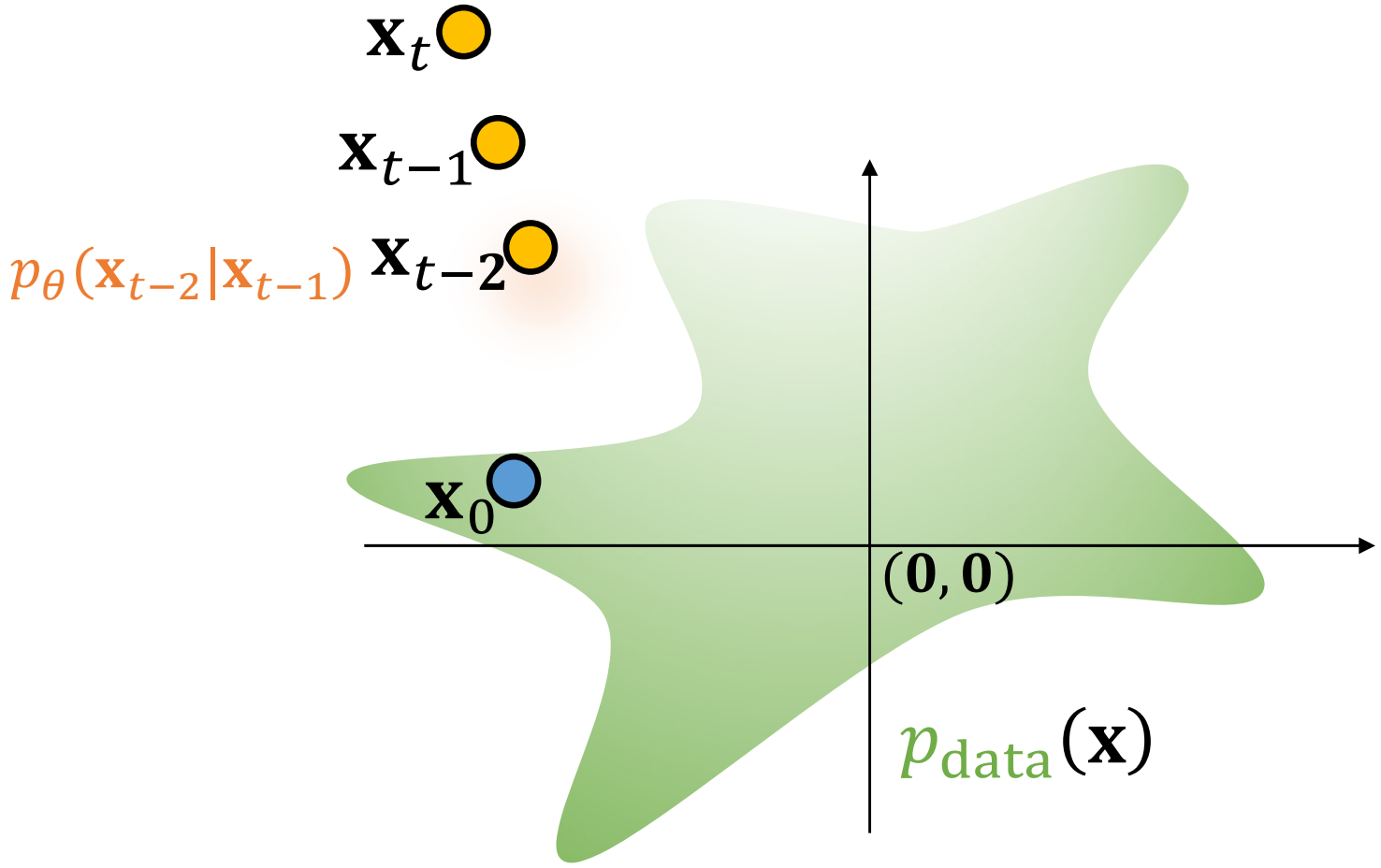
$$\mu_\theta(\mathrm{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$$



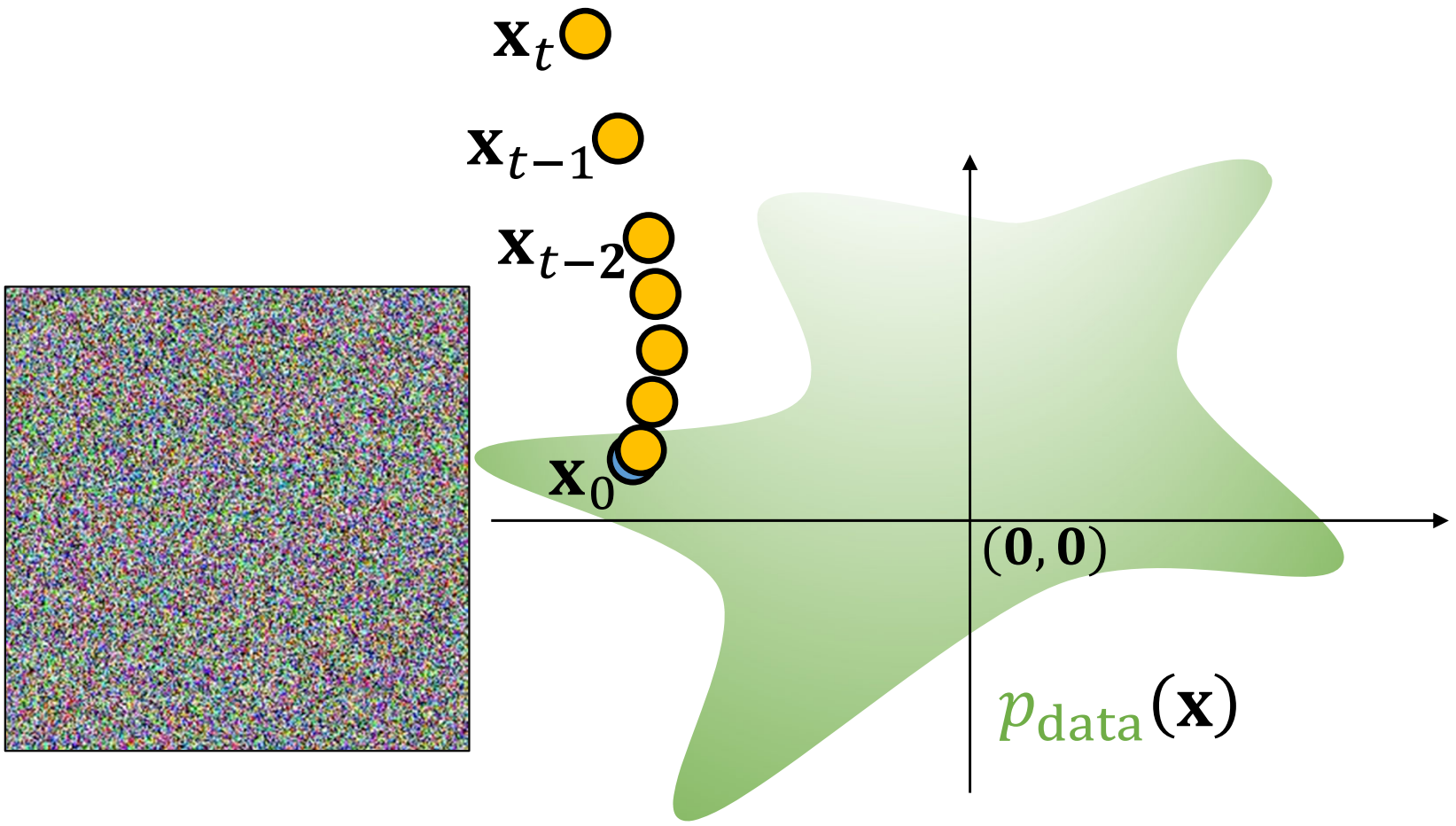
$$\mu_\theta(\mathrm{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$$



