Diffusion model

Presented By
201924124 전자공학과 김희수

Outline

❑ 1 | Introduction
- Generative model : Latent variable model
- Overview of generative model

❑ 2 | Diffusion model
- Diffusion model의 활용
- What is diffusion model?
- Prerequirement

❑ 3 | Diffusion model
- Forward process

❑ 4 | Diffusion model
- Reverse process

❑ 5 | Conclusion

CVSP

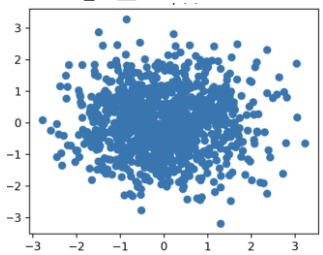
1

Introduction

Introduction

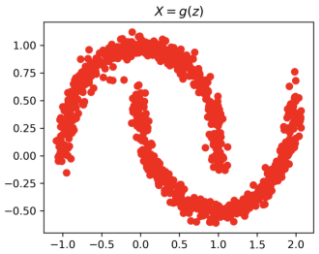
❑ Generative model : Latent variable model

$p_z(z) \sim N$



Trained model

$p'_x(x)$



✓ *Simple distribution*

✓ *Tractable(Gaussian)*

✓ *Complex distribution*

✓ *Visual, Audio patterns*

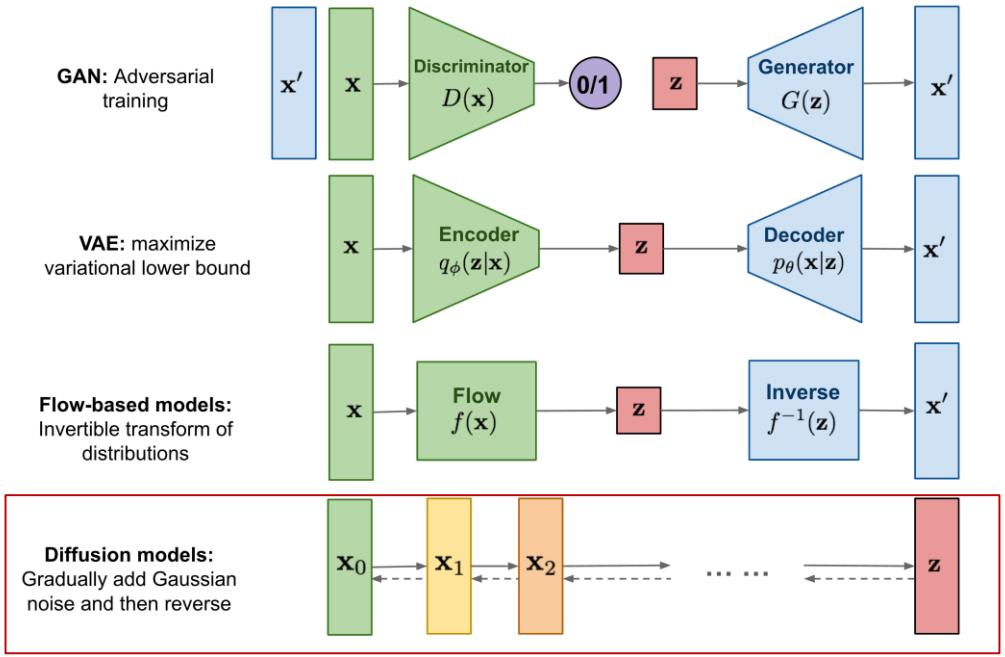
Input

Output

- 결국 생성 모델로부터 **원하는 것은 매우 간단한 분포를 특정한 패턴을 갖는 분포 변환**(mapping, transformation, sampling)하는 것.
- 그렇기에 대부분의 생성모델은 주어진 입력 데이터로부터 **latent variable(Z)**를 얻어내고, 이를 변환하는 역량을 학습하고자 함.
- latent variable model ~: 모든 변수를 알고 참조하는 방식으로 학습하는 것이 아니라 잠재된 변수가 있다고 상정하고 학습을 진행하는 모델

Introduction

❑ Overview of generative model



기존 generative model의 단점
GAN:학습의 불안정성
VAE: 직접 likelihood최대화 불가
Flow-based model: 역함수가 존재하는model(구조)가 필요

Diffusion model

- 반복적인 변화(iterative transformation)를 활용한다는 점에서 **Flow-based models 와 유사**
- 분포에 대한 변분적 추론을 통한 학습을 진행한다는 점은 **VAE와 유사**

2

Diffusion model

Diffusion model

https://ffighting.net/deep-learning-paper-review/diffusion-model/diffusion-model-basic/

❑ Diffusion model의 활용

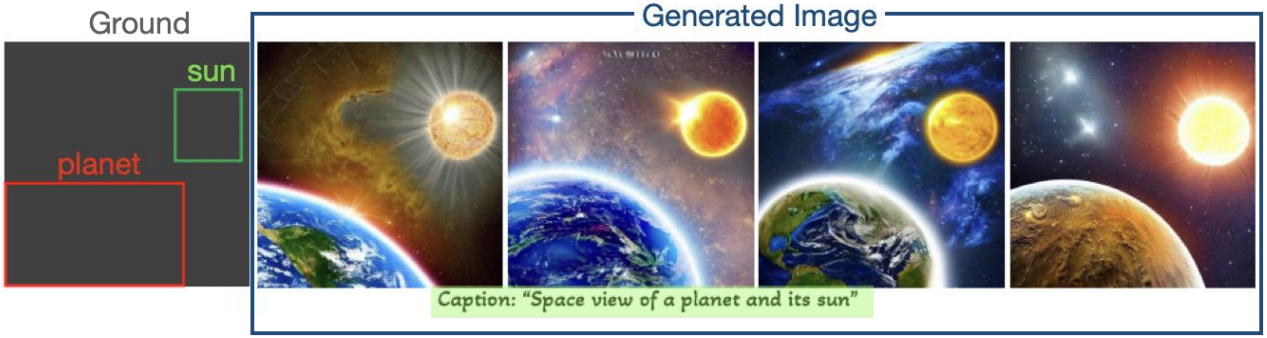
DALLE2가 생성한 이미지



a teddy bear on a skateboard in times square

- Diffusion model은 대표적인 generative model 중 하나로, 기존 Image Generative Model과는 달리 다양한 요구조건에 맞는 고화질 이미지를 생성할 수 있는 모습을 보여주면서 Diffusion Model은 큰 화제가 되었다.

GLIGEN이 생성한 이미지



Diffusion model

❑ Diffusion model의 활용

Text to image generation



Fig. 2. Images generated by GLIDE [15].

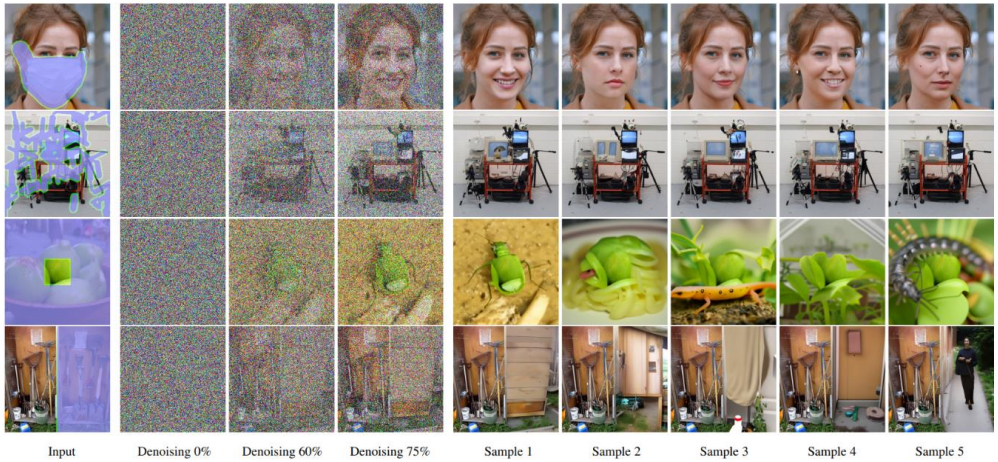
Diffusion Models for Natural Language Generation

Original sentence: *How do I make friends.*		***Paraphrase reference***: *How to make friends ?*
GPT2-large finetune	**GPVAE-T5**	**DIFFUSEQ**
How can I make friends?	How can I make friends?	How can I make friends better?
How can I make friends?	How do I make friends?	How can I make friends?
How can I make friends?	How can I make friends?	How do you make friends?
How can I make friends?	How can I make friends?	What is the best way to make friends?
How do I make friends and keep them?	What's the best way to make friends and make make friends?	How can I make friends and more something?

