

Generative Adversarial Network (GAN)

Jinho Park

POSTECH Graduate School of Artificial Intelligence (GSAI)

jhpark756@postech.ac.kr

2023. 6. 21

Contents

■ Part 1 - Generative Adversarial Network (GAN)

- 생성 모델 (Generative Model)
- Generative Adversarial Network (GAN)
- GAN의 학습 과정
- GAN의 단점
- GAN의 응용
- Practice 1

■ Part 2 - CycleGAN

- CycleGAN
- CycleGAN의 학습 과정
- Practice 2

Part 1

Generative Adversarial Network (GAN)

기계학습 알고리즘의 분류

학습 방식에 따라 크게 3가지로 분류됨!

- **지도학습 (Supervised Learning)**

라벨이 주어진 데이터로 모델을 학습시키는 방식

입력 데이터 x 에 대한 출력 값 y 를 예측

작업 예시: 분류 (classification), 회귀 (regression), ...

- **비지도학습 (Unsupervised Learning)**

라벨이 주어지지 않은 데이터로 모델을 학습시키는 방식

데이터 자체의 특성과 정보를 학습

작업 예시: 생성 (generation), 군집화 (clustering), ...

- **강화학습 (Reinforcement Learning)**

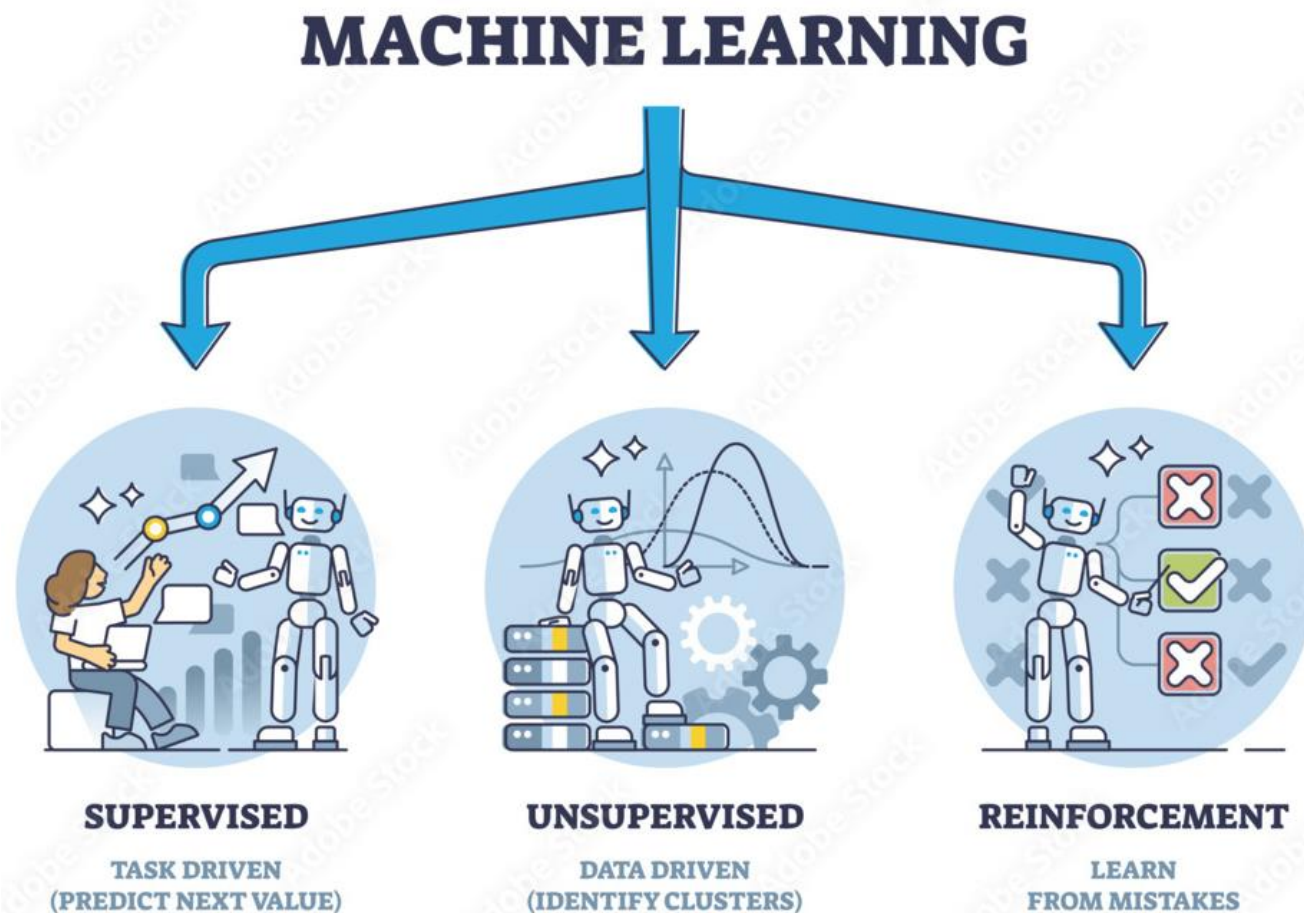
모델이 환경과 상호작용하면서 스스로 최적의 행동을 학습하는 방식

환경과의 상호작용은 Reward나 Punishment의 형태로 표현됨

작업 예시: 인공지능 (artificial intelligence), ...

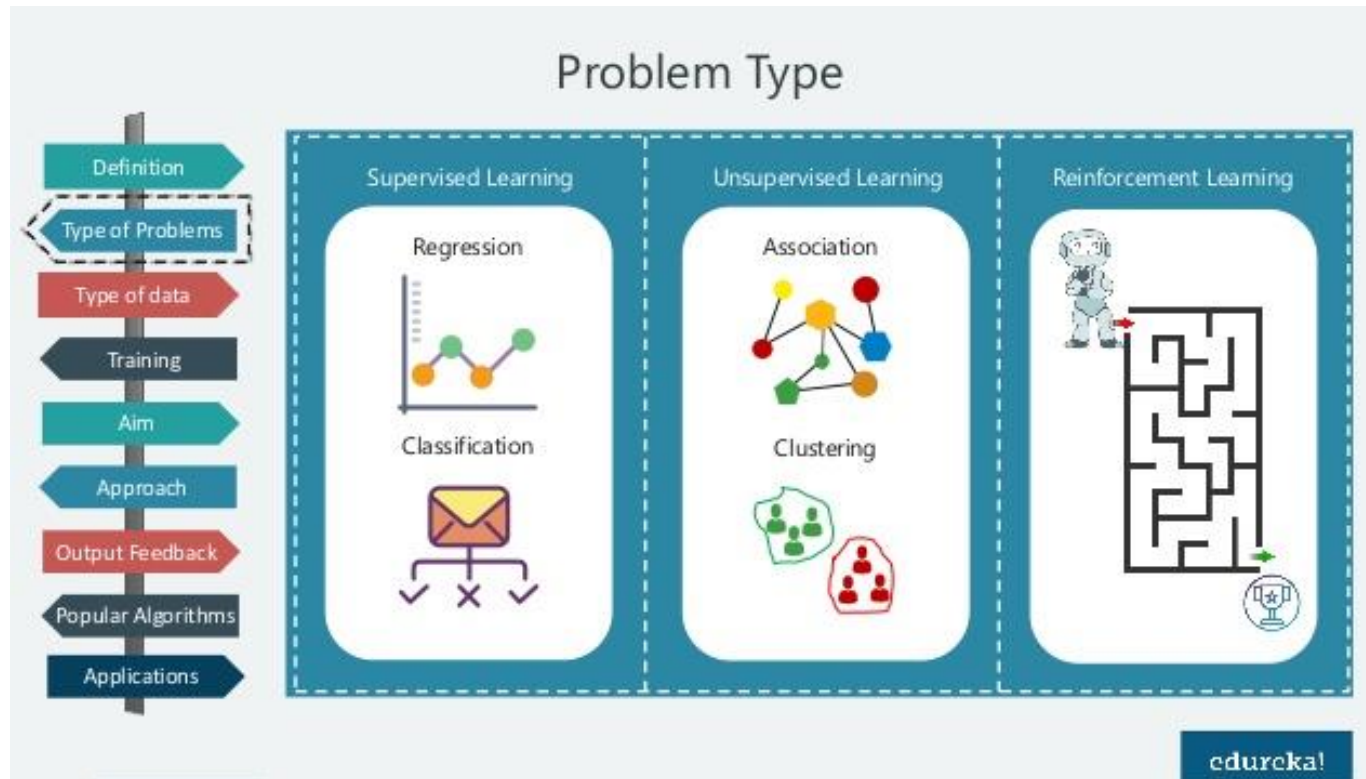
기계학습 알고리즘의 분류

기계학습의 3가지 알고리즘



기계학습 알고리즘의 분류

기계학습의 3가지 알고리즘



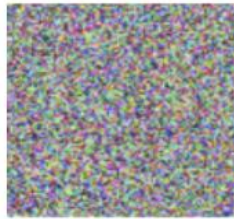
생성 모델 (Generative Model)

비지도학습 (Unsupervised Learning)의 응용 방향 중 하나!