$$\mu_\theta(\mathrm{x}_t, t) = \frac{1}{\sqrt{\alpha_t}}\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{(1-\bar{\alpha}_t)\alpha_t}}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\mathbf{x}_t, t)$$



$\mu_\theta(\mathrm{x}_t, t) =$



Training vs. Inference

- Summary

Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \text{Uniform}(\{1, \dots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Take gradient descent step on
$$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$$
6: **until** converged

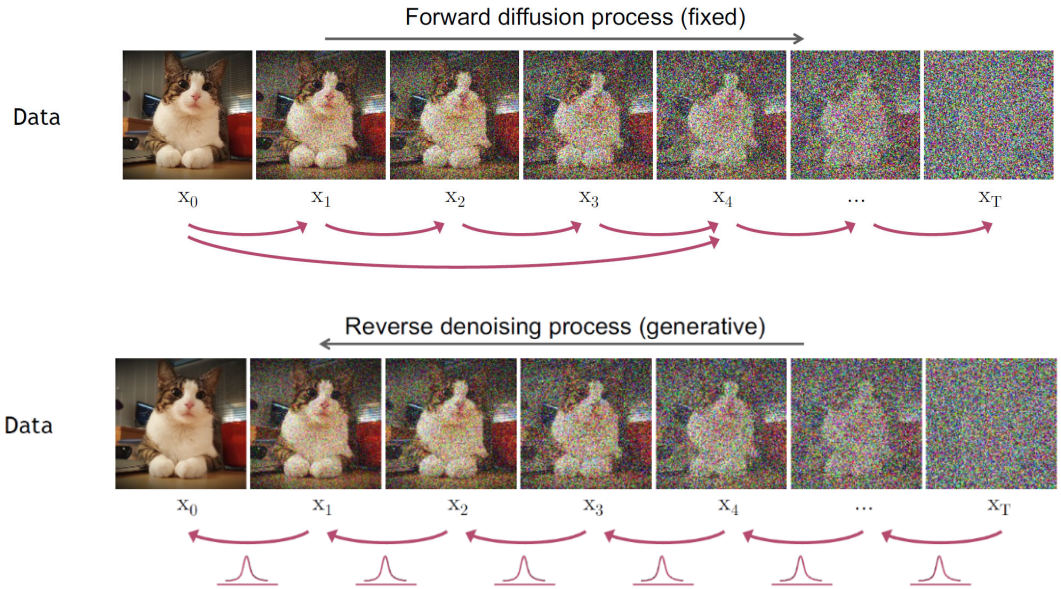
x_t

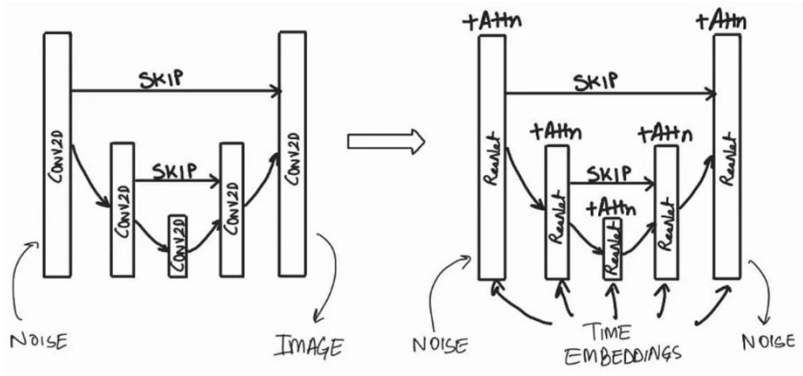
Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \dots, 1$ **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

allowing generation diversity

$$x_t \xrightarrow{model} \epsilon_\theta(x_t, t) \xrightarrow{P(x_t|x_0) \to P(x_0|x_t, \epsilon_\theta)} \hat{x}_0(x_t, \epsilon_\theta) \longrightarrow \mu(x_t, \hat{x}_0), \beta_t \xrightarrow{P(x_{t-1}|x_t, x_0)} \hat{x}_{t-1}$$





https://medium.com/@vedantjumle/class-conditioned-diffusion-models-using-keras-and-tensorflow-9997fa6d958c

