

20기 정규세션

ToBig's 19기 강연자

위 성 진

# 19. Generative Advanced

# Contents

---

Unit 01 | Introduction

---

Unit 02 | What is diffusion?

---

Unit 03 | Diffusion Process

---

Unit 04 | Latent Diffusion

---

Unit 05 | Assignments

---

## Unit 01 | Introduction

## 18·19기 투빅스 컨퍼런스

## Sketchbook – 간단한 스케치로 로고 제공 모델 (InstructPix2Pix)

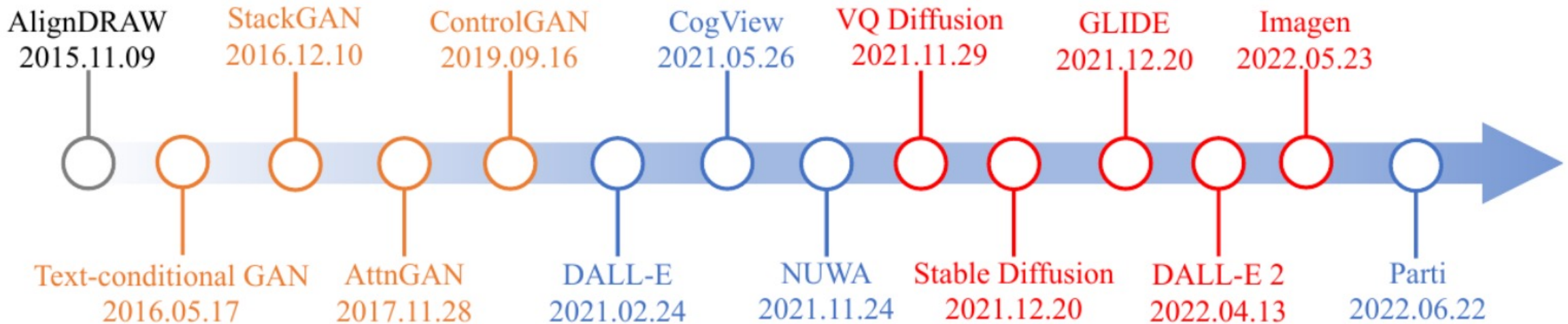


## ToAnywhere – Segmentation task를 활용한 Conditional Inpainting (StableDiffusion Inpainting)

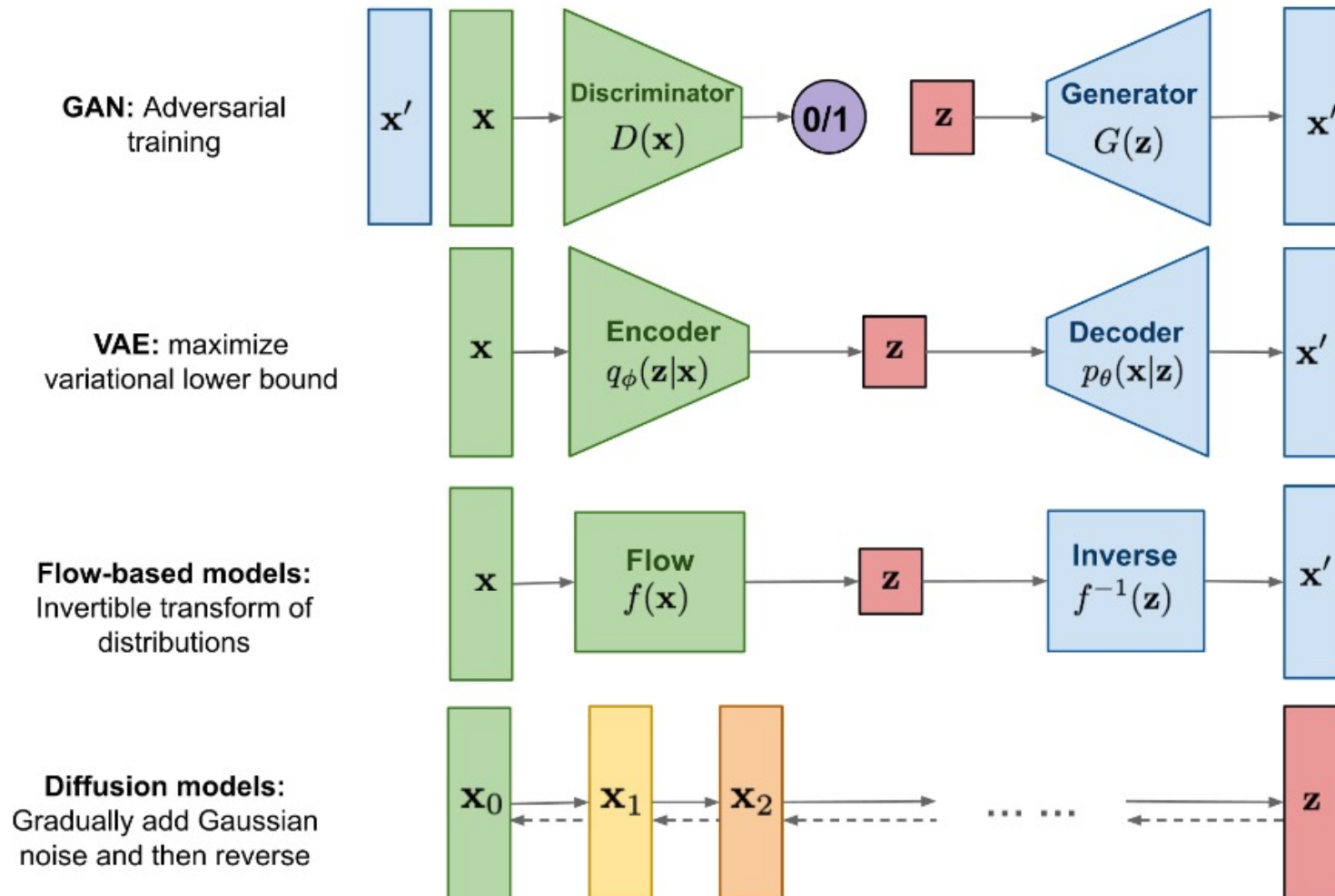


## Unit 01 | Introduction

## Representative works on text-to-image task over time



## Unit 01 | Introduction



## Unit 02 | What is diffusion?

## Markov Chain: Markov 성질을 가지는 이산 확률과정

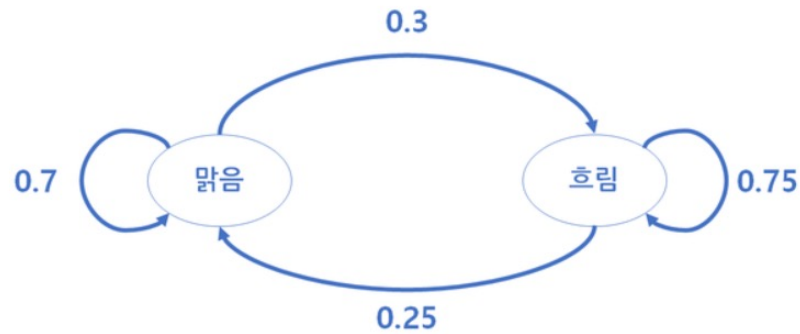
$$P(x^{(t+1)} | x^{(0)}, \dots, x^{(t-1)}, x^{(t)}) = P(x^{(t+1)} | x^{(t)})$$

$\Rightarrow x^{(t+1)}$  is independent of  $\{x^{(0)}, \dots, x^{(t-1)}\}$  given  $x^{(t)}$

- 특정 시점( $t+1$ )의 상태 확률은 그 직전( $t$ ) 상태에만 의존
- 한 상태에서 다른 상태로의 전이(transition)는 그동안 상태 전이에 대한 긴 이력(history)을 필요로 하지 않고 바로 직전 상태에서의 전이로 추정할 수 있음.