학습에 사용된 데이터들과 유사한 새로운 데이터를 만드는 모델

학습에 사용된 데이터의 분포 $p_{\text{data}}(\mathbf{x})$ 의 학습 후, 이를 기반으로 새로운 데이터 \mathbf{x}' 생성

Noise $\sim N(0,1)$

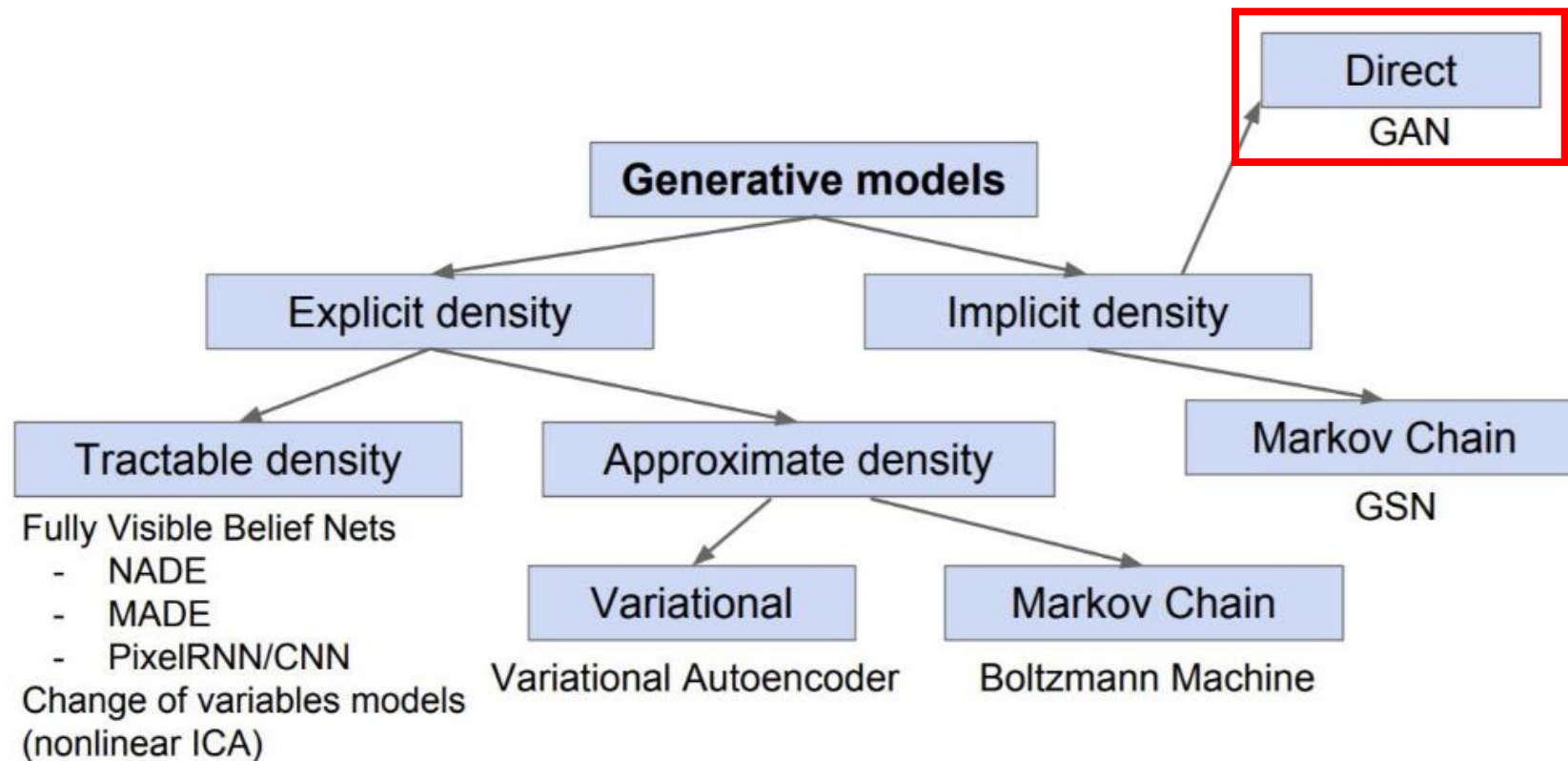


Generative
Model



생성 모델 (Generative Model)

생성 모델의 종류



Generative Adversarial Network (GAN)

GAN의 의미

Generative Adversarial Network

생성적

적대적

신경망

Generative Adversarial Network (GAN)

GAN의 의미

Generative Adversarial Network

생성적

적대적

신경망



학습에 사용된 데이터를 기반으로
이와 유사한 데이터를 스스로 생성할 수 있음
- 모델 내의 생성자 (generator) 가 담당

Generative Adversarial Network (GAN)

GAN의 의미

Generative Adversarial Network

생성적 적대적 신경망



학습에 사용하는 데이터의 분포를 모델이 학습하려면,
학습에 사용되는 진짜 데이터 (true data)와
생성자가 생성하는 가짜 데이터 (fake data)를 구분할 수 있어야 함!
- 이를 담당하는 모델 내의 **판별자 (discriminator)** 는 생성자와 적대적!

Generative Adversarial Network (GAN)

GAN의 의미

Generative Adversarial Network

생성적

적대적

신경망



진짜 데이터와 모조 데이터를 기반으로
학습이 이루어지는 기계학습 모델

Generative Adversarial Network (GAN)

GAN의 의미

Generative Adversarial Network

생성적 적대적 신경망

즉, GAN은 **생성자 (generator) 와 판별자 (discriminator) 사이의 경쟁**을 통해 학습한 뒤,
생성자로 학습에 사용된 데이터와 유사한 데이터를 생성하는 방식의
생성 모델을 의미!

Generative Adversarial Network (GAN)

GAN의 의미

Generative Adversarial Network

생성적 적대적 신경망

생성자와 판별자 사이의 경쟁이란?

생성자: 판별자가 진짜 데이터라고 잘못 판별하게 만들 가짜 데이터 생성

판별자: 진짜 데이터와 가짜 데이터를 성공적으로 구분

Generative Adversarial Network (GAN)

GAN의 비유 - 위조 지폐 제작 과정



진짜 지폐



위조 지폐

어떻게 하면 가짜라는 것을 쉽게 들키지 않는
위조 지폐를 만들 수 있을까?

Generative Adversarial Network (GAN)

GAN의 비유 - 위조 지폐 제작 과정

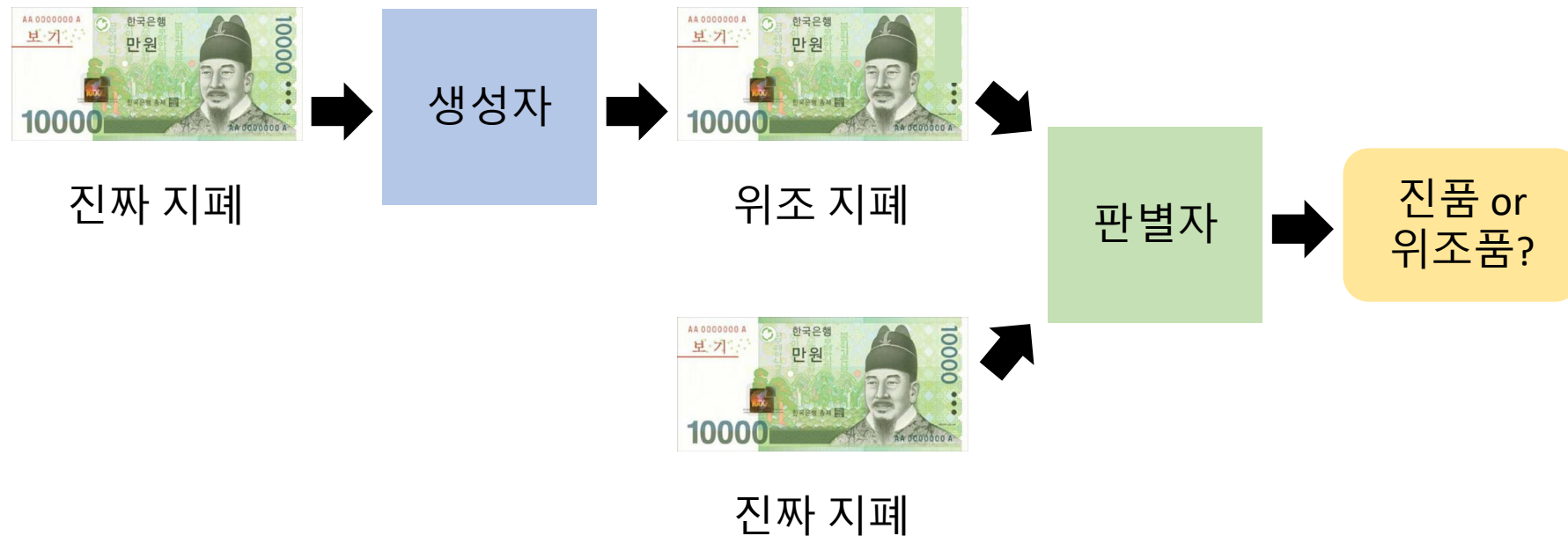
전체 흐름도



Generative Adversarial Network (GAN)