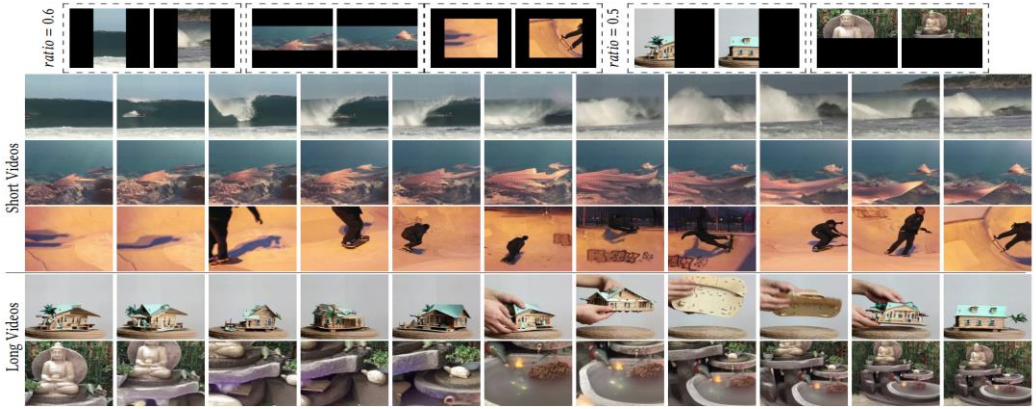
Diffusion model

❑ Diffusion model의 활용

Image inpainting



Video outpainting

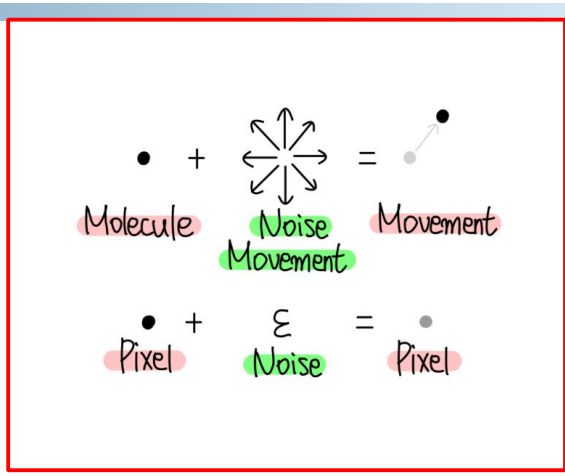


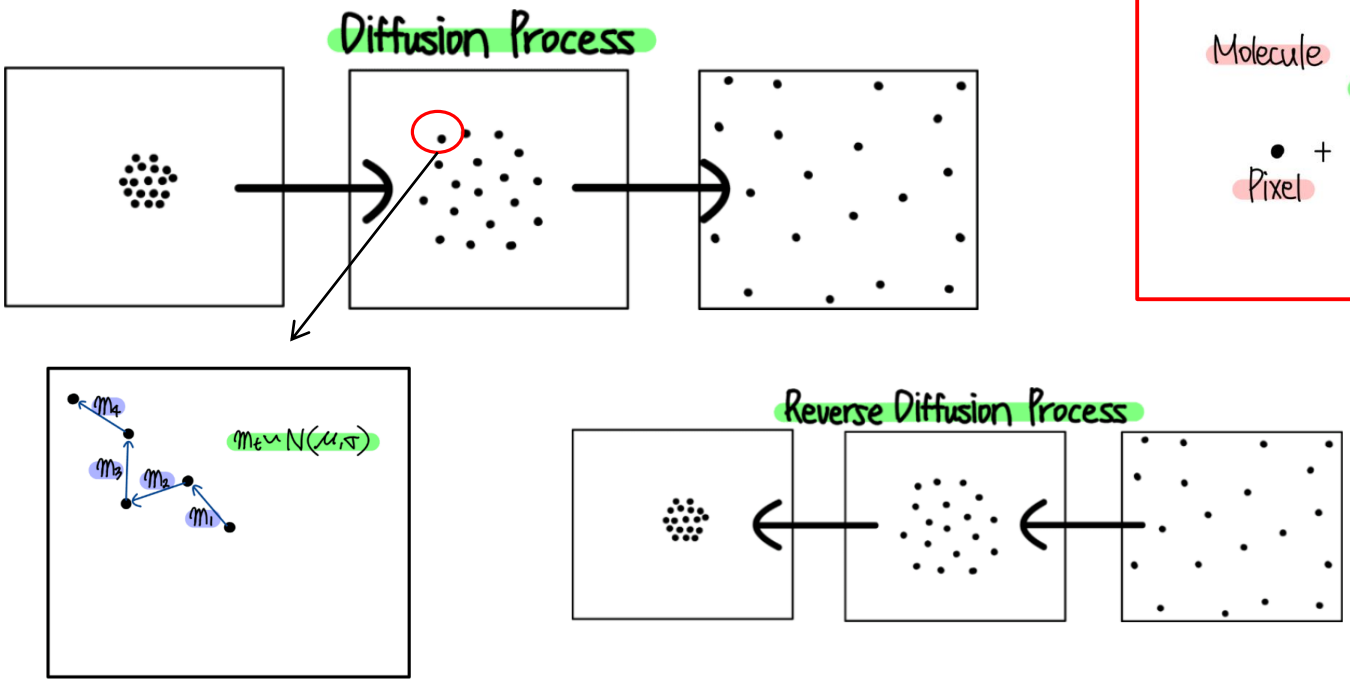


Diffusion model

https://ffighting.net/deep-learning-paper-review/diffusion-model/diffusion-model-basic/

❑ What is diffusion model?





Physical intuition

분자의 움직임은 Gaussian Distribution을 따른다. 즉 위의 그림과 같이 평균(μ)와 표준 편차(σ)를 따르는 정규 분포로 움직임을 표현할 수 있다. 그렇다면, 매시간 t마다 각 분자의 움직임을 계산해낼 수 있다면 어떻게 될까? 그럼 Diffusion Process를 거꾸로 한다면? Reverse diffusion process!

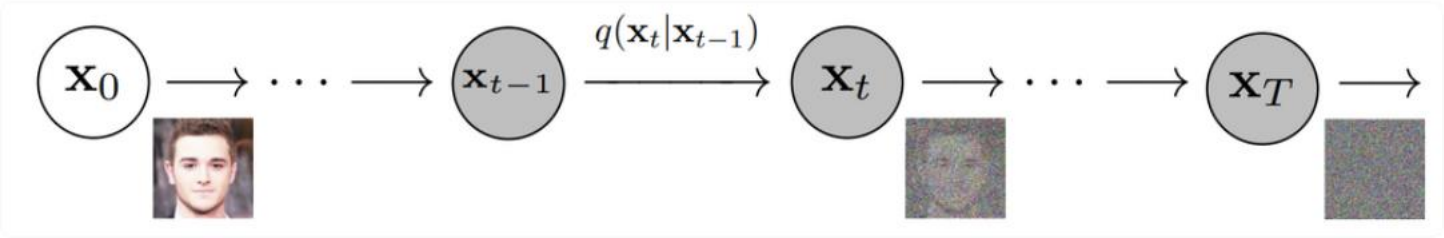
Diffusion model

https://www.assemblyai.com/blog/diffusion-models-for-machine-learning-introduction/

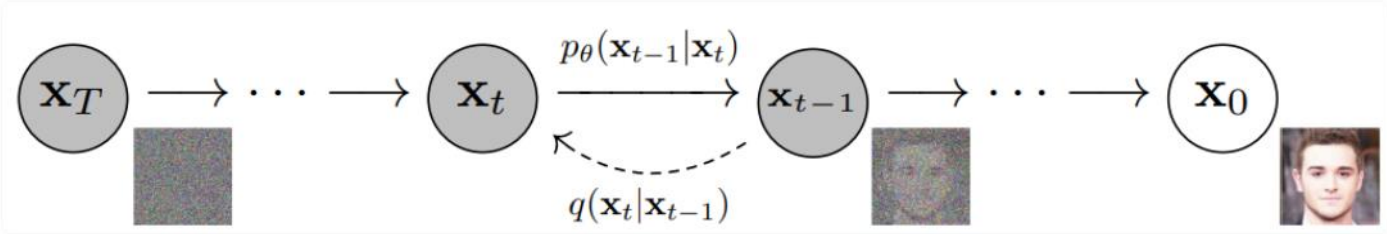
❑ What is diffusion model?

forward process(noising)를 통해 입력 data에 radom하게 매 step noise를 추가해 나가고,

reverse process(denoising)과정을 통해 noise에서 denoising하면서 새로운 data를 generate 하는 모델.



(Modified from source)



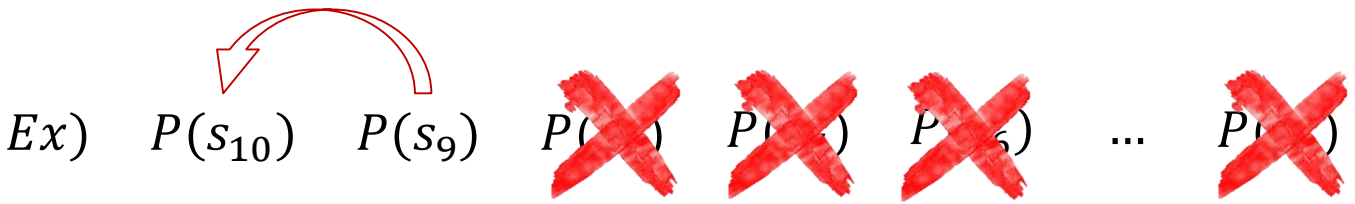
(Modified from source)