## Unit 02 | What is diffusion?

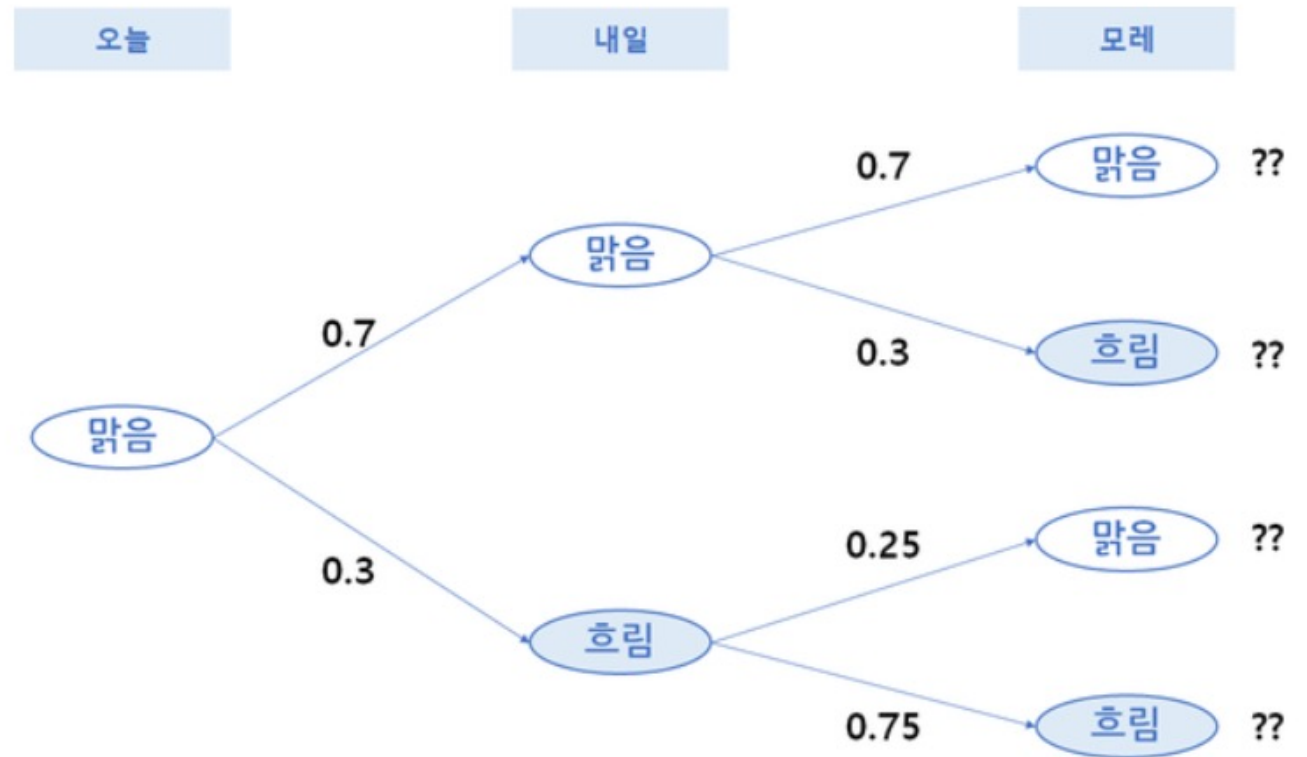
## Markov Chain: Markov 성질을 가지는 이산 확률과정



상태 전이도 예시

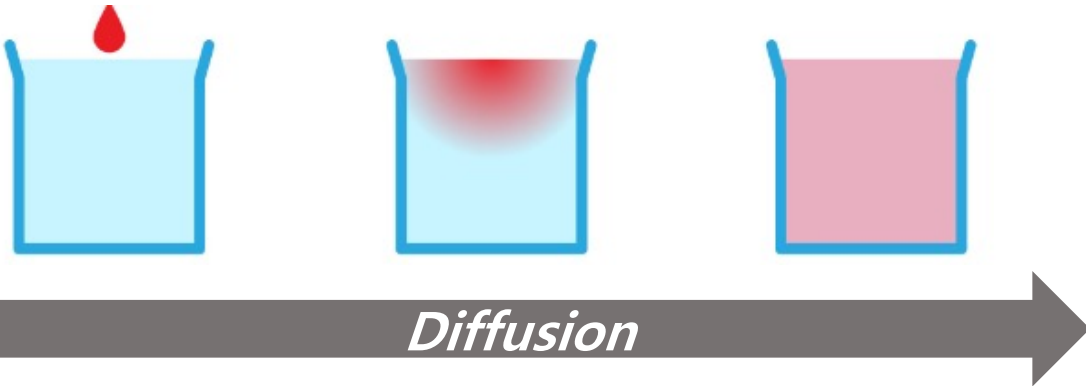
오늘/내일	맑음	흐림
맑음	0.7	0.3
흐림	0.25	0.75

오늘 날씨에 기준하여 내일 특정 날씨가 발생할 확률



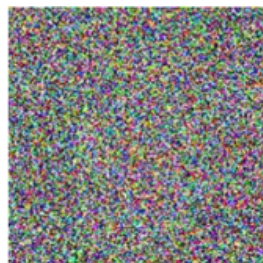