GAN의 비유 - 위조 지폐 제작 과정

학습이 충분히 이루어지기 전의 상황

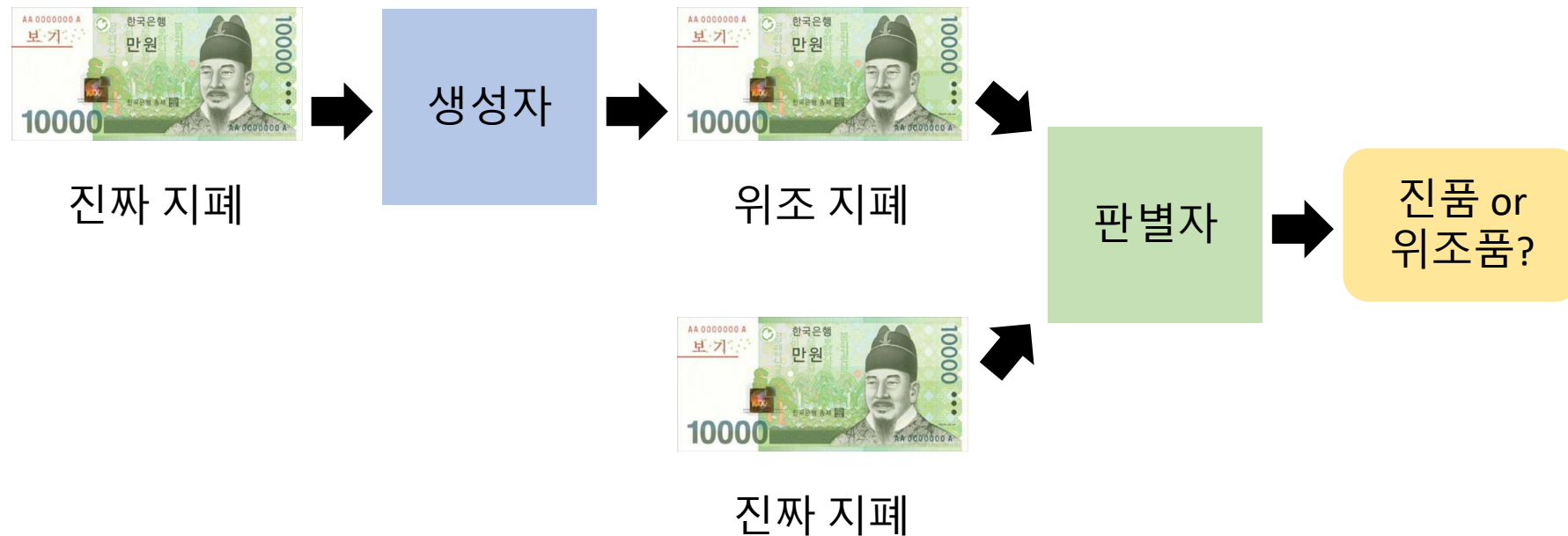


어설픈 위조 지폐 → 위조품임이 쉽게 판별됨!

Generative Adversarial Network (GAN)

GAN의 비유 - 위조 지폐 제작 과정

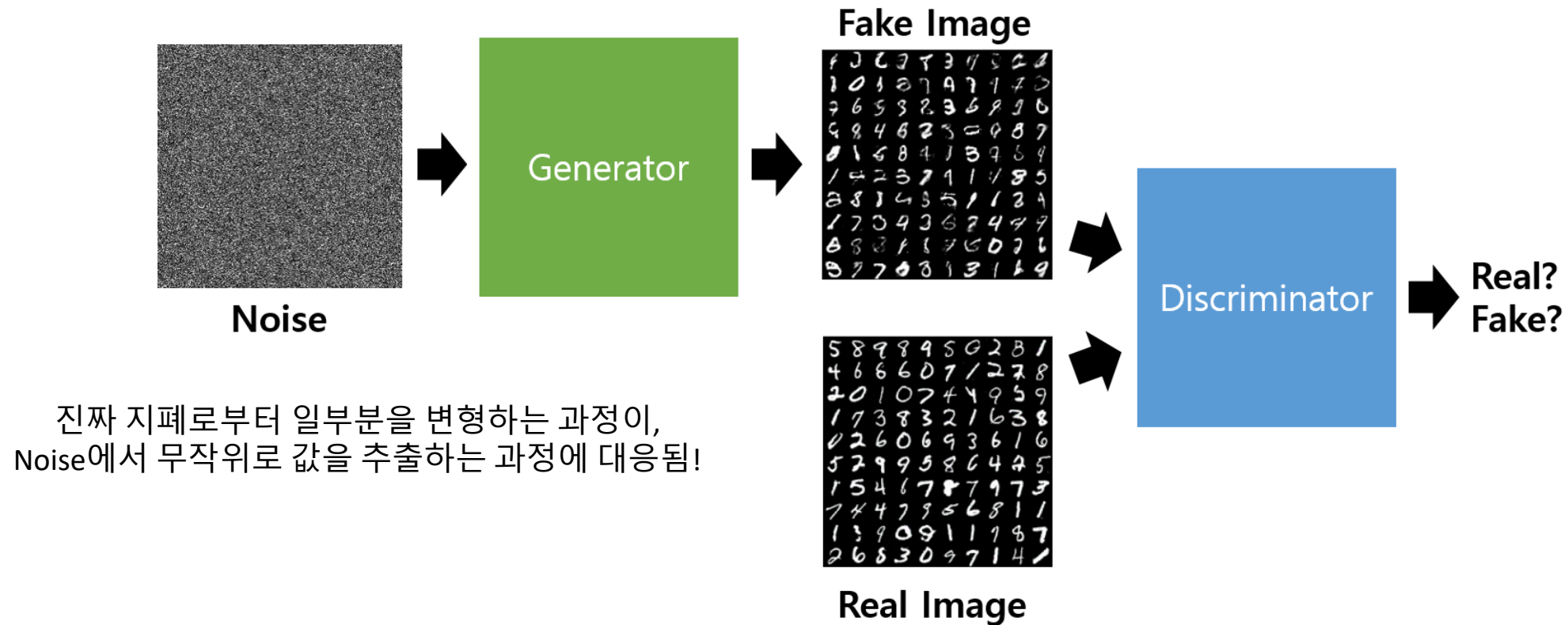
학습이 충분히 이루어진 뒤의 상황



진짜 지폐와 구분이 어려운 위조 지폐 → 진품으로 잘못 판별될 여지가 생김!