Diffusion model

❑ Prerequirement

- Markov chain= Markov 성질을 갖는 이산 확률과정
  - ✓ Markov 성질 : "특정 상태의 확률(t+1)은 오직 현재(t)의 상태에 의존한다"
  - ✓ 이산 확률과정 : 이산적인 시간(0초, 1초, 2초, ..) 속에서의 확률적 현상

$$P[s_{t+1}|s_t = \mathrm{P}[s_{t+1}|s_1, .., s_t]$$



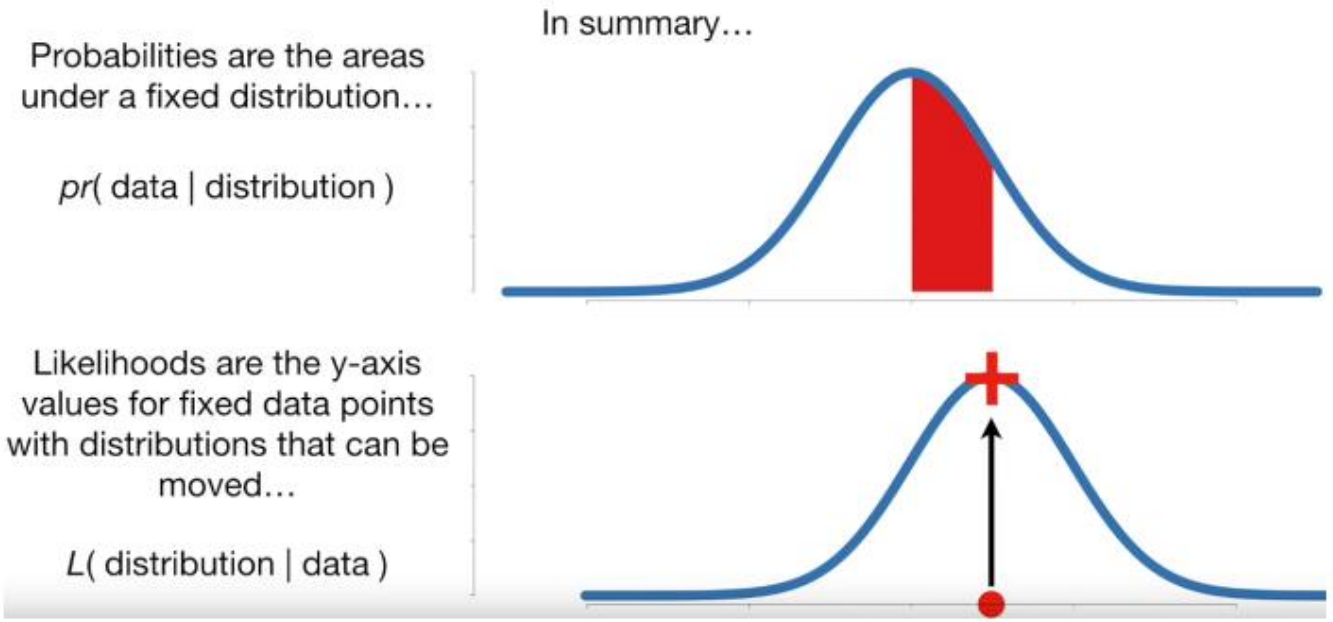
- ✓ 예 : "내일의 날씨는 오늘의 날씨만 보고 알 수 있다."
- ✓ (내일의 날씨는 오로지 오늘의 날씨 만을 조건부로 하는 확률적 과정)

Diffusion model

❑ Prerequirement

- Likelihood : 가능도 즉, 관측된 데이터가 특정 확률분포에서 나왔을 확률

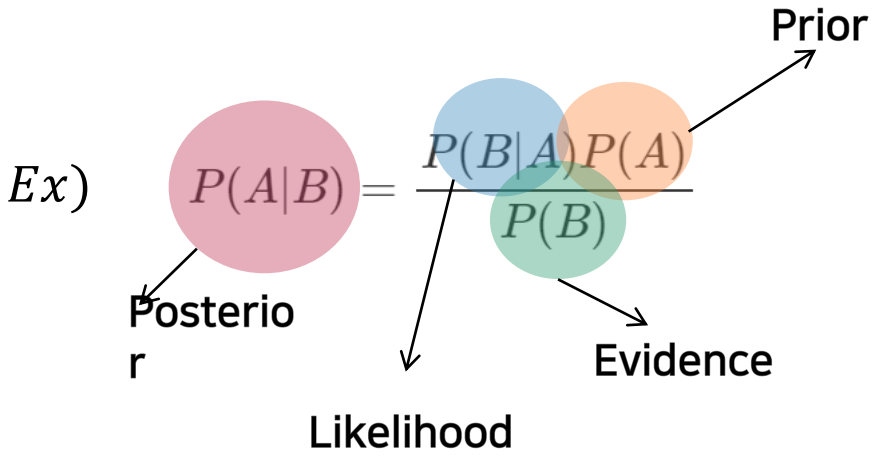
$Ex)$



Diffusion model

❑ Prerequirement

- Bayes rule= 결과를 알고 원인을 예측



수식적 증명

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} = \frac{P(AnB) * P(A)}{P(A) * P(B)}$$

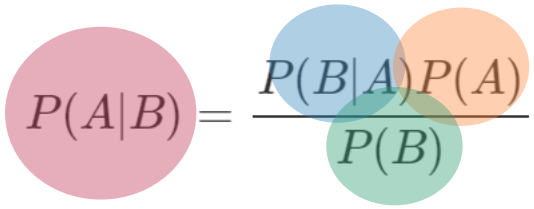
- posterior: 사후확률(사건 B가 발생한 후 갱신된 사건 A의 확률)
- Prior: 사전확률(사건 B가 발생하기 전에 가지고 있던 사건 A의 확률)
- Likelihood: 가능도(사건 A가 발생한 경우 사건 B의 확률)
- Evidence: 증거 (확률의 크기를 조정, 사건 B의 확률)

✓ 사건 B가 발생함으로써(사건 B가 진실이라는 것을 알게 됨으로써, 사건 B의 확률 =1 이라는 것을 알게 됨.) 사건 A의 확률이 어떻게 변화하는지를 표현한 정리 즉, 새로운 정보가 기존의 추론에 어떻게 영향을 미치는지를 나타낸다.

Diffusion model

❑ Prerequirement

- Bayes rule 확장



$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

$$\longrightarrow \quad P(A|B,C) = \frac{P(B|A,C)P(A|C)}{P(B|C)}$$

수식적 증명

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)} = \frac{P(AnB) * P(A)}{P(A) * P(B)}$$

수식적 증명

$P(A|B,C)$는 B와 C가 조건인 A의 확률. 즉 $P(A|(BnC)$

$P(A,B,C) = P(B|A,C)P(A,C)$

$= P(B|A,C)P(A|C)P(C)$

$P(A,B,C) = P(A|B,C)P(B,C)$

$= P(A|B,C)P(B|C)P(C)$

$-> \ P(B|A,C)P(A|C)P(C) = P(A|B,C)P(B|C)P(C)$

$\therefore P(A|B,C) = \frac{P(B|A,C)P(A|C)}{P(B|C)}$

Diffusion model

❑ Prerequirement

- ELBO(Evidence Lower Bound)

$$\mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p(\boldsymbol{x},\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right] \qquad \log p(\boldsymbol{x}) \geq \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p(\boldsymbol{x},\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right]$$

<유도과정1>

$$\begin{aligned}
\log p(\boldsymbol{x}) &= \log \int p(\boldsymbol{x},\boldsymbol{z})d\boldsymbol{z} \\
&= \log \int \frac{p(\boldsymbol{x},\boldsymbol{z})q_\phi(\boldsymbol{z}|\boldsymbol{x})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}d\boldsymbol{z} \\
&= \log \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\frac{p(\boldsymbol{x},\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right] \\
&\geq \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p(\boldsymbol{x},\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right]
\end{aligned}$$

marginal property

*1 Expectation 정의 apply

jensen's inequality

<유도과정2>

$$\begin{aligned}
\log p(\boldsymbol{x}) &= \log p(\boldsymbol{x}) \int q_\phi(\boldsymbol{z}|\boldsymbol{x})d\boldsymbol{z} && \text{(Multiply by } 1 = \int q_\phi(\boldsymbol{z}|\boldsymbol{x})d\boldsymbol{z}) && (9)\\
&= \int q_\phi(\boldsymbol{z}|\boldsymbol{x})(\log p(\boldsymbol{x}))d\boldsymbol{z} && \text{(Bring evidence into integral)} && (10)\\
&= \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}[\log p(\boldsymbol{x})] && \text{(Definition of Expectation)} && (11)\\
&= \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p(\boldsymbol{x},\boldsymbol{z})}{p(\boldsymbol{z}|\boldsymbol{x})}\right] && \text{(Apply Equation 2)} && (12)\\
&= \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p(\boldsymbol{x},\boldsymbol{z})q_\phi(\boldsymbol{z}|\boldsymbol{x})}{p(\boldsymbol{z}|\boldsymbol{x})q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right] && \text{(Multiply by } 1 = \frac{q_\phi(\boldsymbol{z}|\boldsymbol{x})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}) && (13)\\
&= \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p(\boldsymbol{x},\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right] + \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{q_\phi(\boldsymbol{z}|\boldsymbol{x})}{p(\boldsymbol{z}|\boldsymbol{x})}\right] && \text{(Split the Expectation)} && (14)\\
&= \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p(\boldsymbol{x},\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right] + D_{\mathrm{KL}}(q_\phi(\boldsymbol{z}|\boldsymbol{x}) \parallel p(\boldsymbol{z}|\boldsymbol{x})) && \text{(Definition of KL Divergence)} && (15)\\
&\geq \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{p(\boldsymbol{x},\boldsymbol{z})}{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\right] && \text{(KL Divergence always} \geq 0) && (16)
\end{aligned}$$

KLdivergence

$$\begin{aligned}
KL(q(z) \parallel p(z \mid x;\theta)) &:= E_{Z\sim q}\left[\log \frac{q(Z)}{p(Z \mid x;\theta)}\right] \\
&= E_{Z\sim q}[\log q(Z)] - E_{Z\sim q}\left[\log \frac{p(x,Z;\theta)}{p(x;\theta)}\right] \\
&= E_{Z\sim q}[\log q(Z)] - E_{Z\sim q}[\log p(x,Z;\theta)] + E_{Z\sim q}[\log p(x;\theta)] \\
&= \log p(x;\theta) - E_{Z\sim q}\left[\log \frac{p(x,Z;\theta)}{q(z)}\right] \\
&= \text{evidence} - \text{ELBO}
\end{aligned}$$