## Unit 02 | What is diffusion?

## Diffusion (확산)



## Dynamics

- 액체나 기체에 다른 물질이 섞이고, 그것이 조금씩 번져가다가 마지막에는 일률적인 농도로 바뀌는 현상



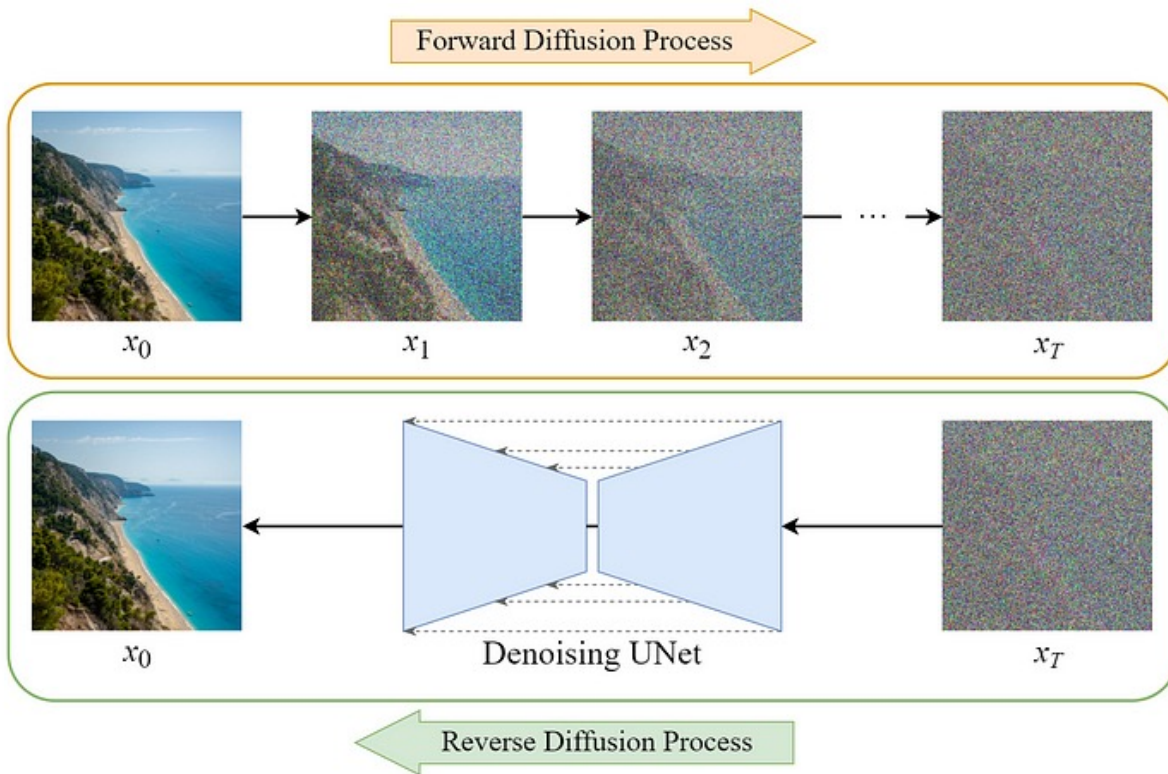
## Generative AI

- 입력 데이터에 Noise를 반복적으로 추가해 데이터의 패턴이 서서히 와해 되는 과정 (forward diffusion process)