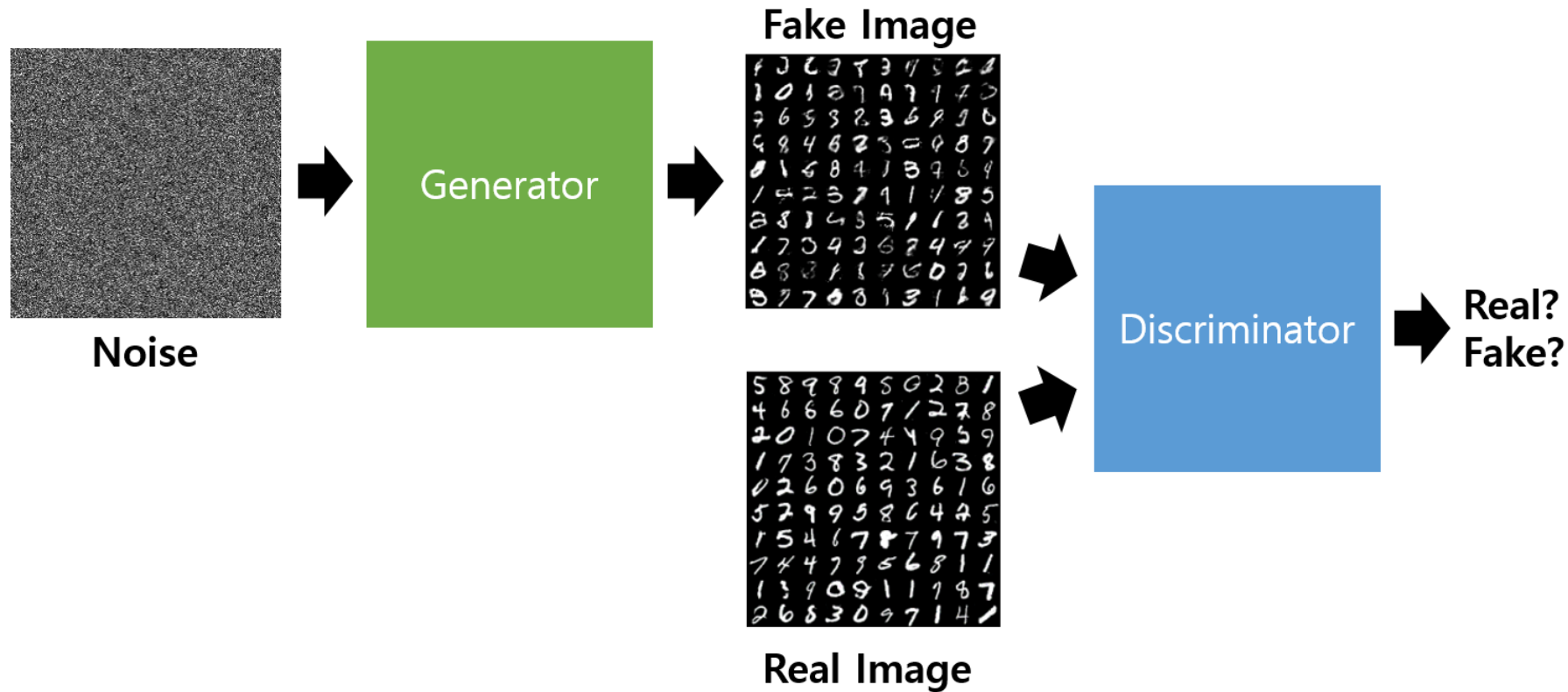
Generative Adversarial Network (GAN)

GAN의 전체 구조



Generative Adversarial Network (GAN)

GAN의 전체 구조



생성자와 판별자 사이의 경쟁을 통해, 생성자는 실제 데이터와 유사한 데이터를 생성할 수 있게 됨!

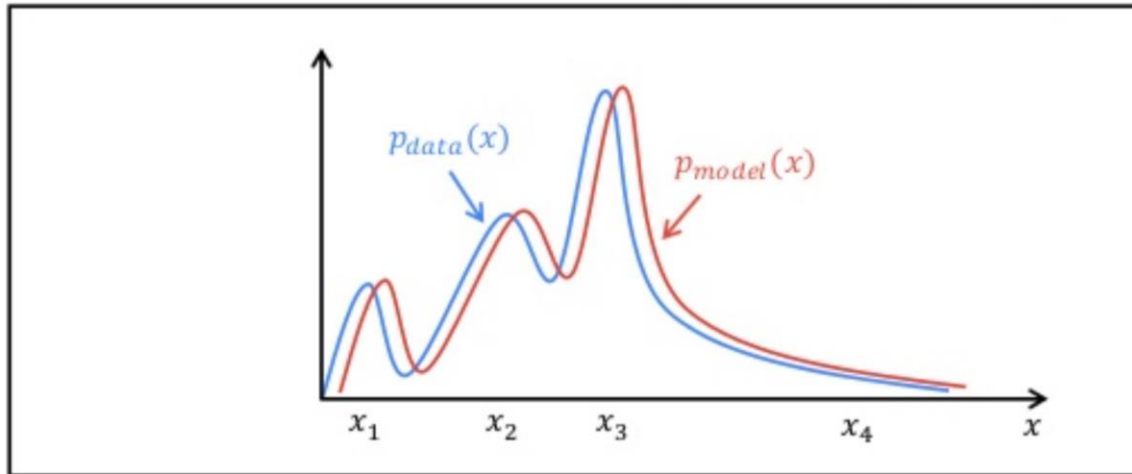
GAN의 학습 과정

생성 모델의 학습 목표

The goal of the generative model is to find a $p_{model}(x)$ that approximates $p_{data}(x)$ well.

↗ Distribution of images generated by the model

↘ Distribution of actual images

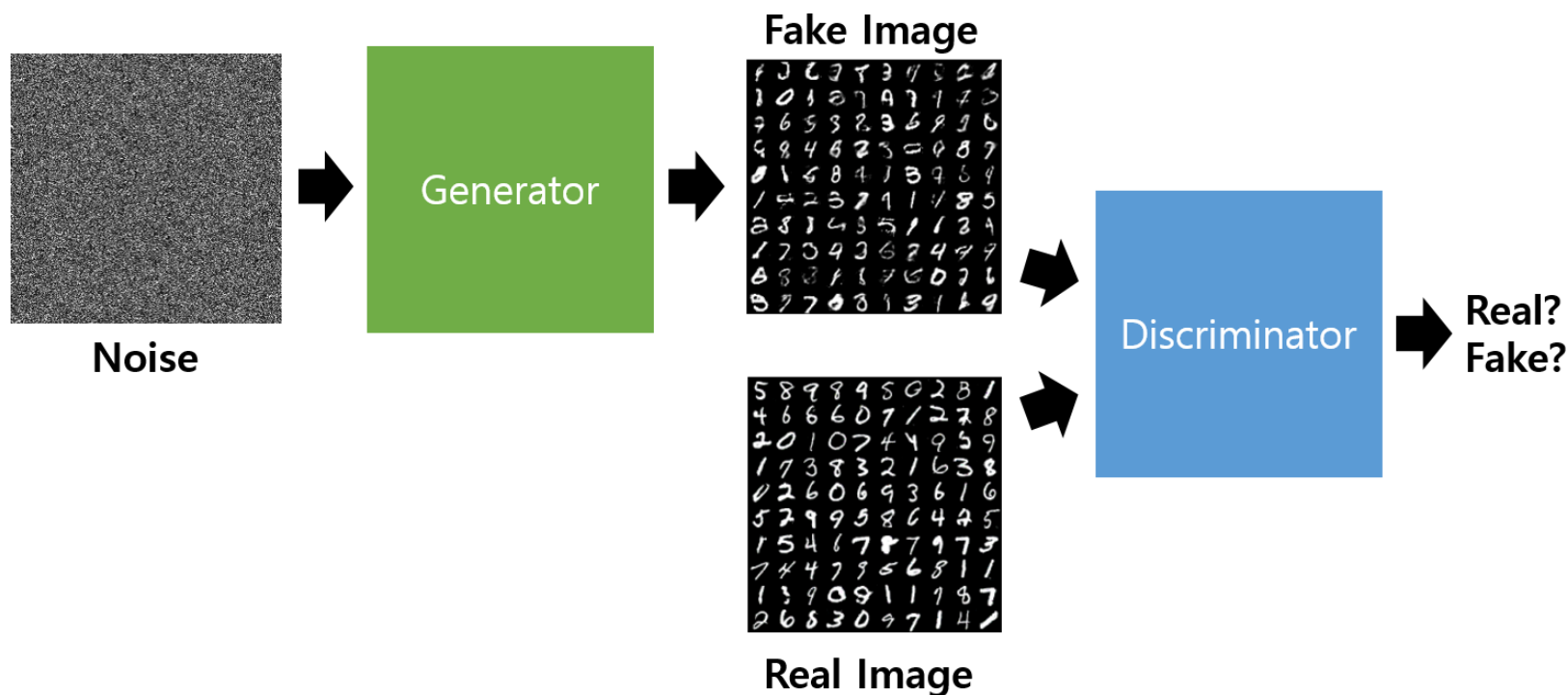


$p_{data}(x)$: 실제 데이터의 분포
 $p_{model}(x)$: 모델이 생성하는 데이터의 분포

$p_{data}(x)$ 와 유사한 $p_{model}(x)$ 를 얻는 것이 생성 모델의 학습 목표!