$\text{evidence} := \log p(x;\theta)$

고정된 세타에서의 log likelihood function = evidence

(세타=parameters, p=model, x=observed data)

(주어진 parameter값으로부터 데이터가 나타날 확률)

$p(x,z)$

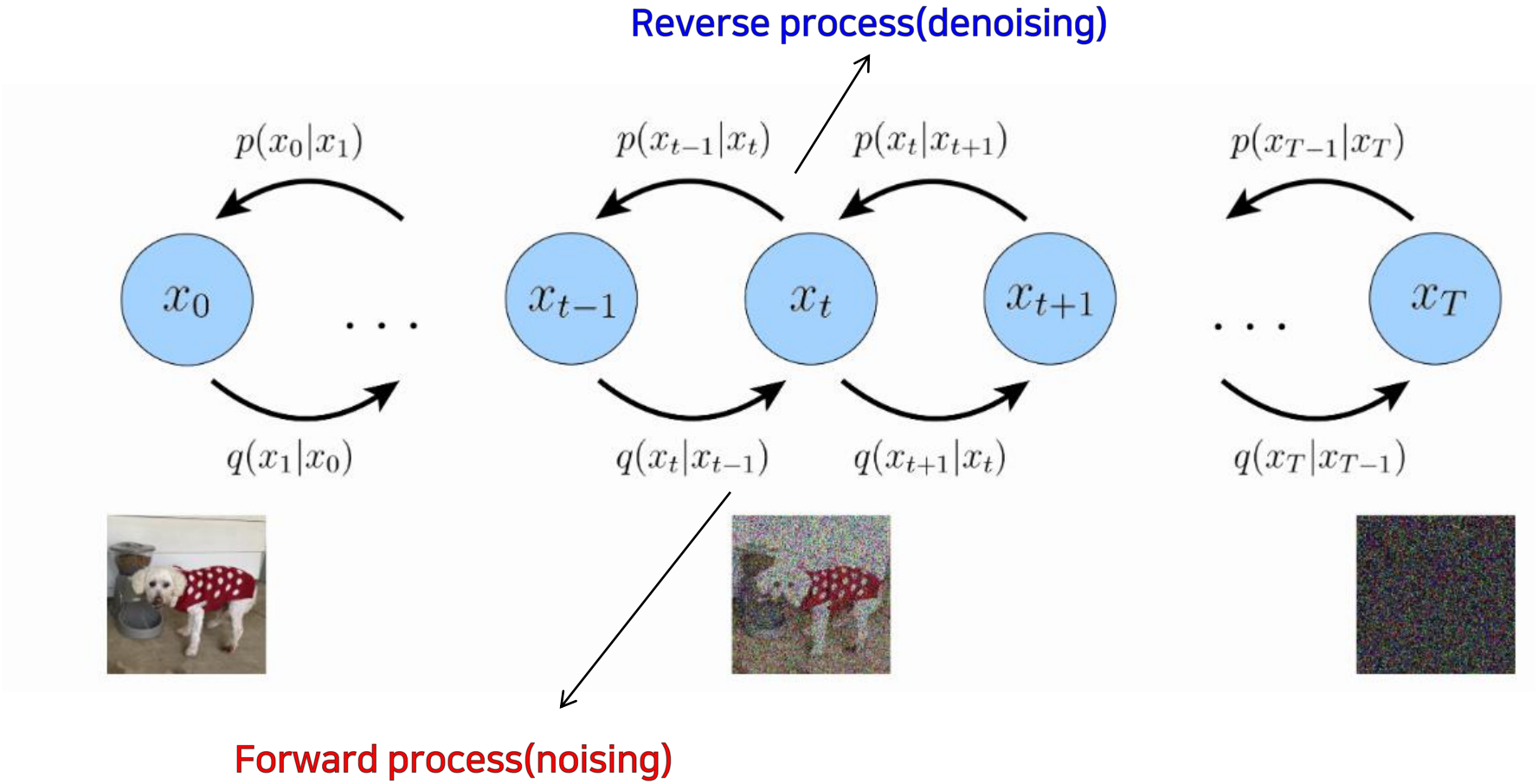
(주어진 데이터 x 및 잠재 변수z에 대한 결합 확률 분포.관측데이터와잠재변수의 결합 확률)

$q_\emptyset(z|x)$

(주어진 데이터 x에 대한 잠재 변수z에 대한 z의 변분 추론 분포. 관측데이터로부터 잠재변수의 가능한 값에 대한 확률 분포)

Diffusion model

- Diffusion model flow



3

Diffusion model

(Forward process)

Diffusion model

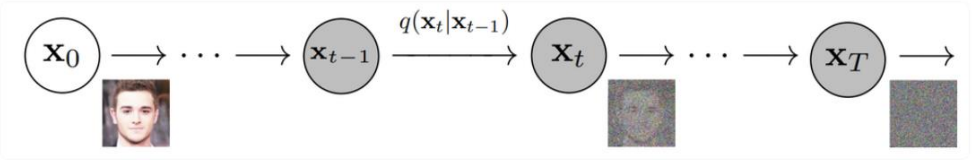
❑ Forward process

$$q(x_t|x_{t-1} = N(x_t; \sqrt{1-\beta_t} * x_{t-1}, \beta_t * I)$$

매 step의 data distribution을 의미, 즉, noise를 추가하는 단계. 임의의 랜덤한 variance schedule 값을 사용하여 $\{\beta_t \in (0,1)\}_{t=1}^{T}$ 입력은 점점 희미하게 즉, 전 단은 step의 반복에 따라 distinguishable한 feature를 잃어가고, 노이즈는 점점 계속해서 증가시켜 나간다. 이후 최종으로 t=T인 순간에는 gaussian noise를 가지게 된다. (이때 $\beta_1 < \beta_2 < \beta_t$ ....)

즉, 매 step의 distribution은 markov 성질을 따르고, 전체 t=1에서 t=T까지의 과정은 $q(x_{1:T}|x_0 = \Pi_{t=1}^{T} q(x_t|x_{t-1})$ given x0 즉, input data가 주어졌을때, t=1에서 t=T까지의 노이즈를 더해가는 과정은 markov chain 형태로 나타난다.

즉, forward process는 자연스러운 확산 과정으로, 잉크를 떨어트렸을때 자연스럽게 퍼져나가서 최종적으로 uniform하게 퍼져있는 그러한 형태로 자연스러운 과정으로 생각할 수 있다.



(Modified from source)

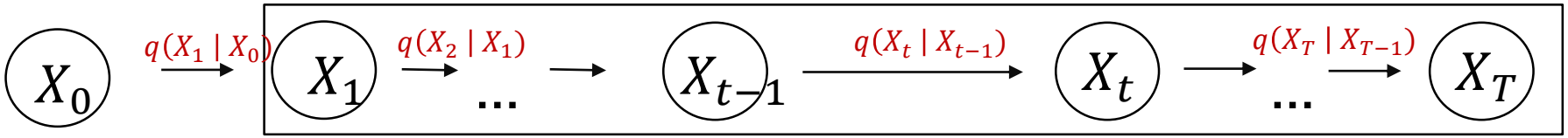
x1~xT: noise sample

x0 : input data

Diffusion model

lilianweng.github.io,
https://www.youtube.com/watch?v=_JQSMhqXw-4

- Forward process



$Origianl\ Data$ → $X_1 = Z_1, X_2 = Z_2, \dots, X_T = Z_T$

X_0



pure isotrophic gaussian noise
(모든 방향에서 동일한 특성을 가지는 gaussian 분포를 따르는 noise)

$$q(x_t|x_{t-1}) = N(x_t; \sqrt{1-\beta_t} * x_{t-1}, \beta_t * I)$$

β_t

: $Noise$ 주입 정도

CVSP

4

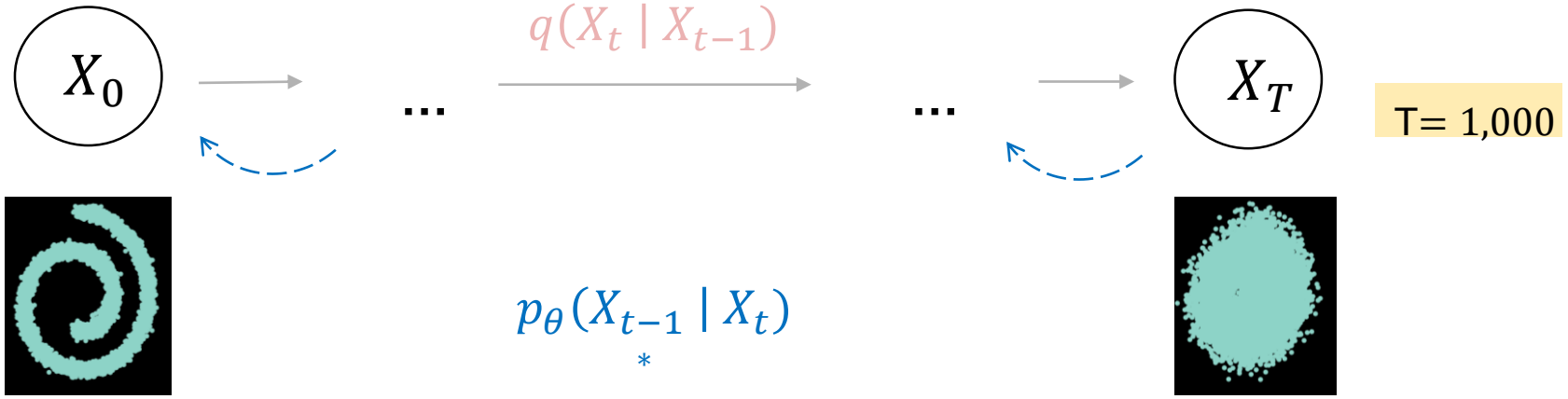
Diffusion model

(Reverse process)