## Unit 02 | What is diffusion?

## Diffusion Model



- **Generative model** – 학습 데이터의 패턴(확률 분포)을 따르는 데이터를 생성
- **Forward diffusion process**에서 noise를 추가해 데이터의 패턴을 와해시키고, **Reverse diffusion process**에서 이를 복원하는 과정을 학습함으로써 패턴의 생성 과정을 학습.
- 이 noising, denoising을 하나의 단일 step transformation으로 학습하는 것은 매우 어려운 과제이기 때문에 **여러 단계로 쪼개어진 Markov Chain**으로 각 Process를 구성
- 위 과정을 통해 **gaussian과 같은 간단한 분포(z)**를 학습 데이터의 패턴을 갖는 분포로 **변환(mapping)**하는 모델을 얻고자 함

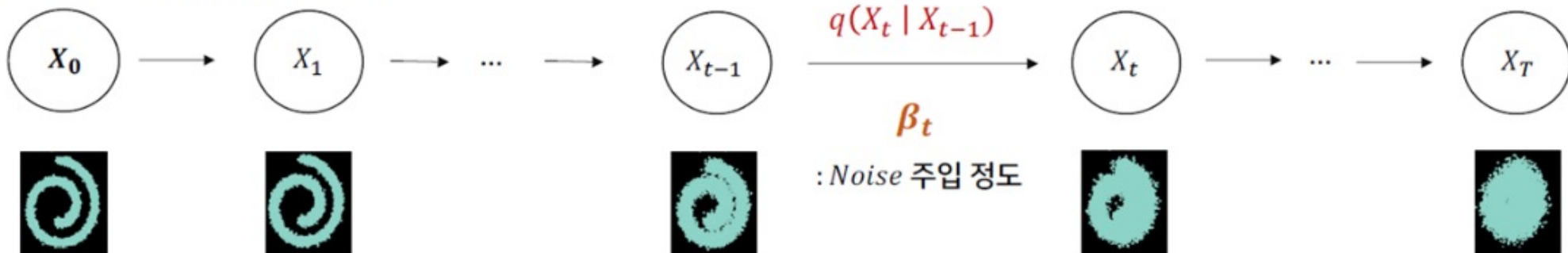
## Unit 03 | Diffusion Process

## Forward Diffusion Process

- 데이터에 gaussian noise를 점진적으로 더해가며 isotropic gaussian으로 만들어가는 과정
- 더해지는 noise는 사전 정의(fixed)된 gaussian 분포에서 생성됨

$$q(X_t | X_{t-1}) := N(X_t; \mu_{X_{t-1}}, \Sigma_{X_{t-1}}) := N(X_t; \sqrt{1 - \beta_t} X_{t-1}, \beta_t \cdot I)$$

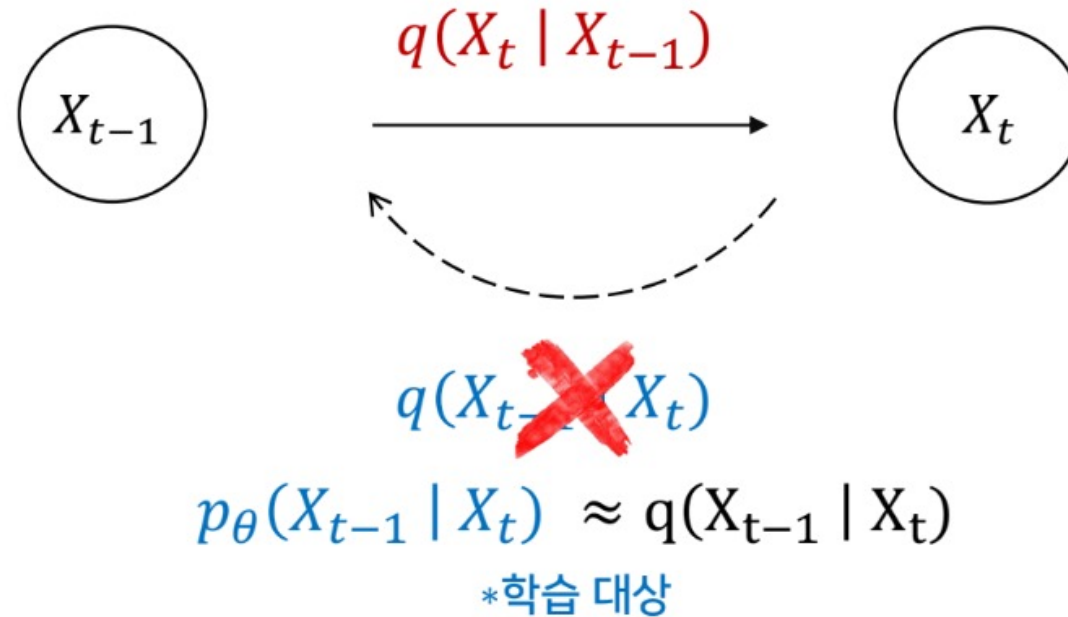
## Forward Diffusion Process



## Unit 03 | Diffusion Process

## Reverse Diffusion Process

- Forward diffusion process의 역 과정(Denoising)을 학습
- $p_{\theta}(X_{t-1}|X_t)$ 를 통해  $q(X_{t-1}|X_t)$ 를 approximation