GAN의 학습 과정

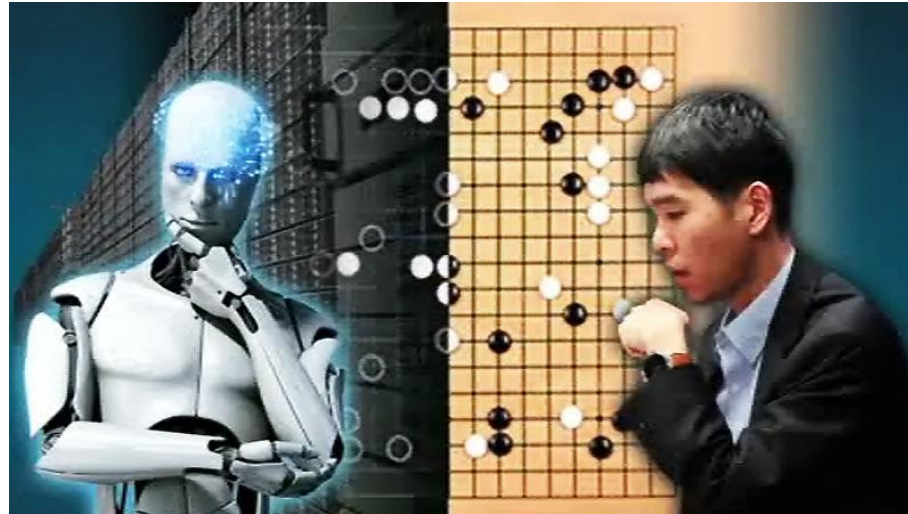
GAN의 학습



GAN의 경우 $p_{\text{model}}(\mathbf{x})$ 를 얻는 과정은 $p_{\text{data}}(\mathbf{x})$ 를 기반으로 이루어지지 않고, 완전히 다른 방식으로 이루어짐!

GAN의 학습 과정

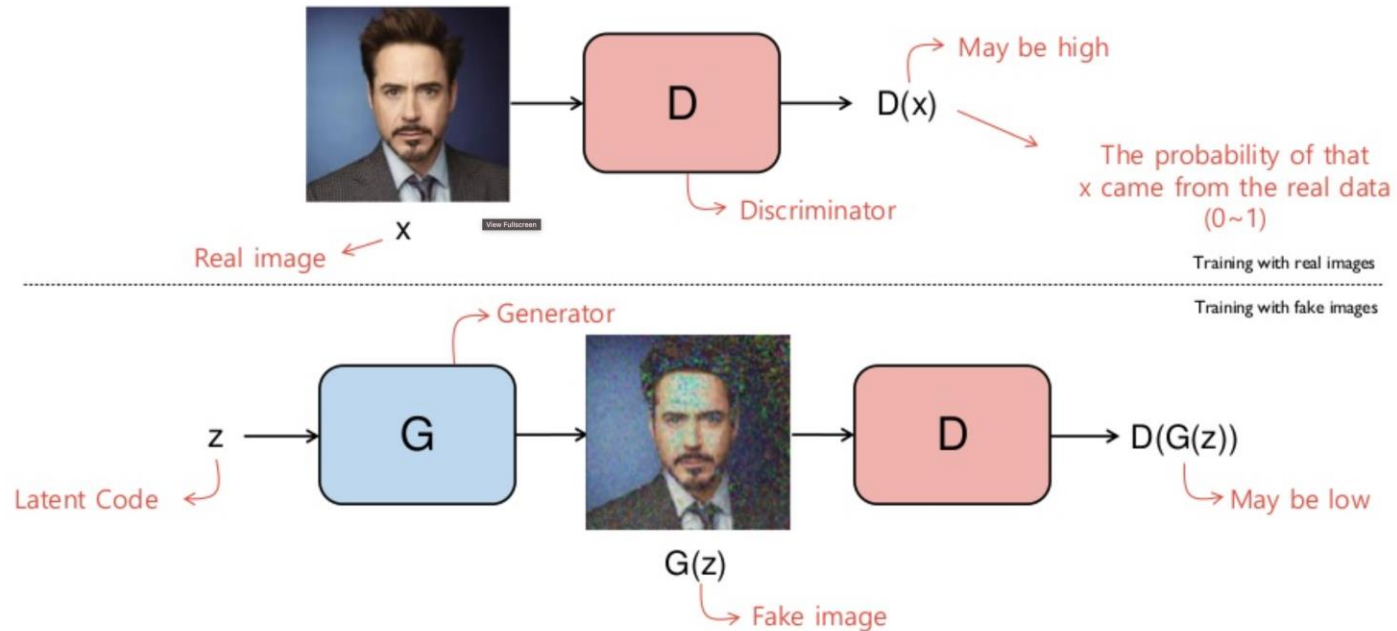
GAN의 학습 - Minimax 알고리즘



서로 경쟁하는 두 대상 사이에서 각자 채택할 수 있는 전략 중 하나로,
자기 자신은 나의 점수를 최대화 (혹은 손실을 최소화) 하는 선택을 하고,
상대는 나의 점수를 최소화 (혹은 손실을 최대화) 하는 선택을 하는 원리의 알고리즘

GAN의 학습 과정

Minimax 알고리즘과 GAN의 학습 원리



판별자의 출력 값은 가짜 데이터에 가까울수록 0에 가까워져야 하고,
진짜 데이터에 가까울수록 1에 가까워져야 함!
생성자의 경우, 판별자의 출력이 1이 되게 하는 가짜 데이터를 생성하는 것이 목표!

GAN의 학습 과정

GAN의 Loss Function

$$\min_{\theta_g} \max_{\theta_d} \left[\mathbb{E}_{x \sim p_{data}} \log D_{\theta_d}(x) + \mathbb{E}_{z \sim p(z)} \log (1 - D_{\theta_d}(G_{\theta_g}(z))) \right]$$

Discriminator outputs likelihood in (0,1) of real image

Discriminator output for real data x Discriminator output for generated fake data $G(z)$

GAN의 loss function은 Minimax 알고리즘에 의존적!

GAN의 학습 과정

GAN의 Loss Function

Discriminator outputs likelihood in (0,1) of real image

$$\min_{\theta_g} \max_{\theta_d} \left[\mathbb{E}_{x \sim p_{data}} \log D_{\theta_d}(x) + \mathbb{E}_{z \sim p(z)} \log (1 - D_{\theta_d}(G_{\theta_g}(z))) \right]$$

Discriminator output Discriminator output

for real data x for generated fake data $G(z)$

생성자 기준	진짜 데이터의 생성에는 관여하지 않으므로 해당 식의 값과 생성자의 성능은 무관함	생성한 가짜 데이터가 진짜 (1) 로 판별되길 원함 따라서, 해당 식의 값이 작아질수록 좋음
판별자 기준	진짜 데이터는 진짜 (1) 로 판별해야 함 따라서, 해당 식의 값이 커질수록 좋음	생성자가 생성한 데이터는 가짜 (0) 로 판별해야 함 따라서, 해당 식의 값이 커질수록 좋음

GAN의 학습 과정

GAN의 Loss Function

$$\min_{\theta_g} \max_{\theta_d} \left[\mathbb{E}_{x \sim p_{data}} \log D_{\theta_d}(x) + \mathbb{E}_{z \sim p(z)} \log (1 - D_{\theta_d}(G_{\theta_g}(z))) \right]$$

Discriminator outputs likelihood in (0,1) of real image

Discriminator output for real data x Discriminator output for generated fake data $G(z)$

생성자는 Loss를 최소화하려는 반면, 판별자는 Loss를 최대화하려고 함!

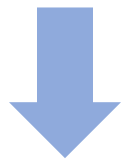
GAN의 학습 과정

Generator Loss와 Discriminator Loss

앞의 Loss Function은 Generator Loss와 Discriminator Loss로 나뉘어질 수 있음

$$\min_{\theta_g} \max_{\theta_d} \left[\mathbb{E}_{x \sim p_{data}} \log \underbrace{D_{\theta_d}(x)}_{\substack{\text{Discriminator output} \\ \text{for real data } x}} + \mathbb{E}_{z \sim p(z)} \log (1 - \underbrace{D_{\theta_d}(G_{\theta_g}(z))}_{\substack{\text{Discriminator output} \\ \text{for generated fake data } G(z)}}) \right]$$

Discriminator outputs likelihood in (0,1) of real image



DISCRIMINATOR LOSS

$$\mathcal{L}_D^{\text{GAN}} = -\mathbb{E}_{x \sim p_d} [\log(D(x))] - \mathbb{E}_{\hat{x} \sim p_g} [\log(1 - D(\hat{x}))]$$

GENERATOR LOSS

$$\mathcal{L}_G^{\text{GAN}} = \mathbb{E}_{\hat{x} \sim p_g} [\log(1 - D(\hat{x}))]$$

GAN의 학습 과정

Generator Loss와 Discriminator Loss

앞의 Loss Function은 Generator Loss와 Discriminator Loss로 나뉘어질 수 있음

$$\min_{\theta_g} \max_{\theta_d} \left[\mathbb{E}_{x \sim p_{data}} \log \underbrace{D_{\theta_d}(x)}_{\substack{\text{Discriminator output} \\ \text{for real data } x}} + \mathbb{E}_{z \sim p(z)} \log (1 - \underbrace{D_{\theta_d}(G_{\theta_g}(z))}_{\substack{\text{Discriminator output} \\ \text{for generated fake data } G(z)}}) \right]$$