Slide credit: Kreis, Gao, & Vahdat

Training vs. Inference

- Summary

Algorithm 1 Training

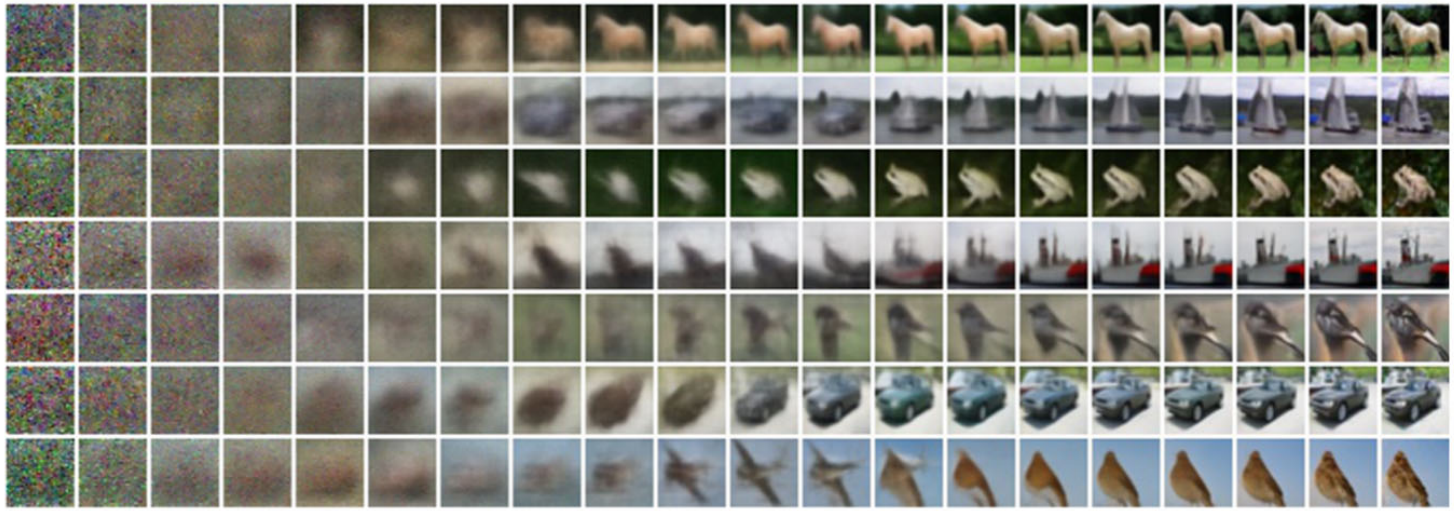
1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \text{Uniform}(\{1, \dots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Take gradient descent step on
$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$
6: **until** converged

x_t

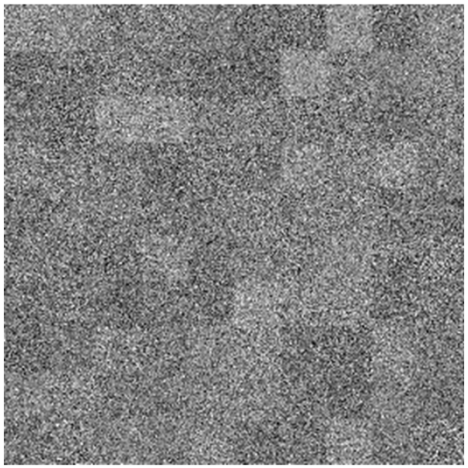
Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \dots, 1$ **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

allowing generation diversity



More steps

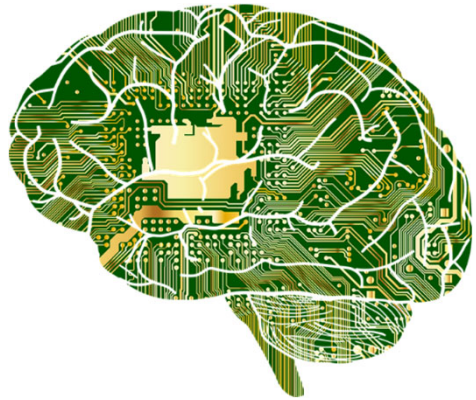


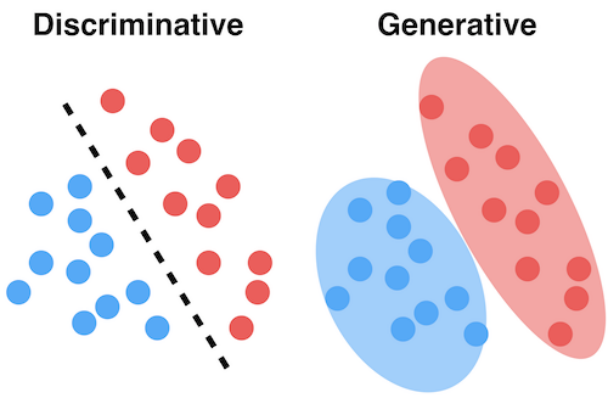
MNIST handwritten image data

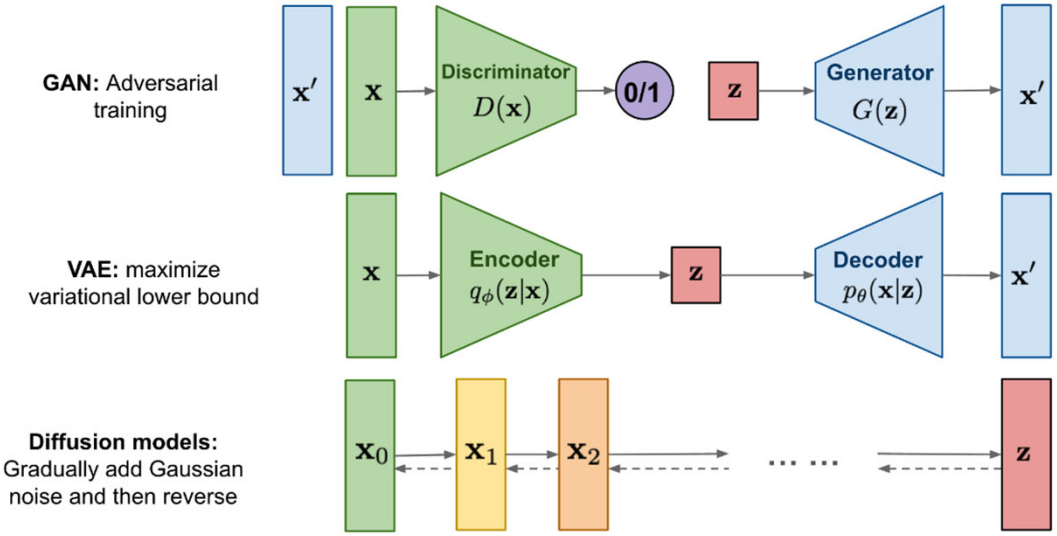
Slide credit: Kreis, Gao, & Vahdat
https://medium.com/ai-blog-tw/%E9%82%8A%E5%AF%A6%E4%BD%9C%E9%82%8A%E5%AD%B8%E7%BF%92diffusion-model-%E5%BE%9Eddpm%E7%9A%84%E7%B0%A1%E5%8C%96%E6%A6%82%E5%BF%B5%E7%90%86%E8%A7%A3-4c565a1c09c