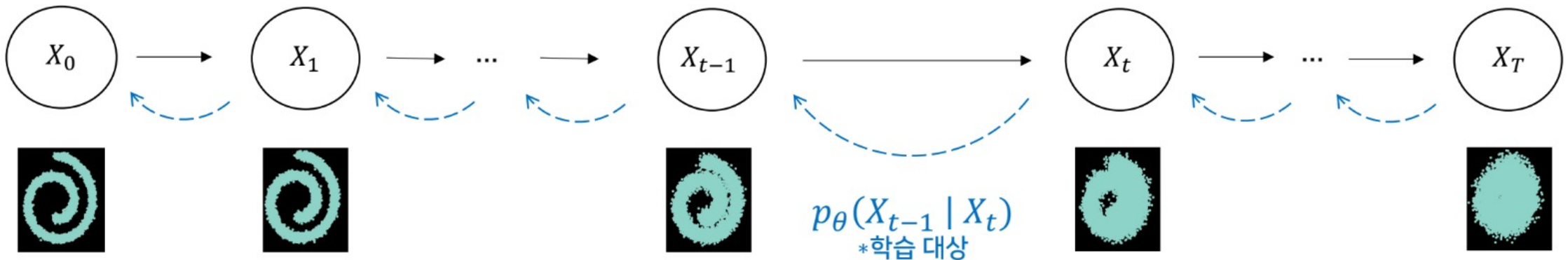


## Unit 03 | Diffusion Process

## Reverse Diffusion Process

- Gaussian noise를 제거해가며 특정한 패턴을 만들어가는 forward diffusion process의 역과정

$$p_{\theta}(X_{0:T}) := p(X_T) \prod_{t=1}^T q(X_{t-1} | X_t), \quad p_{\theta}(X_{t-1} | X_t) := N(X_{t-1}; \underbrace{\mu_{\theta}(X_t, t)}_{\text{학습 대상}}, \underbrace{\Sigma_{\theta}(X_t, t)}_{\text{(mean \& variance function)}})$$



## Unit 03 | Diffusion Process

## Diffusion Loss

[Deep Unsupervised Learning using Nonequilibrium Thermodynamics](#)(2015)

$$\begin{aligned}
 LOSS_{Diffusion} &= D_{KL}(q(z | x_0) || P_{\theta}(x_0 | z)) - E_{z \sim q(z|x)} [\log P_{\theta}(z)] \\
 &= \underbrace{D_{KL}(q(z | x_0) || P_{\theta}(z))}_{\text{Regularizer on Encoder}} + \underbrace{\sum_{t=2} D_{KL}(q(x_{t-1} | x_t, x_0) || P_{\theta}(x_{t-1} | x_t))}_{\text{Denoising Process}} - \underbrace{E_q[\log P_{\theta}(x_0 | x_1)]}_{\text{Reconstruction on Decoder}}
 \end{aligned}$$



[Denoising Diffusion Probabilistic Models](#)(2020)

$$LOSS_{DDPM} = \mathbb{E}_{x_0, \epsilon} \left[ \left| \epsilon - \epsilon_{\theta} \left( \sqrt{\tilde{\alpha}_t} + \sqrt{1 - \tilde{\alpha}_t} \epsilon, t \right) \right|^2 \right]$$