Discriminator outputs likelihood in (0,1) of real image



DISCRIMINATOR LOSS

$$\mathcal{L}_D^{\text{GAN}} = -\mathbb{E}_{x \sim p_d} [\log(D(x))] - \mathbb{E}_{\hat{x} \sim p_g} [\log(1 - D(\hat{x}))]$$

GENERATOR LOSS

$$\mathcal{L}_G^{\text{GAN}} = \mathbb{E}_{\hat{x} \sim p_g} [\log(1 - D(\hat{x}))]$$

GAN의 학습 과정

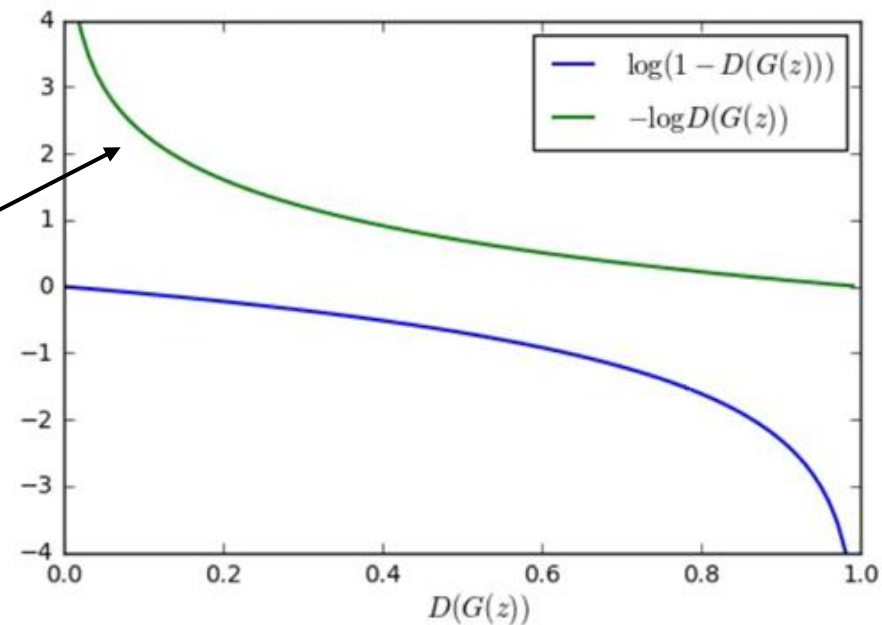
Generator Loss와 Discriminator Loss

Generator Loss의 경우 식의 변형을 통해 학습이 더 잘 이루어지게 할 수 있음

$$\min_{\theta_g} \mathbb{E}_{z \sim p(z)} \log(1 - D_{\theta_d}(G_{\theta_g}(z))) \Rightarrow \max_{\theta_g} \mathbb{E}_{z \sim p(z)} \log(D_{\theta_d}(G_{\theta_g}(z)))$$

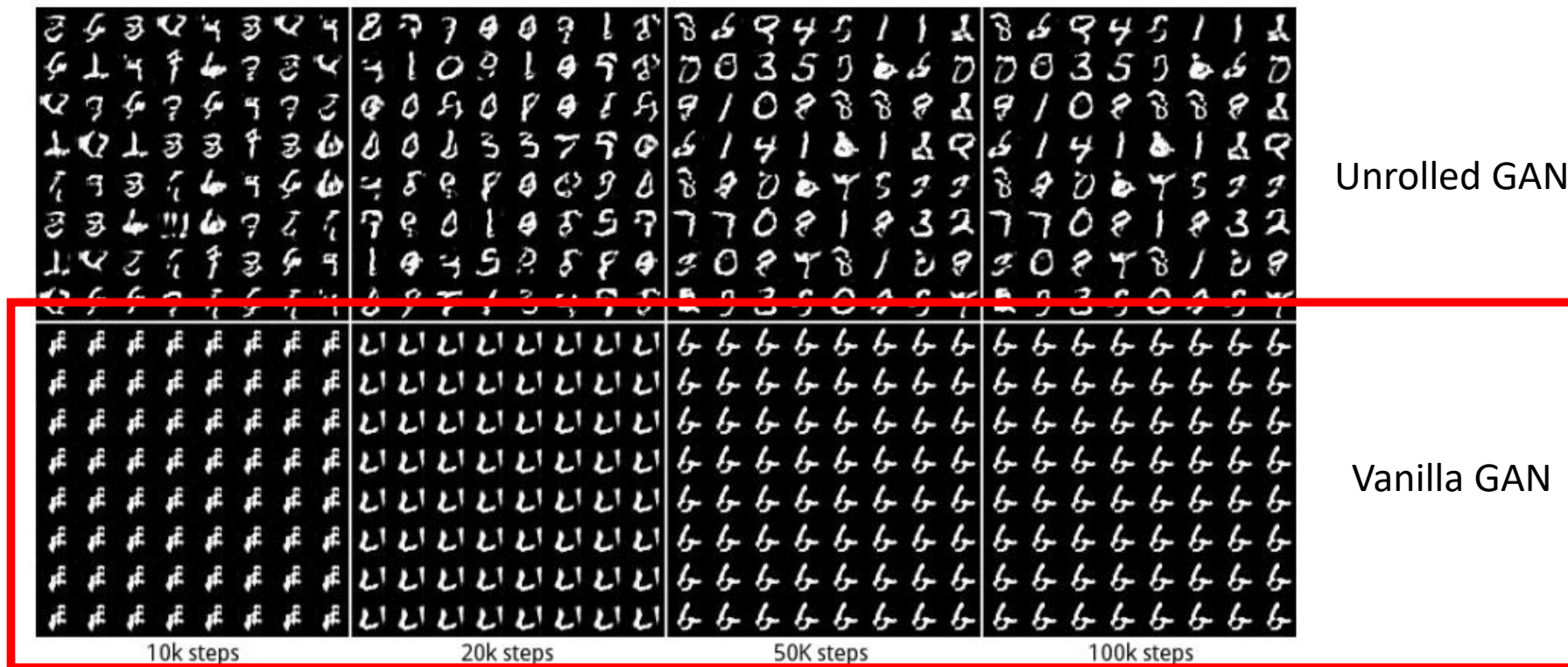
학습 초기에는 생성자의 성능이 나빠
 $D(G(z))$ 의 값이 0에 가까움!