What's to Be Covered Today…

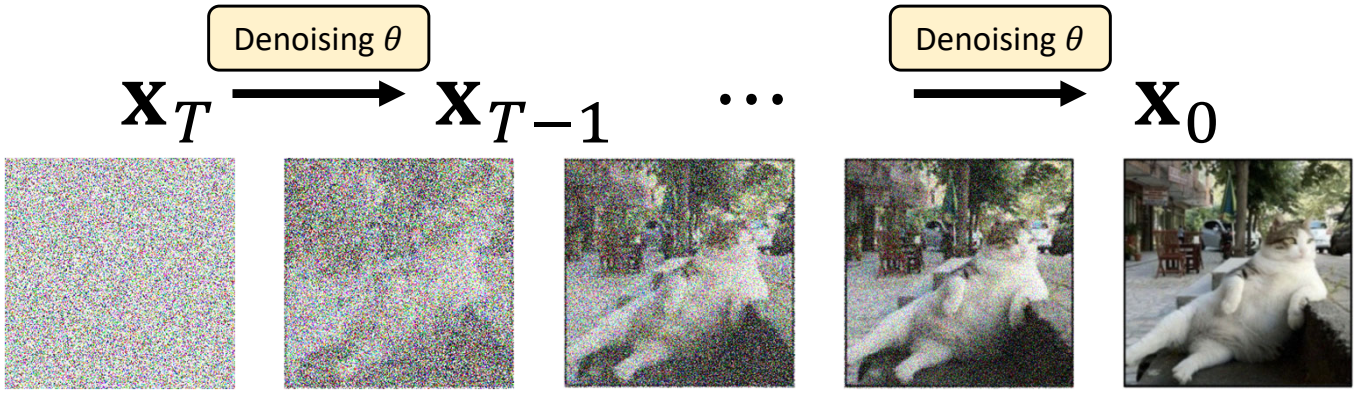
- Generative Models
  - Autoencoder
  - Variational Autoencoder
  - Diffusion Model
    - Denoising Diffusion Probabilistic Model (DDPM)
    - Latent Diffusion Model
    - Denoising Diffusion Implicit Model (DDIM)