Diffusion model

- Reverse process

Reverse process 는 diffusion process의 역 과정(denoising)을 학습



Diffusion model

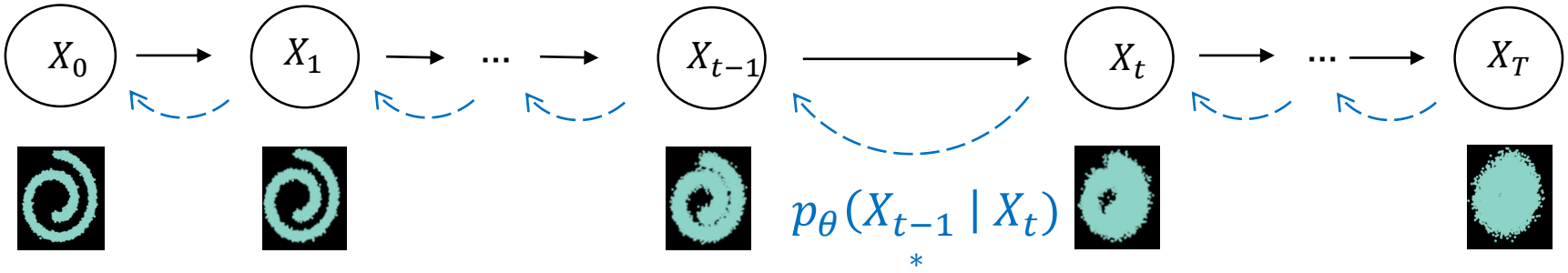
❑ Reverse process

Reverse process 는 diffusion process의 역과정으로, gaussian nosie를 제거하면서 특정한 패턴을 만들어가는 과정

$$p_\theta(X_{0:T}) := p(X_T) \prod_{t=1}^{T} p_\theta(X_{t-1} \mid X_t) \quad / \quad p_\theta(X_{t-1} \mid X_t) := N(X_{t-1}; \mu_\theta(X_t, t), \Sigma_\theta(X_t, t))$$

학습 대상
$(mean\ \&\ variance\ function)$

$$p(X_T) = N(X_T;\ \mathbf{0}, \boldsymbol{I})$$



Diffusion model

Understanding Diffusion Models: A Unifted Perspective, Calvin Luo(google Research, brain team), August 26, 2022

❑ Reverse process

Find ELBO

$$
\begin{aligned}
\log p(\boldsymbol{x}) &\geq \mathbb{E}_{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p(\boldsymbol{x}_{0:T})}{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\right] \\
&= \mathbb{E}_{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p(\boldsymbol{x}_T)\prod_{t=1}^{T} p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)}{\prod_{t=1}^{T} q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})}\right] \\
&= \mathbb{E}_{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p(\boldsymbol{x}_T)p_{\boldsymbol{\theta}}(\boldsymbol{x}_0|\boldsymbol{x}_1)\prod_{t=2}^{T} p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)}{q(\boldsymbol{x}_1|\boldsymbol{x}_0)\prod_{t=2}^{T} q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})}\right] \\
&= \mathbb{E}_{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p(\boldsymbol{x}_T)p_{\boldsymbol{\theta}}(\boldsymbol{x}_0|\boldsymbol{x}_1)\prod_{t=2}^{T} p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)}{q(\boldsymbol{x}_1|\boldsymbol{x}_0)\prod_{t=2}^{T} q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\boldsymbol{x}_0)}\right] \\
&= \mathbb{E}_{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_T)p_{\boldsymbol{\theta}}(\boldsymbol{x}_0|\boldsymbol{x}_1)}{q(\boldsymbol{x}_1|\boldsymbol{x}_0)} + \log\prod_{t=2}^{T}\frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)}{q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\boldsymbol{x}_0)}\right] \\
&= \mathbb{E}_{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p(\boldsymbol{x}_T)p_{\boldsymbol{\theta}}(\boldsymbol{x}_0|\boldsymbol{x}_1)}{q(\boldsymbol{x}_1|\boldsymbol{x}_0)} + \log\prod_{t=2}^{T}\frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)}{\frac{q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)q(\boldsymbol{x}_t|\boldsymbol{x}_0)}{q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_0)}}\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \mathbb{E}_{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p(\boldsymbol{x}_T)p_{\boldsymbol{\theta}}(\boldsymbol{x}_0|\boldsymbol{x}_1)}{q(\boldsymbol{x}_1|\boldsymbol{x}_0)} + \log\prod_{t=2}^{T}\frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)}{\frac{q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)\cancel{q(\boldsymbol{x}_t|\boldsymbol{x}_0)}}{\cancel{q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_0)}}}\right] \\
&= \mathbb{E}_{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p(\boldsymbol{x}_T)p_{\boldsymbol{\theta}}(\boldsymbol{x}_0|\boldsymbol{x}_1)}{\cancel{q(\boldsymbol{x}_1|\boldsymbol{x}_0)}} + \log\frac{\cancel{q(\boldsymbol{x}_1|\boldsymbol{x}_0)}}{q(\boldsymbol{x}_T|\boldsymbol{x}_0)} + \log\prod_{t=2}^{T}\frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)}{q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)}\right] \\
&= \mathbb{E}_{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p(\boldsymbol{x}_T)p_{\boldsymbol{\theta}}(\boldsymbol{x}_0|\boldsymbol{x}_1)}{q(\boldsymbol{x}_T|\boldsymbol{x}_0)} + \sum_{t=2}^{T}\log\frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)}{q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)}\right] \\
&= \mathbb{E}_{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}_0|\boldsymbol{x}_1)\right] + \mathbb{E}_{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p(\boldsymbol{x}_T)}{q(\boldsymbol{x}_T|\boldsymbol{x}_0)}\right] + \sum_{t=2}^{T}\mathbb{E}_{q(\boldsymbol{x}_{1:T}|\boldsymbol{x}_0)}\left[\log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)}{q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)}\right] \\
&= \mathbb{E}_{q(\boldsymbol{x}_1|\boldsymbol{x}_0)}\left[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}_0|\boldsymbol{x}_1)\right] + \mathbb{E}_{q(\boldsymbol{x}_T|\boldsymbol{x}_0)}\left[\log \frac{p(\boldsymbol{x}_T)}{q(\boldsymbol{x}_T|\boldsymbol{x}_0)}\right] + \sum_{t=2}^{T}\mathbb{E}_{q(\boldsymbol{x}_t,\boldsymbol{x}_{t-1}|\boldsymbol{x}_0)}\left[\log \frac{p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)}{q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)}\right] \\
&= \underbrace{\mathbb{E}_{q(\boldsymbol{x}_1|\boldsymbol{x}_0)}\left[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}_0|\boldsymbol{x}_1)\right]}_{\text{reconstruction term}} - \underbrace{D_{\mathrm{KL}}(q(\boldsymbol{x}_T|\boldsymbol{x}_0)\parallel p(\boldsymbol{x}_T))}_{\text{prior matching term}} - \sum_{t=2}^{T}\underbrace{\mathbb{E}_{q(\boldsymbol{x}_t|\boldsymbol{x}_0)}\left[D_{\mathrm{KL}}(q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)\parallel p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t))\right]}_{\text{denoising matching term}}
\end{aligned}
$$

노이즈가 추가된 입력의 분포가 표준 가우시안 정규 분포에 얼마나 가까운지를 알 수 있는 term
P분포가 generate하는 XT와 ground truth를 알고 xT를 generate하는 q분포와의 차이
(논문에서는 0으로 가정0

x0(노이즈가 없는 이미지)를 찾는 과정 = denoising 과정에서 ground truth denoising 과정과 ground truth 를 모를 때 denoising 과정 간의 차이를 보는 term(즉, 0에 수렴할 수록 ground truth에 가까움)
Forward proces와 Reverse process 분포간의 차이 (최대한 같게 학습이 목표)

최종 생성하는 목표였던 생성 이미지를 만드는 term(마지막 noise(latent xl)로부터 새로운 이미지(data x0)를 만드는 likelihood)-> maximize 필요

$$p_\theta(X_{0:T}) := p(X_T)\prod_{t=1}^{T} p_\theta(X_{t-1}\mid X_t) \quad / \quad p_\theta(X_{t-1}\mid X_t) := N(X_{t-1}; \mu_\theta(X_t,t), \Sigma_\theta(X_t,t))$$

$$q(x_{1:T}|x_0) = \Pi_{t=1}^{T} q(x_t|x_{t-1})$$

CVSP

Diffusion model

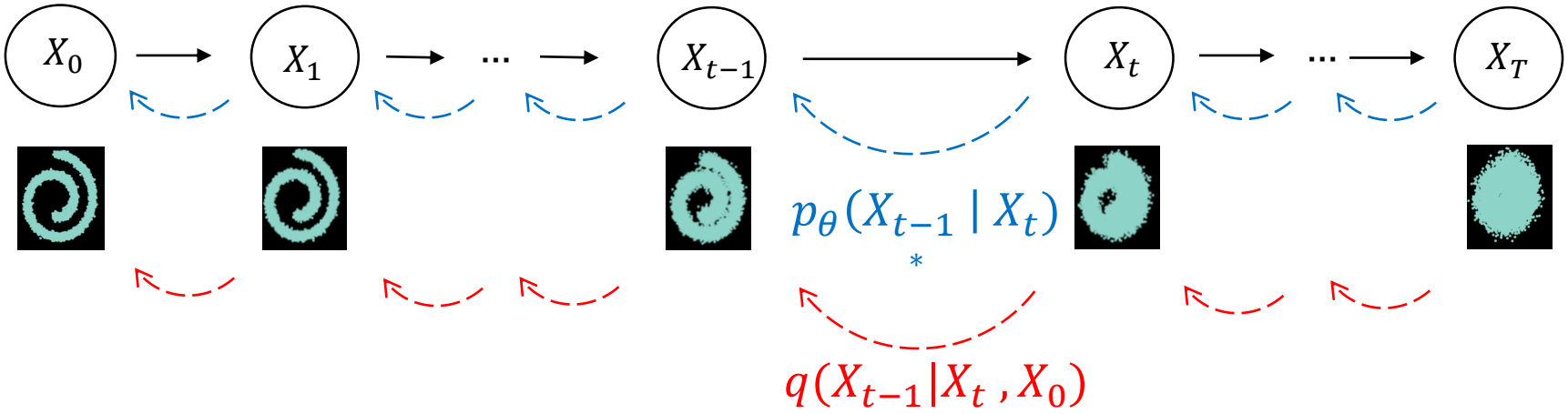
Denoising Diffusion Probabilistic Model(Jonathan Ho, 2020)
https://www.youtube.com/watch?v=_JQSMhqXw-4
Understanding Diffusion Models: A Unifted Perspective,
Calvin Luo(google Research, brain team), August 26, 2022