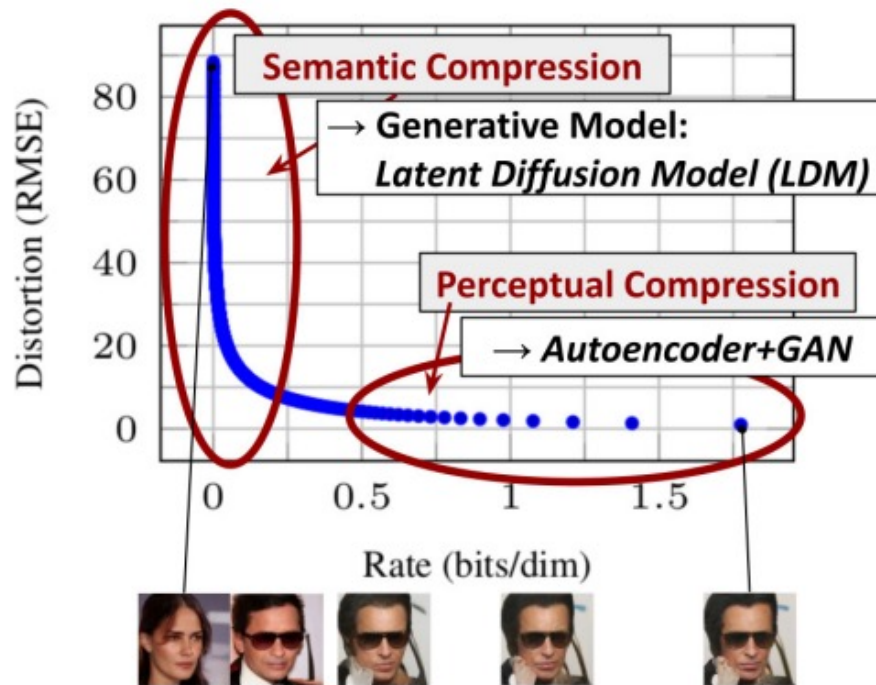
## Unit 04 | Latent Diffusion

### Latent diffusion Motivation

- **Pixel space**에서 모델을 학습하는 것은 매우 방대한 컴퓨팅 리소스를 필요로 함.
- 기존 diffusion model의 성능을 유지하면서 **연산량을 줄일 필요**가 있음

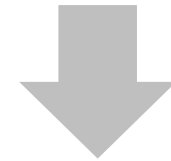
## Unit 04 | Latent Diffusion

## Latent diffusion Motivation



## Rate-distortion trade-off of a trained model

- 이미지의 많은 bits가 imperceptible detail에 대응
- 기존 Diffusion model은 semantic한 의미가 없는 정보를 suppress하는데 전체 pixel에 대한 loss, gradient를 사용. 때문에 불필요하게 많은 연산이 이루어짐

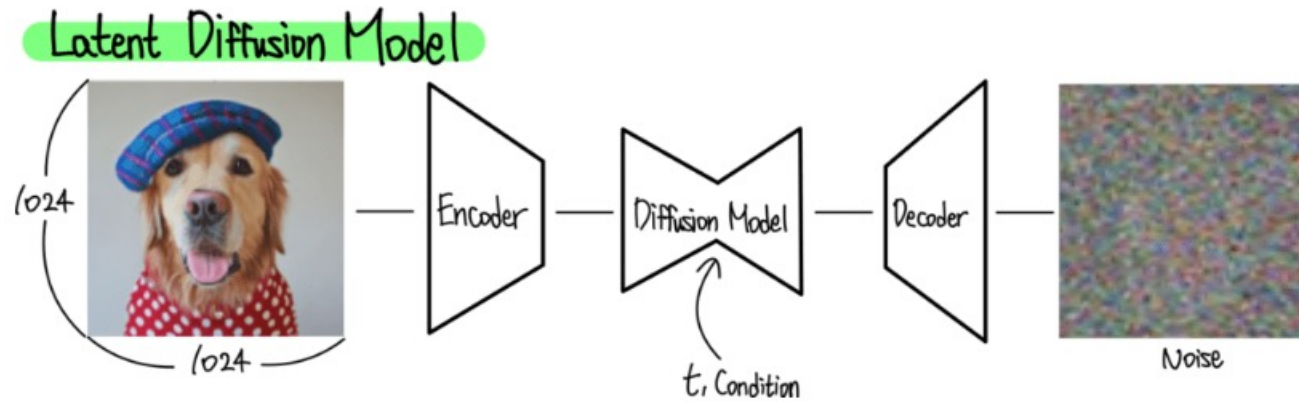
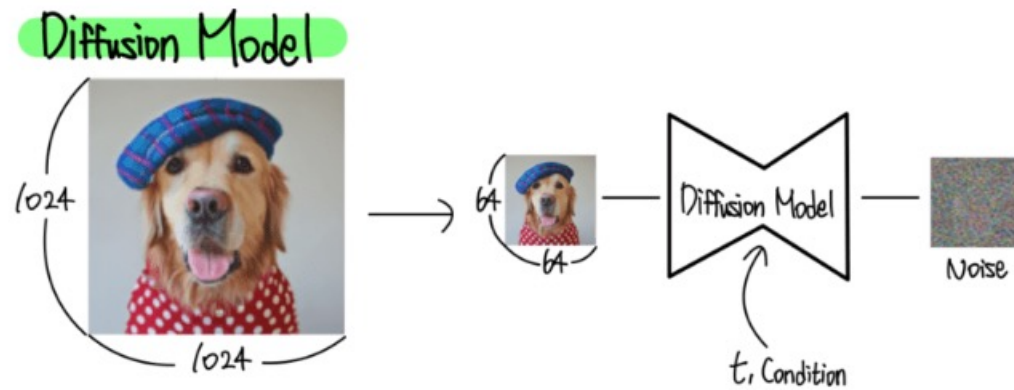


## 학습을 2단계로 구성

- **Perceptual compression:** high-frequency detail들을 제거하지만 의미(semantic)는 거의 학습하지 않는 단계
- **Semantic compression:** 데이터의 의미론적(semantic) 구성과 개념적(conceptual) 구성을 학습하는 실질적인 생성모델의 단계