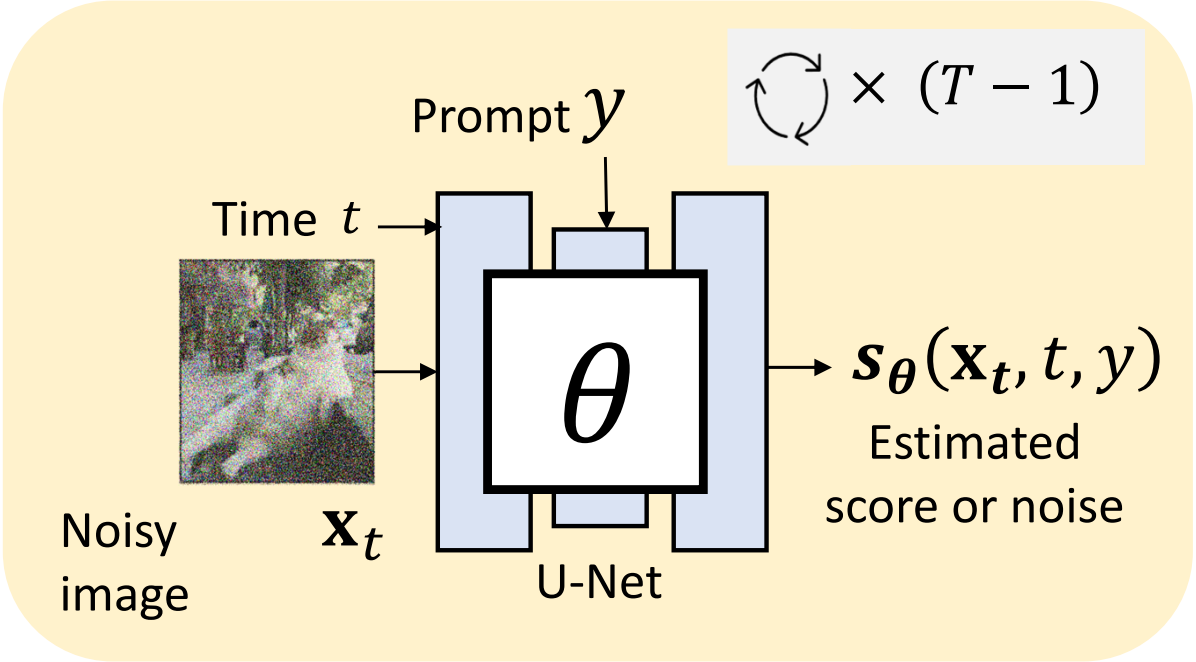




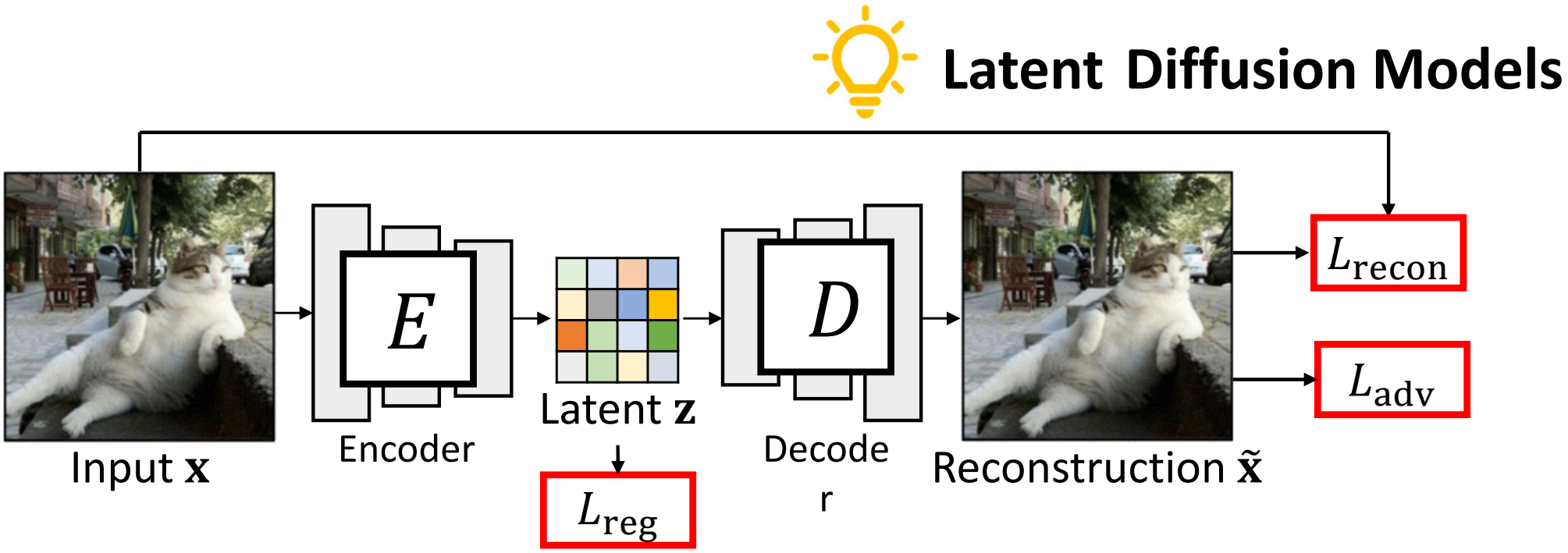


Lecture/slide credit: Prof. Jia-Bin Huang @ U Maryland
https://youtu.be/i2qSxMVeVLI?si=ZIwgo9T-6lEg3kJM


Denoising θ
$\mathbf{x}_T$
$\mathbf{x}_{T-1}$
$\cdots$
Denoising θ
$\mathbf{x}_0$

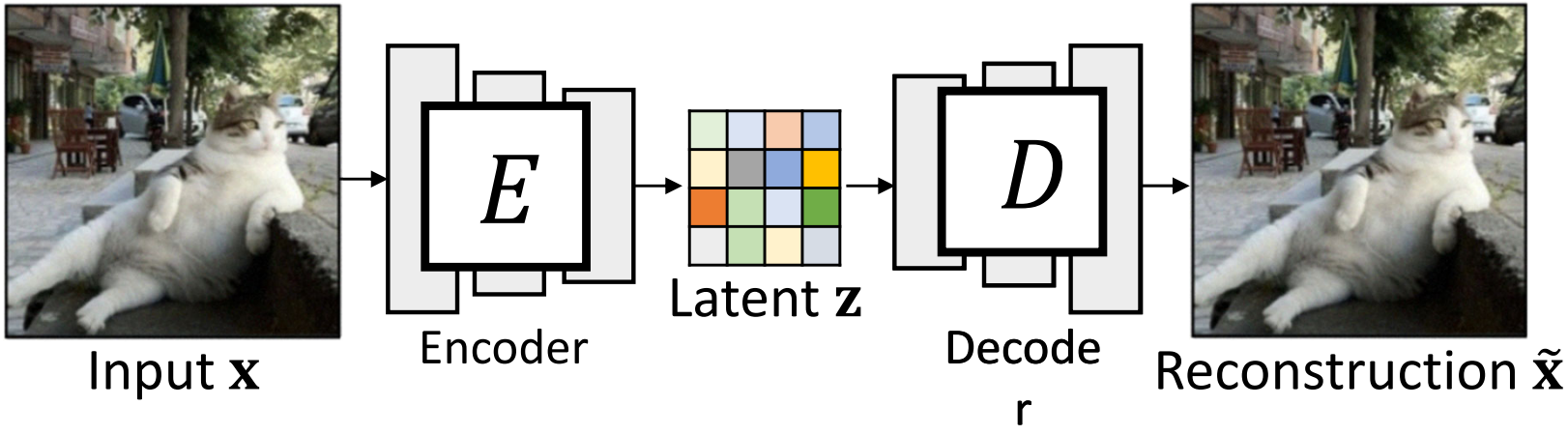

$\times\ (T-1)$
Prompt y
Time t
$\boldsymbol{\theta}$
$\boldsymbol{s_\theta(\mathbf{x}_t, t, y)}$
Estimated score or noise
Noisy image
$\mathbf{x}_t$
U-Net

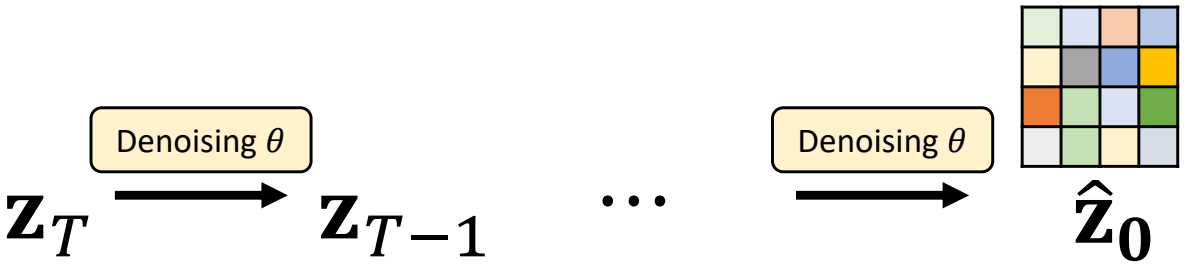
Any concern?