❑ Reverse process

Ground truth = X_0 를 알기 때문에 이를 이용해서 reverse process를 해보자!

$$p_\theta(X_{0:T}) \coloneqq p(X_T) \prod_{t=1}^{T} q(X_{t-1} \mid X_t) \quad \approx \quad p_\theta(X_{0:T}) \coloneqq p(X_T) \prod_{t=1}^{T} p_\theta(X_{t-1} \mid X_t)$$

$$p(X_T) = N(X_T;\ \mathbf{0}, \boldsymbol{I})$$



CVSP

Diffusion model

❑ Reverse process

Ground truth = X_0 를 알기 때문에 이를 이용해서 reverse process를 해보자!

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

(by bayes rule)

$$\begin{aligned} q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) &= \frac{q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{x}_0)q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_0)}{q(\boldsymbol{x}_t|\boldsymbol{x}_0)} \\ &= \frac{\mathcal{N}(\boldsymbol{x}_t; \sqrt{\alpha_t}\boldsymbol{x}_{t-1}, (1-\alpha_t)\mathbf{I})\mathcal{N}(\boldsymbol{x}_{t-1}; \sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_0, (1-\bar{\alpha}_{t-1})\mathbf{I})}{\mathcal{N}(\boldsymbol{x}_t; \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0, (1-\bar{\alpha}_t)\mathbf{I})} \end{aligned}$$

(by forward process)

$$q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{x}_0) = q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \mathcal{N}(\boldsymbol{x}_t; \sqrt{\alpha_t}\boldsymbol{x}_{t-1}, (1-\alpha_t)\mathbf{I})$$

(by reparameterization trick)

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

$$\boldsymbol{x}_t \sim q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1})$$
$$\beta_t = 1 - \alpha_t$$
$$\{\boldsymbol{\epsilon}_t^*, \boldsymbol{\epsilon}_t\}_{t=0}^T \overset{\text{iid}}{\sim} \mathcal{N}(\boldsymbol{\epsilon}; \mathbf{0}, \mathbf{I})$$

(random noise variables)

$$\begin{aligned} \boldsymbol{x}_t &= \sqrt{\alpha_t}\boldsymbol{x}_{t-1} + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_{t-1}^* \\ &= \sqrt{\alpha_t}\left(\sqrt{\alpha_{t-1}}\boldsymbol{x}_{t-2} + \sqrt{1-\alpha_{t-1}}\boldsymbol{\epsilon}_{t-2}^*\right) + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_{t-1}^* \\ &= \sqrt{\alpha_t\alpha_{t-1}}\boldsymbol{x}_{t-2} + \sqrt{\alpha_t - \alpha_t\alpha_{t-1}}\boldsymbol{\epsilon}_{t-2}^* + \sqrt{1-\alpha_t}\boldsymbol{\epsilon}_{t-1}^* \\ &= \sqrt{\alpha_t\alpha_{t-1}}\boldsymbol{x}_{t-2} + \sqrt{\sqrt{\alpha_t - \alpha_t\alpha_{t-1}}^2 + \sqrt{1-\alpha_t}^2}\boldsymbol{\epsilon}_{t-2} \\ &= \sqrt{\alpha_t\alpha_{t-1}}\boldsymbol{x}_{t-2} + \sqrt{\alpha_t - \alpha_t\alpha_{t-1} + 1 - \alpha_t}\boldsymbol{\epsilon}_{t-2} \\ &= \sqrt{\alpha_t\alpha_{t-1}}\boldsymbol{x}_{t-2} + \sqrt{1-\alpha_t\alpha_{t-1}}\boldsymbol{\epsilon}_{t-2} \\ &= \dots \\ &= \sqrt{\prod_{i=1}^{t}\alpha_i}\boldsymbol{x}_0 + \sqrt{1-\prod_{i=1}^{t}\alpha_i}\boldsymbol{\epsilon}_0 \\ &= \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_0 \\ &\sim \mathcal{N}(\boldsymbol{x}_t; \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0, (1-\bar{\alpha}_t)\,\mathbf{I}) \end{aligned}$$

CVSP

Diffusion model

Understanding Diffusion Models: A Unifted Perspective,
Calvin Luo(google Research, brain team), August 26, 2022

- Reverse process

$$q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0) = \frac{q(\boldsymbol{x}_t|\boldsymbol{x}_{t-1},\boldsymbol{x}_0)q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_0)}{q(\boldsymbol{x}_t|\boldsymbol{x}_0)} \tag{71}$$

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

Gaussian pdf

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$\propto \exp\left\{-\left[\frac{(\boldsymbol{x}_t - \sqrt{\alpha_t}\boldsymbol{x}_{t-1})^2}{2(1-\alpha_t)} + \frac{(\boldsymbol{x}_{t-1} - \sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_0)^2}{2(1-\bar{\alpha}_{t-1})} - \frac{(\boldsymbol{x}_t - \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0)^2}{2(1-\bar{\alpha}_t)}\right]\right\} \tag{73}$$

$$= \exp\left\{-\frac{1}{2}\left[\frac{(\boldsymbol{x}_t - \sqrt{\alpha_t}\boldsymbol{x}_{t-1})^2}{1-\alpha_t} + \frac{(\boldsymbol{x}_{t-1} - \sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_0)^2}{1-\bar{\alpha}_{t-1}} - \frac{(\boldsymbol{x}_t - \sqrt{\bar{\alpha}_t}\boldsymbol{x}_0)^2}{1-\bar{\alpha}_t}\right]\right\} \tag{74}$$

1/2 밖으로 빼기

$$= \exp\left\{-\frac{1}{2}\left[\frac{(-2\sqrt{\alpha_t}\boldsymbol{x}_t\boldsymbol{x}_{t-1} + \alpha_t\boldsymbol{x}_{t-1}^2)}{1-\alpha_t} + \frac{(\boldsymbol{x}_{t-1}^2 - 2\sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_{t-1}\boldsymbol{x}_0)}{1-\bar{\alpha}_{t-1}} + C(\boldsymbol{x}_t,\boldsymbol{x}_0)\right]\right\} \tag{75}$$

X_t-1과 무관한 term = constant term

$$\propto \exp\left\{-\frac{1}{2}\left[-\frac{2\sqrt{\alpha_t}\boldsymbol{x}_t\boldsymbol{x}_{t-1}}{1-\alpha_t} + \frac{\alpha_t\boldsymbol{x}_{t-1}^2}{1-\alpha_t} + \frac{\boldsymbol{x}_{t-1}^2}{1-\bar{\alpha}_{t-1}} - \frac{2\sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_{t-1}\boldsymbol{x}_0}{1-\bar{\alpha}_{t-1}}\right]\right\} \tag{76}$$

$$= \exp\left\{-\frac{1}{2}\left[\left(\frac{\alpha_t}{1-\alpha_t} + \frac{1}{1-\bar{\alpha}_{t-1}}\right)\boldsymbol{x}_{t-1}^2 - 2\left(\frac{\sqrt{\alpha_t}\boldsymbol{x}_t}{1-\alpha_t} + \frac{\sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_0}{1-\bar{\alpha}_{t-1}}\right)\boldsymbol{x}_{t-1}\right]\right\} \tag{77}$$

공통항 끼리 묶기

$$= \exp\left\{-\frac{1}{2}\left[\frac{\alpha_t(1-\bar{\alpha}_{t-1}) + 1 - \alpha_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}\boldsymbol{x}_{t-1}^2 - 2\left(\frac{\sqrt{\alpha_t}\boldsymbol{x}_t}{1-\alpha_t} + \frac{\sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_0}{1-\bar{\alpha}_{t-1}}\right)\boldsymbol{x}_{t-1}\right]\right\} \tag{78}$$