따라서, Generator loss 식의 변형을 통해
초기 기울기 값을 크게 해 줌으로써
생성자의 학습을 더 빠르게 만들 수 있음



GAN의 단점

Mode Collapse



Vanilla GAN의 경우 동일한 데이터가 반복적으로 생성되는 현상이 발생할 수 있는데,

이와 같은 현상을 Mode Collapse라 함

해당 현상은 생성자와 판별자 사이의 경쟁 관계가 무너지고, 한 쪽의 학습만 잘 이루어지는 경우 발생

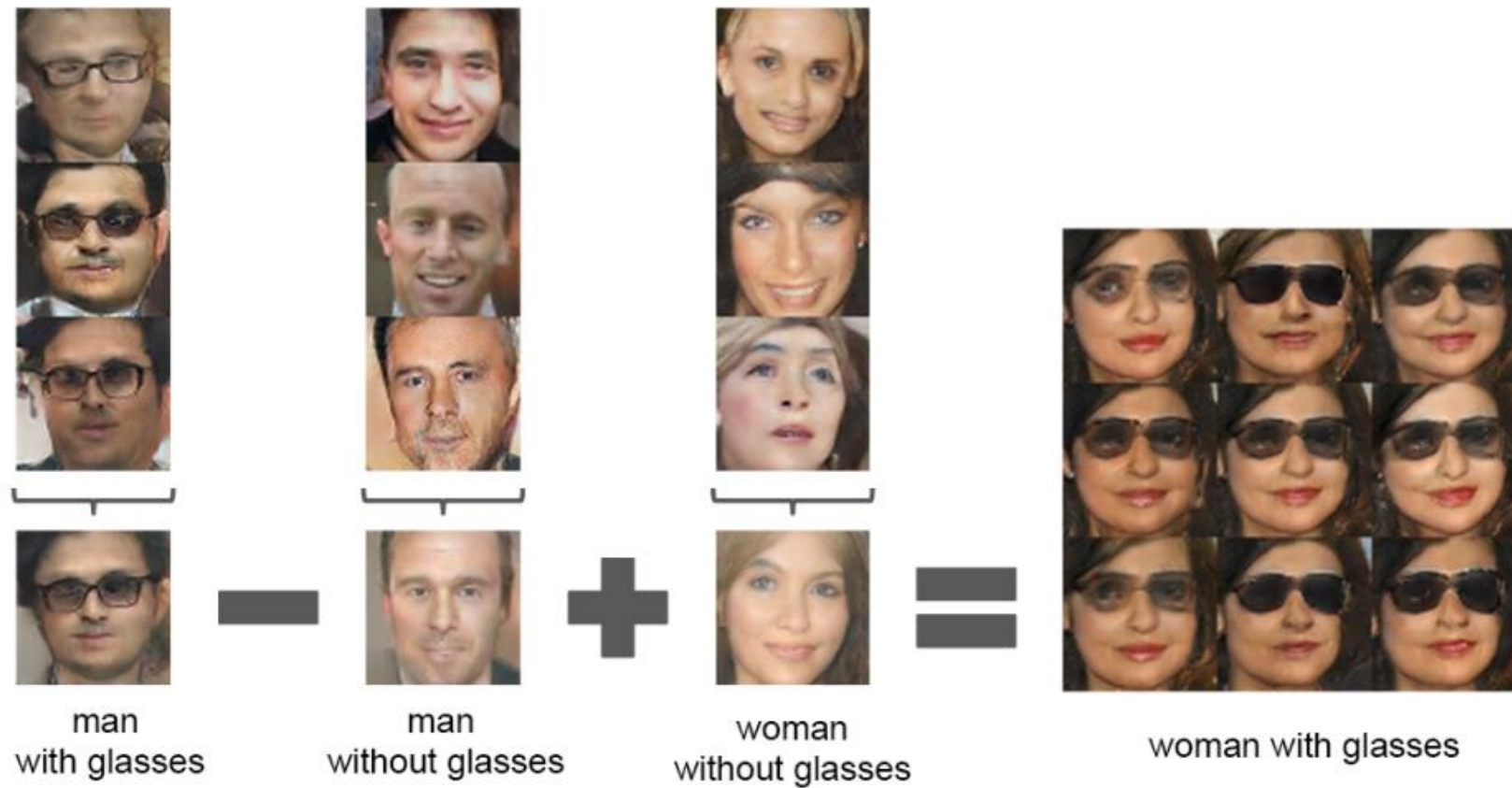
GAN의 응용

이미지 모방 (Practice 1)



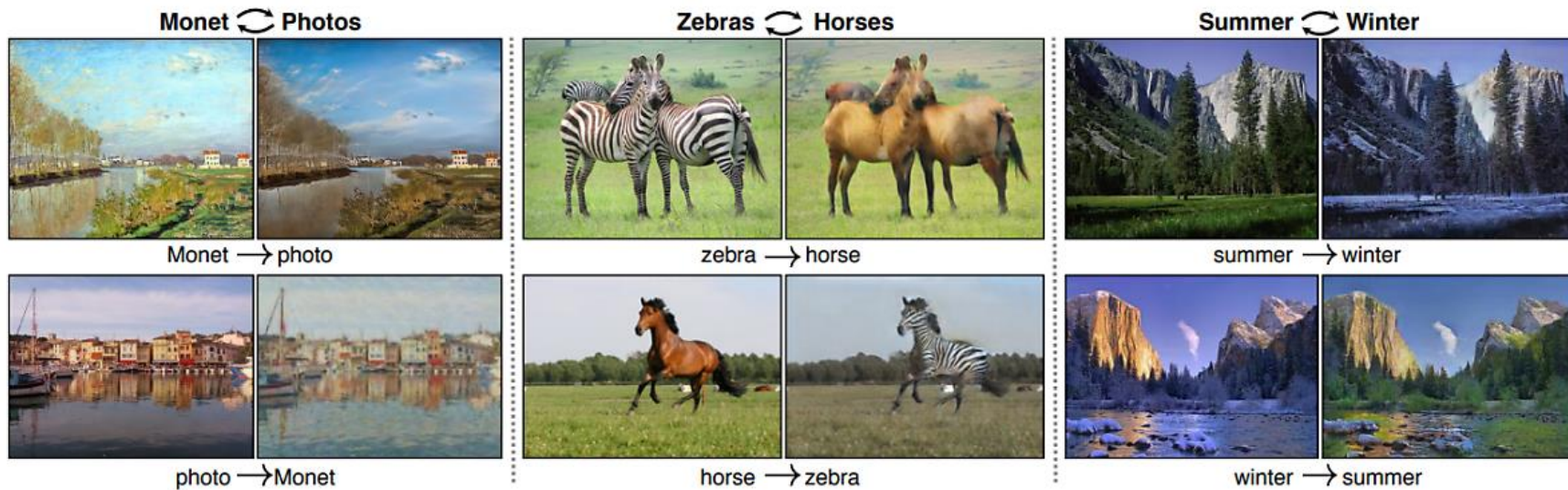
GAN의 응용

이미지 합성



GAN의 응용

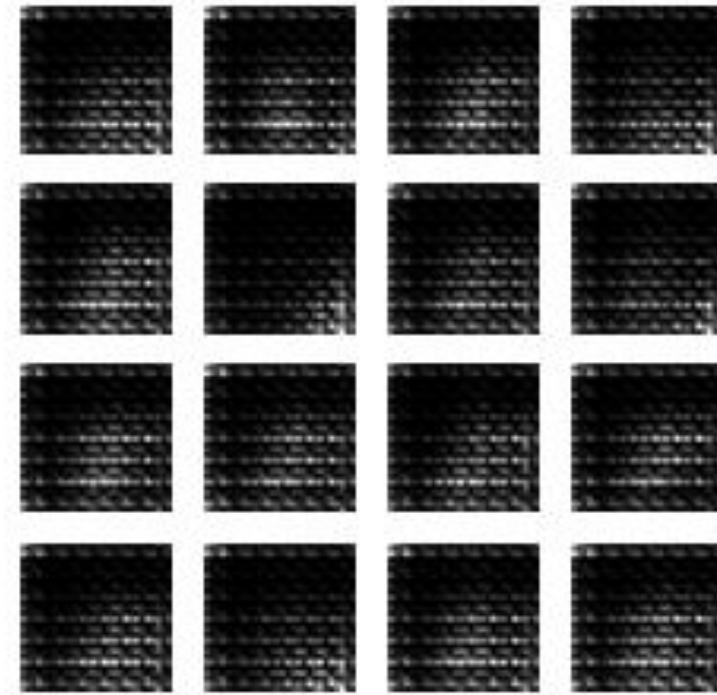
이미지 변환 (Practice 2)



여기서는 CycleGAN이라는 특수한 구조의 GAN이 활용됨!