## Unit 04 | Latent Diffusion

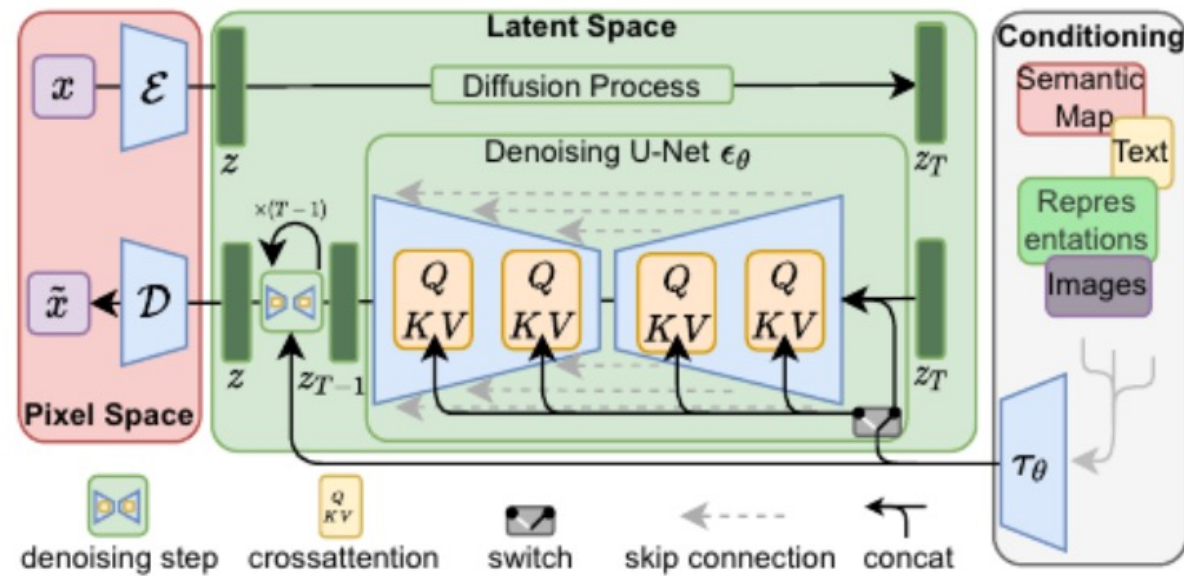
## Latent diffusion Model





## Unit 04 | Latent Diffusion

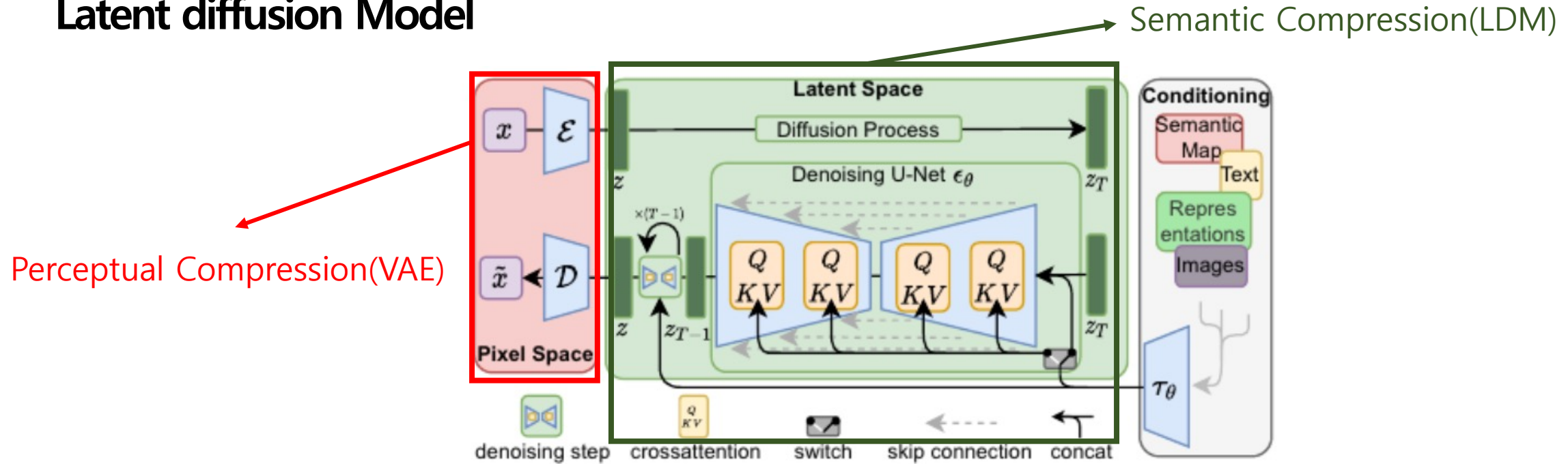
## Latent diffusion Model



- **Pixel space**(high dimensional) 대신 **Latent space**(Low dimensional)에서 diffusion process를 수행함으로써 연산량을 크게 줄임
- 보다 중요한, Semantic 정보들에 더 집중할 수 있음