통분,분배,정리

$$= \exp\left\{-\frac{1}{2}\left[\frac{\alpha_t - \bar{\alpha}_t + 1 - \alpha_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}\boldsymbol{x}_{t-1}^2 - 2\left(\frac{\sqrt{\alpha_t}\boldsymbol{x}_t}{1-\alpha_t} + \frac{\sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_0}{1-\bar{\alpha}_{t-1}}\right)\boldsymbol{x}_{t-1}\right]\right\} \tag{79}$$

$$= \exp\left\{-\frac{1}{2}\left[\frac{1-\bar{\alpha}_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}\boldsymbol{x}_{t-1}^2 - 2\left(\frac{\sqrt{\alpha_t}\boldsymbol{x}_t}{1-\alpha_t} + \frac{\sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_0}{1-\bar{\alpha}_{t-1}}\right)\boldsymbol{x}_{t-1}\right]\right\} \tag{80}$$

$$= \exp\left\{-\frac{1}{2}\left(\frac{1-\bar{\alpha}_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}\right)\left[\boldsymbol{x}_{t-1}^2 - 2\frac{\left(\frac{\sqrt{\alpha_t}\boldsymbol{x}_t}{1-\alpha_t} + \frac{\sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_0}{1-\bar{\alpha}_{t-1}}\right)}{\frac{1-\bar{\alpha}_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}}\boldsymbol{x}_{t-1}\right]\right\} \tag{81}$$

공통항 묶기

$$= \exp\left\{-\frac{1}{2}\left(\frac{1-\bar{\alpha}_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}\right)\left[\boldsymbol{x}_{t-1}^2 - 2\frac{\left(\frac{\sqrt{\alpha_t}\boldsymbol{x}_t}{1-\alpha_t} + \frac{\sqrt{\bar{\alpha}_{t-1}}\boldsymbol{x}_0}{1-\bar{\alpha}_{t-1}}\right)(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\boldsymbol{x}_{t-1}\right]\right\} \tag{82}$$

범분수

$$= \exp\left\{-\frac{1}{2}\left(\frac{1}{\frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}}\right)\left[\boldsymbol{x}_{t-1}^2 - 2\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\boldsymbol{x}_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\boldsymbol{x}_0}{1-\bar{\alpha}_t}\boldsymbol{x}_{t-1}\right]\right\} \tag{83}$$

분배 및 정리

$$\propto \mathcal{N}(\boldsymbol{x}_{t-1}; \underbrace{\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\boldsymbol{x}_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\boldsymbol{x}_0}{1-\bar{\alpha}_t}}_{\mu_q(\boldsymbol{x}_t,\boldsymbol{x}_0)}, \underbrace{\frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}\mathbf{I}}_{\boldsymbol{\Sigma}_q(t)}) \tag{84}$$

Diffusion model

❑ Reverse process

To meet KL divergence smaller (two gaussian distribuition이 같아지게끔.) => optimization

$$p_\theta(X_{t-1} \mid X_t) \approx q(X_{t-1} \mid X_t, X_0)$$

KL Divergence between two Gaussian distributions

$$D_{\mathrm{KL}}(\mathcal{N}(\boldsymbol{x};\boldsymbol{\mu}_x,\boldsymbol{\Sigma}_x) \parallel \mathcal{N}(\boldsymbol{y};\boldsymbol{\mu}_y,\boldsymbol{\Sigma}_y)) = \frac{1}{2}\left[\log\frac{|\boldsymbol{\Sigma}_y|}{|\boldsymbol{\Sigma}_x|} - d + \mathrm{tr}(\boldsymbol{\Sigma}_y^{-1}\boldsymbol{\Sigma}_x) + (\boldsymbol{\mu}_y-\boldsymbol{\mu}_x)^T\boldsymbol{\Sigma}_y^{-1}(\boldsymbol{\mu}_y-\boldsymbol{\mu}_x)\right]$$

$$\begin{aligned}
&\arg\min_{\boldsymbol{\theta}} D_{\mathrm{KL}}(q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0) \parallel p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)) \\
&= \arg\min_{\boldsymbol{\theta}} D_{\mathrm{KL}}(\mathcal{N}(\boldsymbol{x}_{t-1};\boldsymbol{\mu}_q,\boldsymbol{\Sigma}_q(t)) \parallel \mathcal{N}(\boldsymbol{x}_{t-1};\boldsymbol{\mu}_{\boldsymbol{\theta}},\boldsymbol{\Sigma}_q(t))) \\
&= \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\left[\log\frac{|\boldsymbol{\Sigma}_q(t)|}{|\boldsymbol{\Sigma}_q(t)|} - d + \mathrm{tr}(\boldsymbol{\Sigma}_q(t)^{-1}\boldsymbol{\Sigma}_q(t)) + (\boldsymbol{\mu}_{\boldsymbol{\theta}}-\boldsymbol{\mu}_q)^T\boldsymbol{\Sigma}_q(t)^{-1}(\boldsymbol{\mu}_{\boldsymbol{\theta}}-\boldsymbol{\mu}_q)\right] \\
&= \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\left[\log 1 - d + d + (\boldsymbol{\mu}_{\boldsymbol{\theta}}-\boldsymbol{\mu}_q)^T\boldsymbol{\Sigma}_q(t)^{-1}(\boldsymbol{\mu}_{\boldsymbol{\theta}}-\boldsymbol{\mu}_q)\right] \\
&= \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\left[(\boldsymbol{\mu}_{\boldsymbol{\theta}}-\boldsymbol{\mu}_q)^T\boldsymbol{\Sigma}_q(t)^{-1}(\boldsymbol{\mu}_{\boldsymbol{\theta}}-\boldsymbol{\mu}_q)\right] \\
&= \arg\min_{\boldsymbol{\theta}} \frac{1}{2}\left[(\boldsymbol{\mu}_{\boldsymbol{\theta}}-\boldsymbol{\mu}_q)^T\left(\sigma_q^2(t)\mathbf{I}\right)^{-1}(\boldsymbol{\mu}_{\boldsymbol{\theta}}-\boldsymbol{\mu}_q)\right] \\
&= \arg\min_{\boldsymbol{\theta}} \frac{1}{2\sigma_q^2(t)}\left[\|\boldsymbol{\mu}_{\boldsymbol{\theta}}-\boldsymbol{\mu}_q\|_2^2\right]
\end{aligned}$$

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

$$\boldsymbol{\mu}_q(\boldsymbol{x}_t,\boldsymbol{x}_0) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\boldsymbol{x}_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\boldsymbol{x}_0}{1-\bar{\alpha}_t}$$

Diffusion model

❑ Optimization

To meet KL divergence smaller (two gaussian distribuition이 같아지게끔.) => optimization

$$p_\theta(X_{t-1} \mid X_t) \approx q(X_{t-1} \mid X_t, X_0)$$

$$\arg\min_{\boldsymbol{\theta}} D_{\mathrm{KL}}(q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t, \boldsymbol{x}_0) \parallel p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t))$$

$$= \arg\min_{\boldsymbol{\theta}} D_{\mathrm{KL}}(\mathcal{N}(\boldsymbol{x}_{t-1}; \boldsymbol{\mu}_q, \boldsymbol{\Sigma}_q(t)) \parallel \mathcal{N}(\boldsymbol{x}_{t-1}; \boldsymbol{\mu}_{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_q(t)))$$

$$= \arg\min_{\boldsymbol{\theta}} \frac{1}{2\sigma_q^2(t)} \left[\left\| \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\boldsymbol{x}_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\hat{\boldsymbol{x}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)}{1-\bar{\alpha}_t} - \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\boldsymbol{x}_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\boldsymbol{x}_0}{1-\bar{\alpha}_t} \right\|_2^2 \right]$$

$$= \arg\min_{\boldsymbol{\theta}} \frac{1}{2\sigma_q^2(t)} \left[\left\| \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\hat{\boldsymbol{x}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t)}{1-\bar{\alpha}_t} - \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\boldsymbol{x}_0}{1-\bar{\alpha}_t} \right\|_2^2 \right]$$

$$= \arg\min_{\boldsymbol{\theta}} \frac{1}{2\sigma_q^2(t)} \left[\left\| \frac{\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)}{1-\bar{\alpha}_t} (\hat{\boldsymbol{x}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) - \boldsymbol{x}_0) \right\|_2^2 \right]$$

$$= \arg\min_{\boldsymbol{\theta}} \frac{1}{2\sigma_q^2(t)} \frac{\bar{\alpha}_{t-1}(1-\alpha_t)^2}{(1-\bar{\alpha}_t)^2} \left[\|\hat{\boldsymbol{x}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t, t) - \boldsymbol{x}_0\|_2^2 \right]$$

$\hat{x}_\theta(x_t, t)$ = 노이즈가 있는 이미지 x_t와 시간 인덱스 t를 입력받아 x_0를 예측하려고 하는 신경망에 의해 매개변수화 된 것.