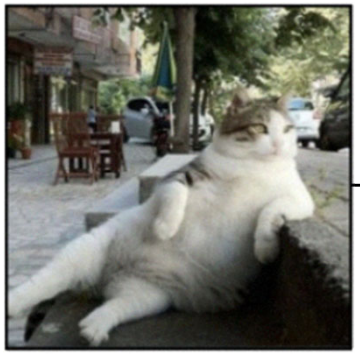
[Rombach et al. 2022] [Vahdat et al. 2021]

Latent Diffusion Models





Latent Diffusion Models



Input $\mathbf{x}$

E

Encoder

Latent $\mathbf{z}$

D

Decoder

Reconstruction $\tilde{\mathbf{x}}$

Denoising θ

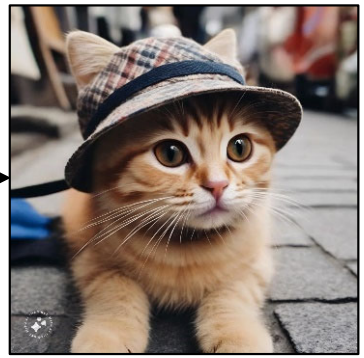
$\mathbf{z}_T$ → $\mathbf{z}_{T-1}$ ⋯

Denoising θ

$\hat{\mathbf{z}}_0$

D

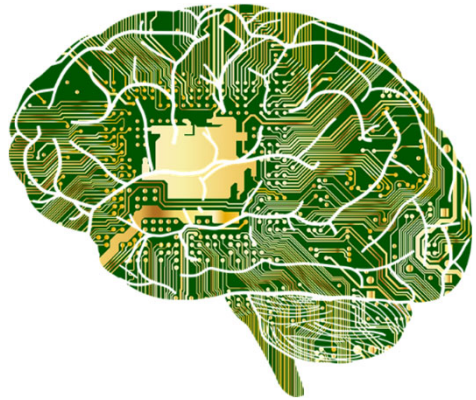
Decoder

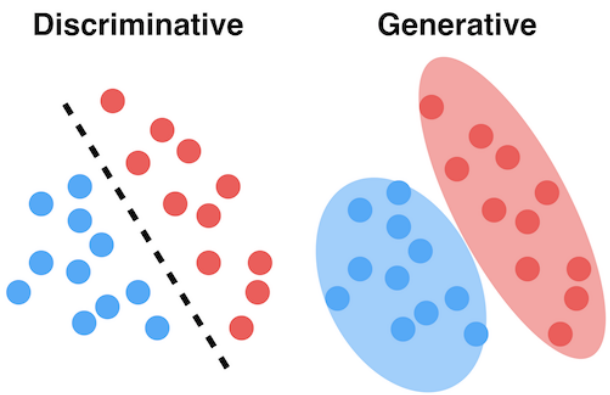


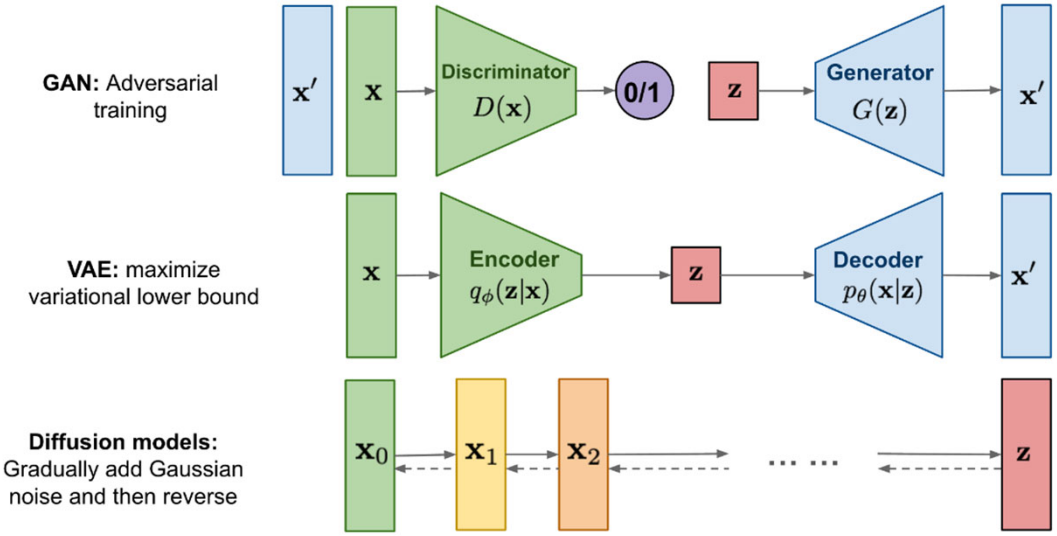
Sample

What's to Be Covered Today…

- Generative Models
  - Autoencoder
  - Variational Autoencoder
  - Diffusion Model
    - Denoising Diffusion Probabilistic Model (DDPM)
    - Latent Diffusion Model
    - Denoising Diffusion Implicit Model (DDIM)







Lecture/slide credit: Prof. Jia-Bin Huang @ U Maryland
https://youtu.be/i2qSxMVeVLI?si=ZIwgo9T-6lEg3kJM

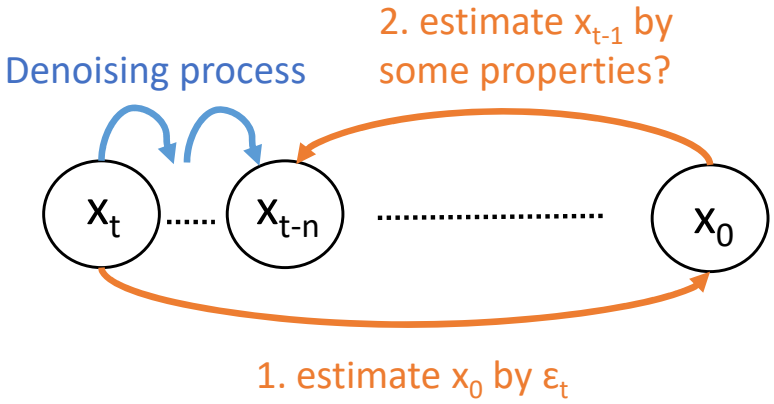
From DDPM to DDIM:
Denoising Diffusion Implict Models (Song *et al.*, ICLR'21)

- Song et al. 2020:
  "*It takes around 20 hours to sample 50k images of size 32 × 32 pixels from a DDPM, but less than a minute to do so from a GAN on an Nvidia 2080 Ti GPU.*"

- Motivation/Goals
  - Accelerating sampling process
  - A deterministic process
    - Inversion
    - Manipulation
    - Interpolation
    - …

- Applications
  - Stable Diffusion

From DDPM to DDIM:
Denoising Diffusion Implict Models



- DDPM
  - Prosn
    - High quality image generation without adversarial training.
  - Cons
    - Require simulating a **Markov chain** for many steps in order to produce a sample. (i.e., x_t relies on x_{t-1})
- DDIM

3.1 Non-Markovian forward processes

Let us consider a family $\mathcal{Q}$ of inference distributions, indexed by a real vector $\sigma \in \mathbb{R}^T_{\geq 0}$:

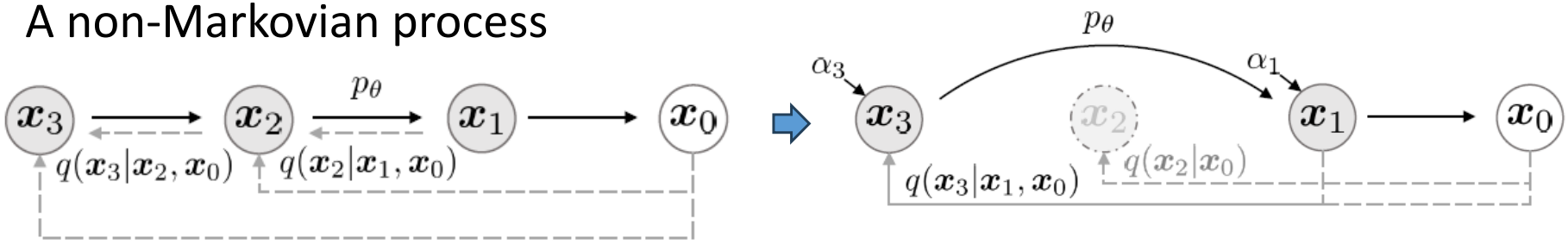
$$q_\sigma(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0) := q_\sigma(\boldsymbol{x}_T|\boldsymbol{x}_0)\prod_{t=2}^{T} q_\sigma(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) \tag{6}$$

where $q_\sigma(\boldsymbol{x}_T|\boldsymbol{x}_0) = \mathcal{N}(\sqrt{\alpha_T}\boldsymbol{x}_0, (1-\alpha_T)\boldsymbol{I})$ and for all $t > 1$, ϵ_t ϵ

$$q_\sigma(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) = \mathcal{N}\left(\sqrt{\alpha_{t-1}}\boldsymbol{x}_0 + \sqrt{1-\alpha_{t-1}-\sigma_t^2}\cdot\frac{\boldsymbol{x}_t - \sqrt{\alpha_t}\boldsymbol{x}_0}{\sqrt{1-\alpha_t}}, \sigma_t^2\boldsymbol{I}\right). \tag{7}$$

From DDPM to DDIM:
Denoising Diffusion Implict Models

- DDIM
  - A non-Markovian process



    - Non-Markovian forward process

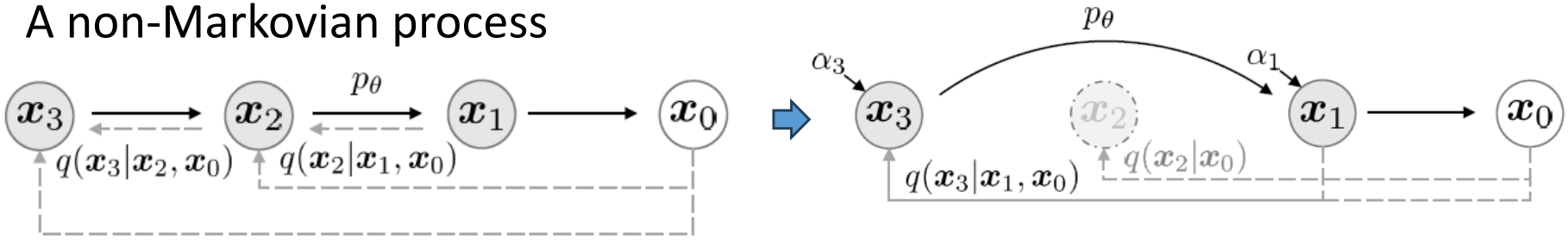
$$q_\sigma(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0) := q_\sigma(\boldsymbol{x}_T|\boldsymbol{x}_0)\prod_{t=2}^{T} q_\sigma(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0)$$

$$\boldsymbol{x}_t = \sqrt{\alpha_t}\boldsymbol{x}_0 + \sqrt{1-\alpha_t}\epsilon, \quad \text{where} \quad \epsilon \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I})$$

$$q_\sigma(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) = \mathcal{N}\left(\sqrt{\alpha_{t-1}}\boldsymbol{x}_0 + \sqrt{1-\alpha_{t-1}-\sigma_t^2}\cdot\frac{\boldsymbol{x}_t - \sqrt{\alpha_t}\boldsymbol{x}_0}{\sqrt{1-\alpha_t}}, \sigma_t^2\boldsymbol{I}\right)$$

From DDPM to DDIM:
Denoising Diffusion Implict Models

- DDIM
  - A non-Markovian process



  - Generative (denoising) process

From the forward process

$$\boldsymbol{x}_t = \sqrt{\alpha_t}\boldsymbol{x}_0 + \sqrt{1-\alpha_t}\epsilon, \quad \text{where} \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$$

Denoised observation

$$f_\theta^{(t)}(\boldsymbol{x}_t) := (\boldsymbol{x}_t - \sqrt{1-\alpha_t} \cdot \epsilon_\theta^{(t)}(\boldsymbol{x}_t))/\sqrt{\alpha_t}$$

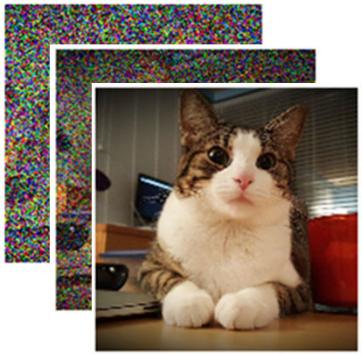
$$\Rightarrow \quad p_\theta^{(t)}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t) = \begin{cases} \mathcal{N}(f_\theta^{(1)}(\boldsymbol{x}_1), \sigma_1^2 \boldsymbol{I}) & \text{if } t = 1 \\ q_\sigma(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, f_\theta^{(t)}(\boldsymbol{x}_t)) & \text{otherwise} \end{cases}$$

  - The above forward/denoising processes results in the same training objectives as those of DDPM (see DDIM paper)