## Unit 04 | Latent Diffusion

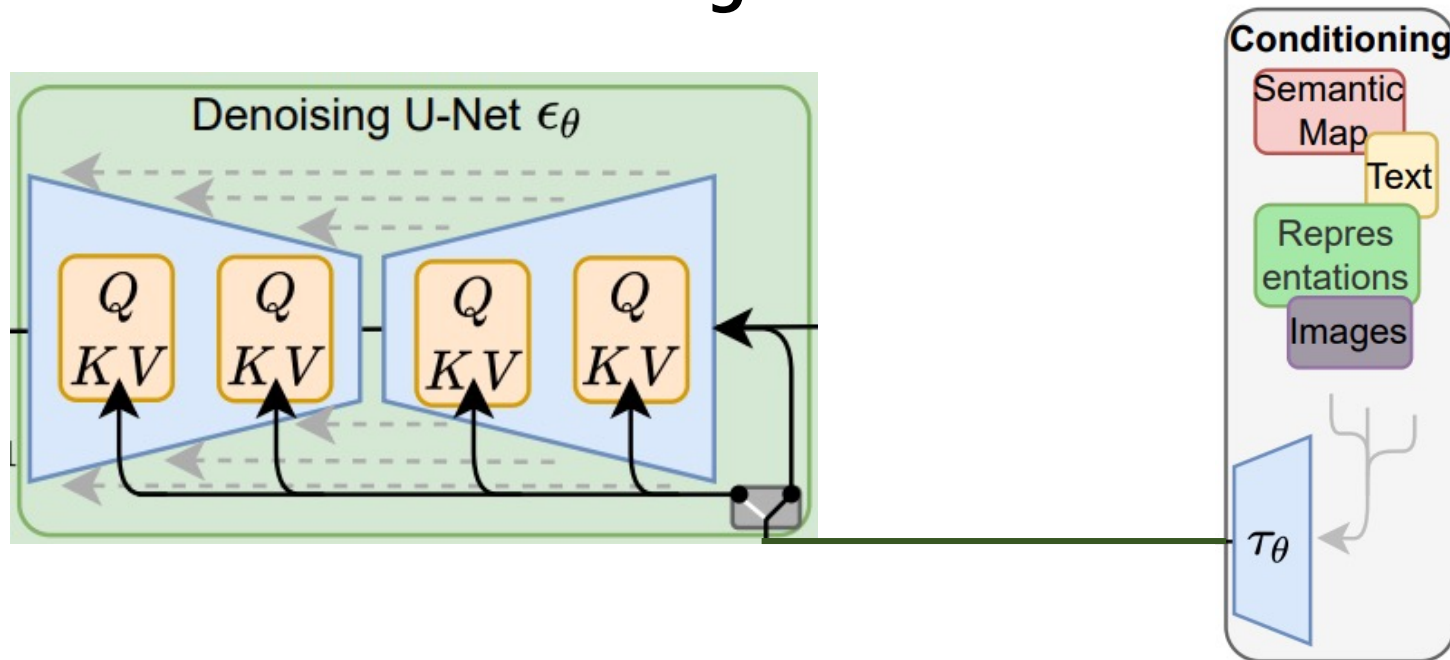
## Latent diffusion Model



- Pixel space(high dimensional) 대신 Latent space(Low dimensional)에서 diffusion process를 수행함으로써 연산량을 크게 줄임
- 보다 중요한, Semantic 정보들에 더 집중할 수 있음

## Unit 04 | Latent Diffusion

## Latent diffusion Model – Conditioning Mechanism



- **Cross-attention**을 UNet backbone에 사용하여 **다양한 modalities의 conditioning**이 가능하도록 하였음
- UNet의 intermediate representation을 flatten하고 이를 Query값으로 **domain specific encoder**를 사용하여 만든 conditioning 정보를 Key, Value로 사용하여 cross-attention 적용

## Unit 04 | Latent Diffusion

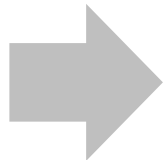
## Latent diffusion Model

$$L_{DM} = \mathbb{E}_{x, \epsilon \sim \mathcal{N}(0,1), t} \left[ \|\epsilon - \epsilon_{\theta}(x_t, t)\|_2^2 \right]$$

$$L_{LDM} := \mathbb{E}_{\mathcal{E}(x), y, \epsilon \sim \mathcal{N}(0,1), t} \left[ \|\epsilon - \epsilon_{\theta}(z_t, t, \tau_{\theta}(y))\|_2^2 \right]$$

## Unit 05 | Assignments

1. 대표적인 이미지 생성 모델들 – GAN, VAE, Flow-base model, Diffusion 각각의 개념, 구조, 차이점을 정리해서 작성
2. Latent diffusion 논문([High-Resolution Image Synthesis with Latent Diffusion Models](#)) 리뷰 작성.  
혹은, 그 외의 diffusion 관련 논문도 괜찮습니다!

 **중 택1**하여 자유로운 형식(Notion, Pdf)등으로 제출해주세요!

# References

## Youtube

- [Stable diffusion 논문 리뷰 \(딥러닝논문읽기모임\)](#)
- [CS 198-126: Lecture 12 - Diffusion Models](#) (Machine Learning at Berkeley)
- [\[Paper Review\] Denoising Diffusion Probabilistic Models](#)

## Posts

- [\[논문리뷰\] High-Resolution Image Synthesis with Latent Diffusion Models \(Stable Diffusion\)](#)
- [\[22' CVPR\] High-Resolution Image Synthesis with Latent Diffusion Models \(Stable Diffusion\)](#)
- [LDM: High-Resolution Image Synthesis with Latent Diffusion Models](#)
- [What are Diffusion Models](#)

## Papers

- [Deep Unsupervised Learning using Nonequilibrium Thermodynamics](#)
- [Denoising Diffusion Probabilistic Models](#)
- [High-Resolution Image Synthesis with Latent Diffusion Models](#)

Q & A

들어주셔서 감사합니다.