Practice 1

MNIST 데이터셋의 숫자 손 글씨들을 모방하는 GAN 모델의 구현



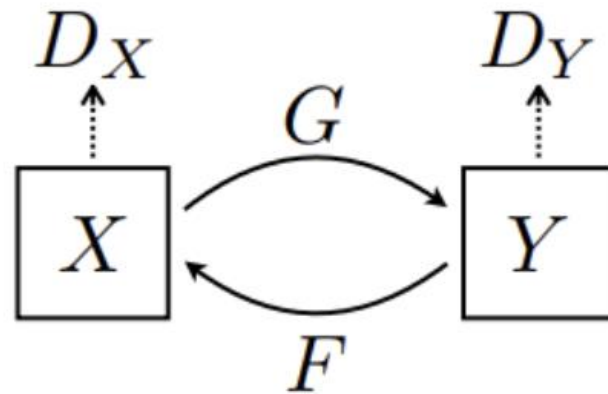
Part 2

CycleGAN

CycleGAN

GAN에서 파생된 구조

2개의 GAN이 한 쌍으로 결합되어 형성된 구조로, 생성자와 판별자 모두 2개씩 존재

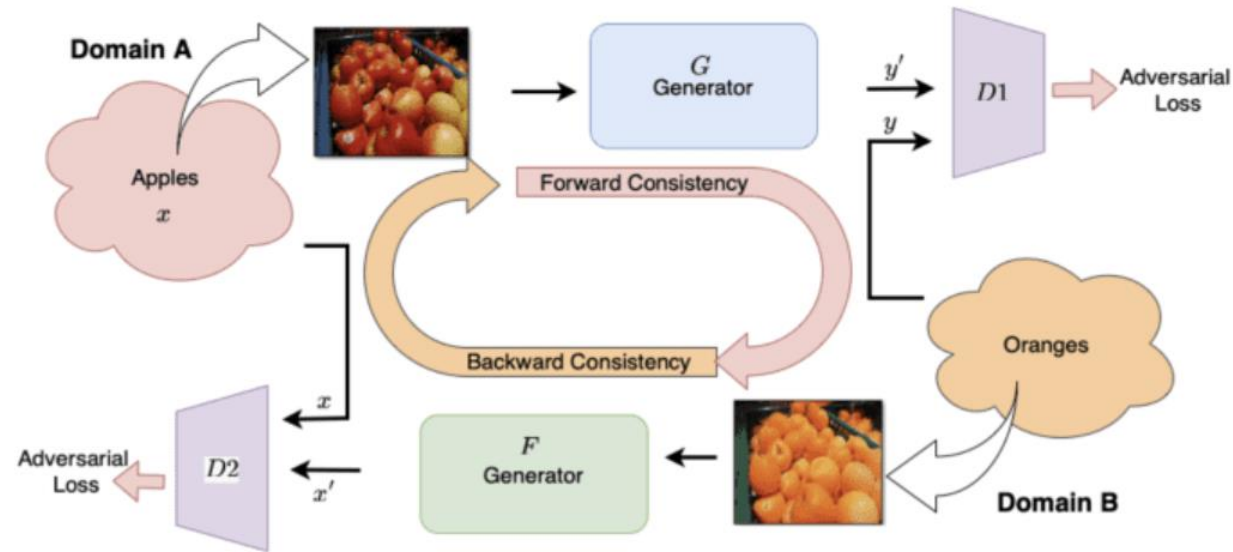


X, Y : 두 종류의 데이터
 F, G : 생성자
 D_X, D_Y : 판별자

CycleGAN

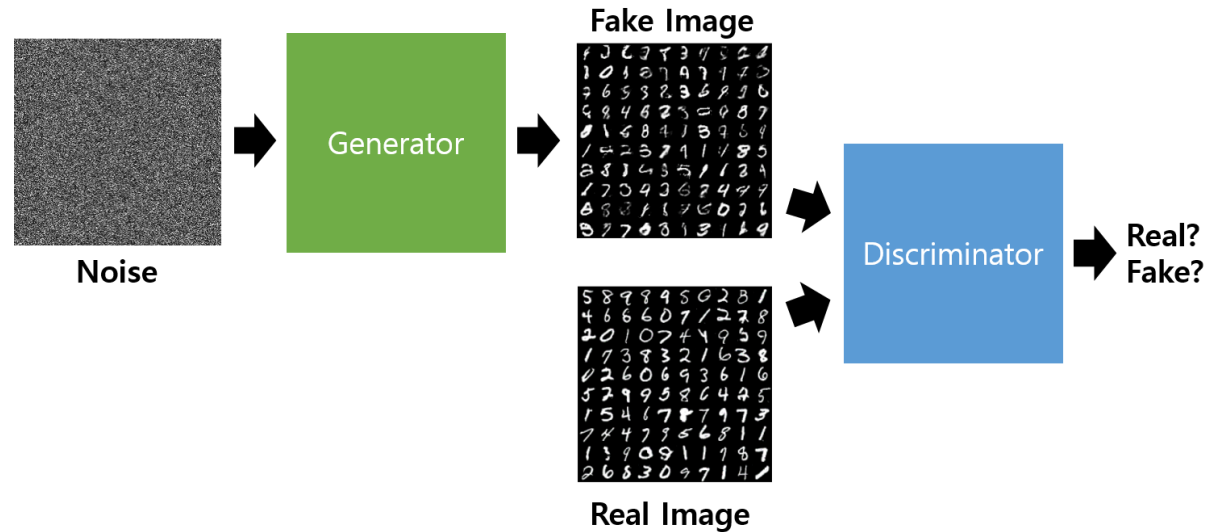
GAN에서 파생된 구조

서로 다른 두 종류의 데이터의 양 방향 변환이 가능!



CycleGAN의 학습 과정

GAN의 학습 과정 및 Loss Function



Discriminator outputs likelihood in (0,1) of real image

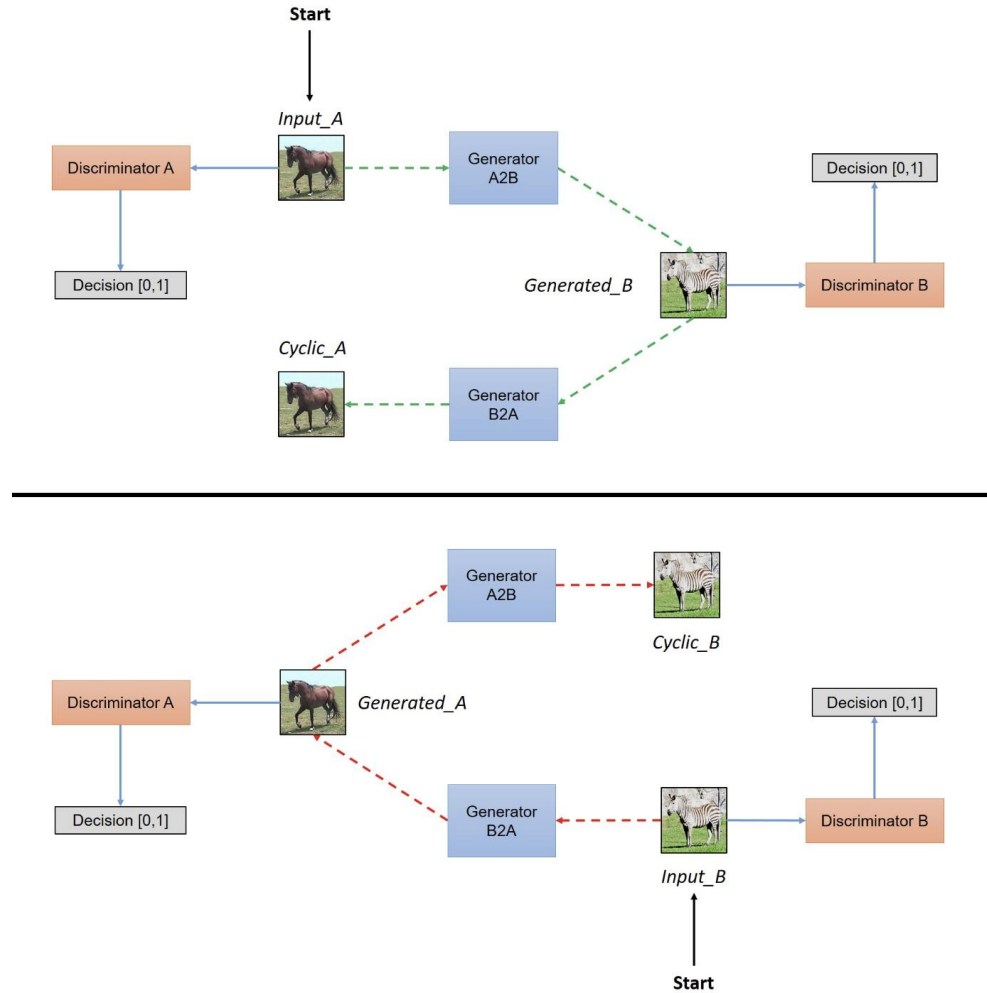
$$\min_{\theta_g} \max_{\theta_d} \left[\mathbb{E}_{x \sim p_{data}} \log D_{\theta_d}(x) + \mathbb{E}_{z \sim p(z)} \log (1 - D_{\theta_d}(G_{\theta_g}(z))) \right]$$

Discriminator output for real data x

Discriminator output for generated fake data $G(z)$

CycleGAN의 학습 과정

CycleGAN의 학습 과정



CycleGAN의 학습 과정

CycleGAN의 Loss Function

총 4개의 부분으로 이루어져 있음!

- Generator loss: 생성자 F 와 G 각각에서 고려
- Discriminator loss: 판별자 D_X 와 D_Y 각각에서 고려
- Cycle loss: 두 가지 가능한 경우 모두 고려

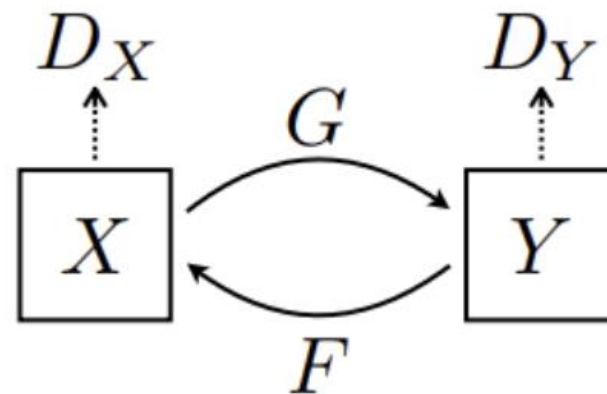
X 자신과 $X \rightarrow Y \rightarrow X$,

Y 자신과 $Y \rightarrow X \rightarrow Y$

- Identity loss: 두 가지 가능한 경우 모두 고려

X 자신과 X 를 생성자 F 에 input한 것,

Y 자신과 Y 를 생성자 G 에 input한 것



Practice 2

말 이미지와 얼룩말 이미지를 서로 변환하는 CycleGAN 모델의 구현