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

$$\boldsymbol{\mu}_q(\boldsymbol{x}_t, \boldsymbol{x}_0) = \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\boldsymbol{x}_t + \sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\boldsymbol{x}_0}{1-\bar{\alpha}_t}$$

Diffusion model

Understanding Diffusion Models: A Unifted Perspective,
Calvin Luo(google Research, brain team), August 26, 2022

❑ Optimization

$$\boldsymbol{x}_0 = \frac{\boldsymbol{x}_t - \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_0}{\sqrt{\bar{\alpha}_t}}$$

$$\begin{aligned}
\frac{1}{2\sigma_q^2(t)}\frac{\bar{\alpha}_{t-1}(1-\alpha_t)^2}{(1-\bar{\alpha}_t)^2}\left[\|\hat{\boldsymbol{x}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t,t)-\boldsymbol{x}_0\|_2^2\right] &= \frac{1}{2\frac{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}}\frac{\bar{\alpha}_{t-1}(1-\alpha_t)^2}{(1-\bar{\alpha}_t)^2}\left[\|\hat{\boldsymbol{x}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t,t)-\boldsymbol{x}_0\|_2^2\right] \\
&= \frac{1}{2}\frac{1-\bar{\alpha}_t}{(1-\alpha_t)(1-\bar{\alpha}_{t-1})}\frac{\bar{\alpha}_{t-1}(1-\alpha_t)^2}{(1-\bar{\alpha}_t)^2}\left[\|\hat{\boldsymbol{x}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t,t)-\boldsymbol{x}_0\|_2^2\right] \\
&= \frac{1}{2}\frac{\bar{\alpha}_{t-1}(1-\alpha_t)}{(1-\bar{\alpha}_{t-1})(1-\bar{\alpha}_t)}\left[\|\hat{\boldsymbol{x}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t,t)-\boldsymbol{x}_0\|_2^2\right] \\
&= \frac{1}{2}\frac{\bar{\alpha}_{t-1}-\bar{\alpha}_t}{(1-\bar{\alpha}_{t-1})(1-\bar{\alpha}_t)}\left[\|\hat{\boldsymbol{x}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t,t)-\boldsymbol{x}_0\|_2^2\right] \\
&= \frac{1}{2}\frac{\bar{\alpha}_{t-1}-\bar{\alpha}_{t-1}\bar{\alpha}_t+\bar{\alpha}_{t-1}\bar{\alpha}_t-\bar{\alpha}_t}{(1-\bar{\alpha}_{t-1})(1-\bar{\alpha}_t)}\left[\|\hat{\boldsymbol{x}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t,t)-\boldsymbol{x}_0\|_2^2\right] \\
&= \frac{1}{2}\frac{\bar{\alpha}_{t-1}(1-\bar{\alpha}_t)-\bar{\alpha}_t(1-\bar{\alpha}_{t-1})}{(1-\bar{\alpha}_{t-1})(1-\bar{\alpha}_t)}\left[\|\hat{\boldsymbol{x}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t,t)-\boldsymbol{x}_0\|_2^2\right] \\
&= \frac{1}{2}\left(\frac{\bar{\alpha}_{t-1}(1-\bar{\alpha}_t)}{(1-\bar{\alpha}_{t-1})(1-\bar{\alpha}_t)}-\frac{\bar{\alpha}_t(1-\bar{\alpha}_{t-1})}{(1-\bar{\alpha}_{t-1})(1-\bar{\alpha}_t)}\right)\left[\|\hat{\boldsymbol{x}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t,t)-\boldsymbol{x}_0\|_2^2\right] \\
&= \frac{1}{2}\left(\frac{\bar{\alpha}_{t-1}}{1-\bar{\alpha}_{t-1}}-\frac{\bar{\alpha}_t}{1-\bar{\alpha}_t}\right)\left[\|\hat{\boldsymbol{x}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t,t)-\boldsymbol{x}_0\|_2^2\right]
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{\mu}_q(\boldsymbol{x}_t,\boldsymbol{x}_0) &= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\boldsymbol{x}_t+\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\boldsymbol{x}_0}{1-\bar{\alpha}_t} \\
&= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\boldsymbol{x}_t+\sqrt{\bar{\alpha}_{t-1}}(1-\alpha_t)\frac{\boldsymbol{x}_t-\sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_0}{\sqrt{\bar{\alpha}_t}}}{1-\bar{\alpha}_t} \\
&= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\boldsymbol{x}_t+(1-\alpha_t)\frac{\boldsymbol{x}_t-\sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_0}{\sqrt{\alpha_t}}}{1-\bar{\alpha}_t} \\
&= \frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})\boldsymbol{x}_t}{1-\bar{\alpha}_t}+\frac{(1-\alpha_t)\boldsymbol{x}_t}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}}-\frac{(1-\alpha_t)\sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}_0}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}} \\
&= \left(\frac{\sqrt{\alpha_t}(1-\bar{\alpha}_{t-1})}{1-\bar{\alpha}_t}+\frac{1-\alpha_t}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}}\right)\boldsymbol{x}_t-\frac{(1-\alpha_t)\sqrt{1-\bar{\alpha}_t}}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}}\boldsymbol{\epsilon}_0 \\
&= \left(\frac{\alpha_t(1-\bar{\alpha}_{t-1})}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}}+\frac{1-\alpha_t}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}}\right)\boldsymbol{x}_t-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}}\boldsymbol{\epsilon}_0 \\
&= \frac{\alpha_t-\bar{\alpha}_t+1-\alpha_t}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}}\boldsymbol{x}_t-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}}\boldsymbol{\epsilon}_0 \\
&= \frac{1-\bar{\alpha}_t}{(1-\bar{\alpha}_t)\sqrt{\alpha_t}}\boldsymbol{x}_t-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}}\boldsymbol{\epsilon}_0 \\
&= \frac{1}{\sqrt{\alpha_t}}\boldsymbol{x}_t-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}}\boldsymbol{\epsilon}_0
\end{aligned}$$

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

$$\begin{aligned}
&\arg\min_{\boldsymbol{\theta}} D_{\mathrm{KL}}(q(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t,\boldsymbol{x}_0)\parallel p_{\boldsymbol{\theta}}(\boldsymbol{x}_{t-1}|\boldsymbol{x}_t)) \\
&= \arg\min_{\boldsymbol{\theta}} D_{\mathrm{KL}}(\mathcal{N}(\boldsymbol{x}_{t-1};\boldsymbol{\mu}_q,\boldsymbol{\Sigma}_q(t))\parallel\mathcal{N}(\boldsymbol{x}_{t-1};\boldsymbol{\mu}_{\boldsymbol{\theta}},\boldsymbol{\Sigma}_q(t))) \\
&= \arg\min_{\boldsymbol{\theta}} \frac{1}{2\sigma_q^2(t)}\left[\left\|\frac{1}{\sqrt{\alpha_t}}\boldsymbol{x}_t-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t,t)-\frac{1}{\sqrt{\alpha_t}}\boldsymbol{x}_t+\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}}\boldsymbol{\epsilon}_0\right\|_2^2\right] \\
&= \arg\min_{\boldsymbol{\theta}} \frac{1}{2\sigma_q^2(t)}\left[\left\|\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}}\boldsymbol{\epsilon}_0-\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}}\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t,t)\right\|_2^2\right] \\
&= \arg\min_{\boldsymbol{\theta}} \frac{1}{2\sigma_q^2(t)}\left[\left\|\frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}\sqrt{\alpha_t}}(\boldsymbol{\epsilon}_0-\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t,t))\right\|_2^2\right] \\
&= \arg\min_{\boldsymbol{\theta}} \frac{1}{2\sigma_q^2(t)}\frac{(1-\alpha_t)^2}{(1-\bar{\alpha}_t)\alpha_t}\left[\|\boldsymbol{\epsilon}_0-\hat{\boldsymbol{\epsilon}}_{\boldsymbol{\theta}}(\boldsymbol{x}_t,t)\|_2^2\right]
\end{aligned}$$

$\hat{\epsilon}_\theta(x_t,t)$ = *neural network that learns to predict the source noise*

$\mu_q(x_t,x_0)$ = *true denoising transition mean*

$\mu_\theta(x_t,t)$ = *approximate denoising transition mean*

Source noise(ground truth's noise) $= \epsilon_0 \sim N(\epsilon;\mathbf{0},\boldsymbol{I})$

CVSP

5

Conclusion

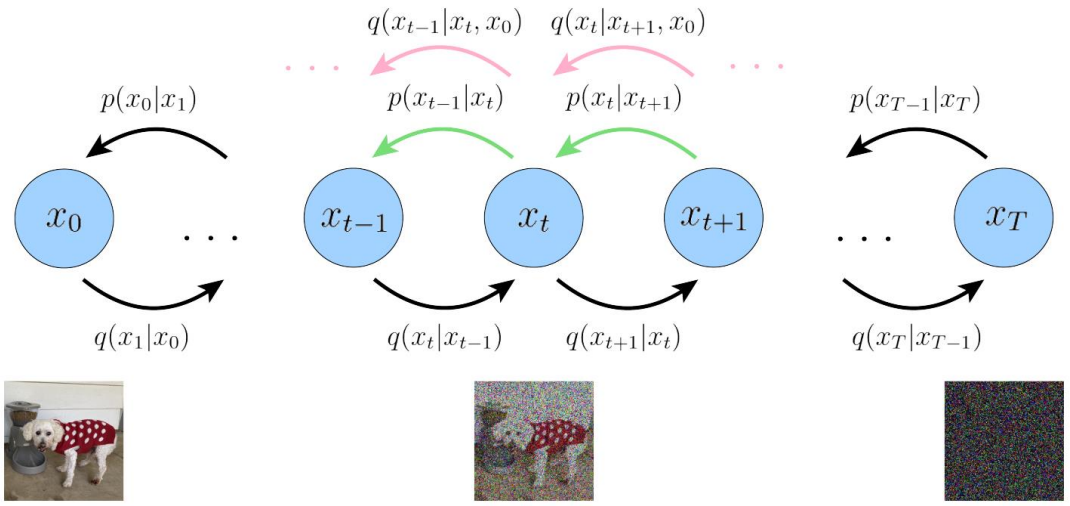
Diffusion model

Understanding Diffusion Models: A Unifted Perspective,
Calvin Luo(google Research, brain team), August 26, 2022

❑ Conclusion

Diffusion model은 forward process(noising)를 통해 gaussian noise를 생성하고 reverse process(denoising)를 통해 gaussian noise로부터 nosie를 제거함으로써 새로운 이미지를 생성하는 generative model이다.

학습목표로는
1) neural network로 하여금 noise로부터 original ground truth(x0)를 예측하는 것.
2) 매 스텝마다 KL divergence를 최소화 시키는 것. (reverse process에서 ground truth를 알고 x0를 예측하는 분포와 ground truth 를 모르고, denoising하는 분포간의 차이를 거의 없게끔 두 분포를 거의 같게끔 만드는 것
※ predict original image = predict noise



CVSP

Thank you for listening