From DDPM to DDIM:
Denoising Diffusion Implict Models

- DDIM
  - Sampling process for generation

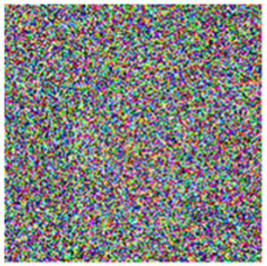
$$\boldsymbol{x}_{t-1} = \sqrt{\alpha_{t-1}} \underbrace{\left(\frac{\boldsymbol{x}_t - \sqrt{1-\alpha_t}\epsilon_\theta^{(t)}(\boldsymbol{x}_t)}{\sqrt{\alpha_t}} \right)}_{\text{“ predicted } \boldsymbol{x}_0\text{”}} + \underbrace{\sqrt{1-\alpha_{t-1}-\sigma_t^2} \cdot \epsilon_\theta^{(t)}(\boldsymbol{x}_t)}_{\text{“direction pointing to } \boldsymbol{x}_t\text{”}} + \underbrace{\sigma_t \epsilon_t}_{\text{random noise}}$$



Step 0

Step 123

Step 456



Step 999

- Additional comment on σ_t^2 : *stochastic* vs. *deterministic* generation process
- Since DDIM and DDPM share the same objective function,
  so one can use a ***pretrained*** DDPM for DDIM generation.

FID (Fréchet inception distance) score:

	CIFAR10 (32 × 32)					CelebA (64 × 64)				
	10	20	50	100	1000	10	20	50	100	1000
DDIM	**13.36**	**6.84**	**4.67**	**4.16**	4.04	**17.33**	**13.73**	**9.17**	**6.53**	3.51
DDPM	367.43	133.37	32.72	9.99	**3.17**	299.71	183.83	71.71	45.20	**3.26**

DDIM [Song et al. 2021]

From DDPM to DDIM

Example results

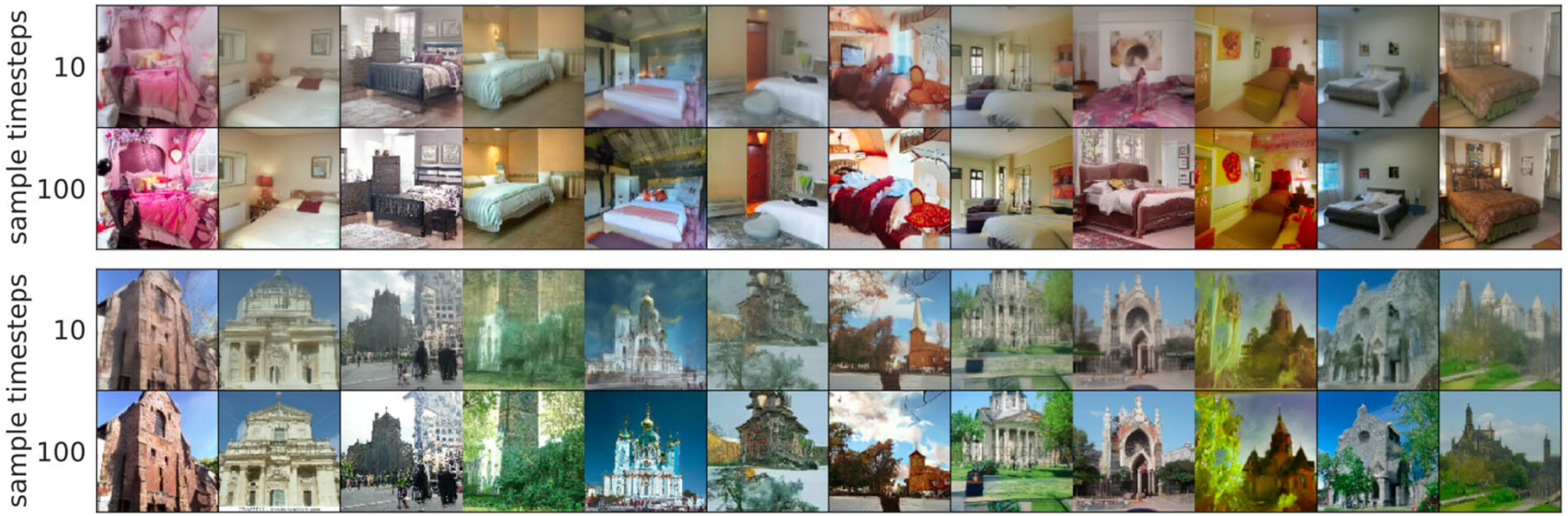


Figure 5: Samples from DDIM with the same random x_T and different number of steps.

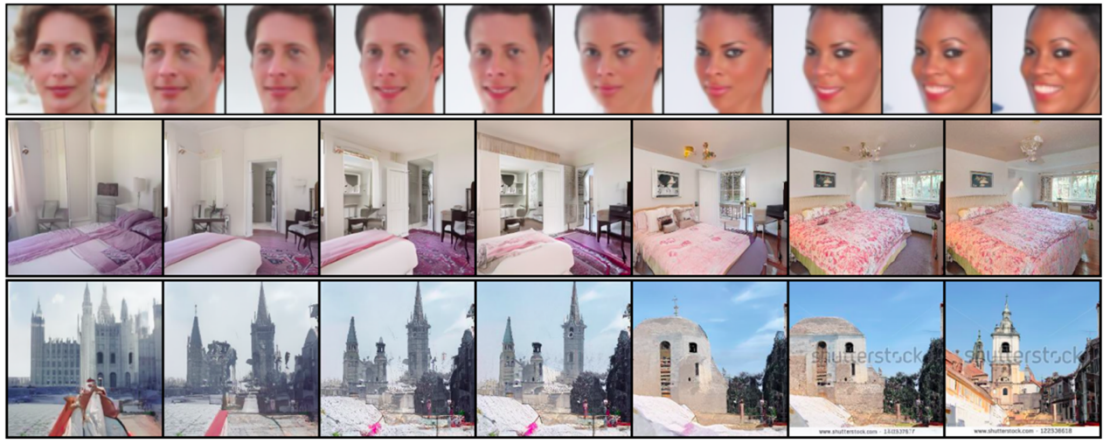


Figure 6: Interpolation of samples from DDIM with $\dim(\tau) = 50$.

What We've Covered Today…

- Generative Models
  - AE & VAE
  - Diffusion Model: DDPM/LDM/DDIM
- Oct. 1st, Tue. (ECCV week)
  - Guest lecture, Dr. Jung-Cheng Chen, Academia Sinica
- What to cover on Oct. 8th
  - Guidance in Diffusion Models
  - Personalization for Diffusion Models
  - Generative Adversarial Network (GAN)