Generative Adversarial Network (GAN)

ML Lab

Machine Learning Laboratory
POSTECH

Contents

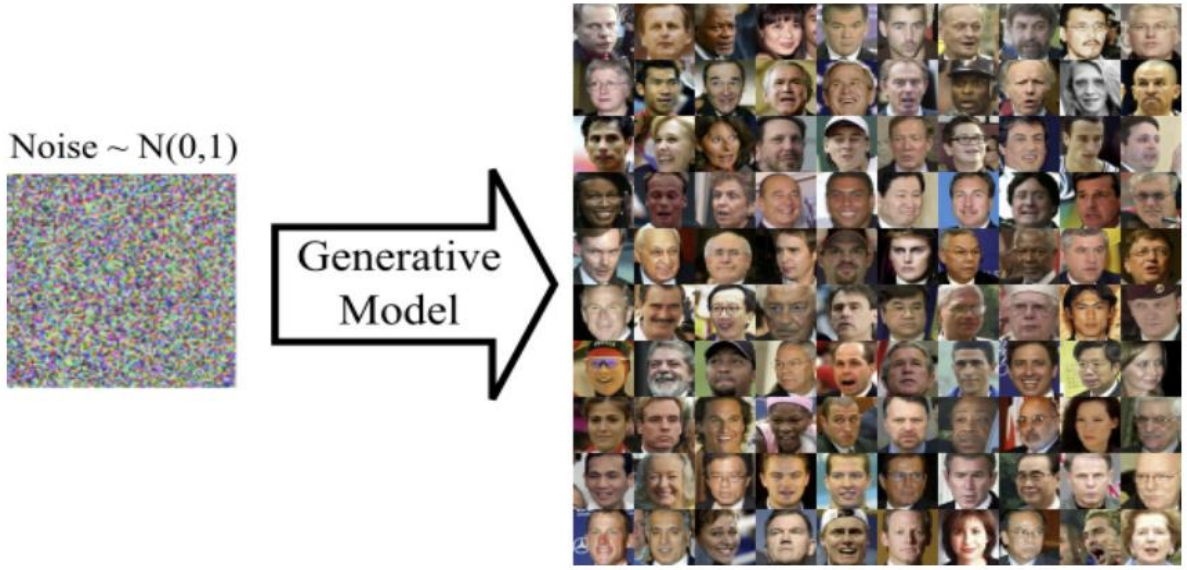
- Generative Adversarial Network (GAN)
  - 생성 모델 (Generative Model)
  - GAN의 학습 과정
  - GAN의 단점
  - GAN의 응용
  - Practice 1

기계학습 알고리즘의 분류

- 학습 방식에 따라 크게 2가지로 분류됨!
- 지도학습 (Supervised Learning)
  - 라벨이 주어진 데이터로 모델을 학습시키는 방식
  - 입력 데이터 x 에 대한 출력 값 y를 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), …

생성 모델 (Generative Model)

- 비지도학습 (Unsupervised Learning)의 응용 방향 중 하나!
  - 학습에 사용된 데이터들과 유사한 새로운 데이터를 만드는 모델
  - 학습에 사용된 데이터의 분포 $p_{data}(x)$ 의 학습 후, 이를 기반으로 새로운 데이터 x' 생성



Generative Adversarial Network (GAN)

- GAN의 의미

Generative Adversarial Network

생성적 적대적 신경망

Generative Adversarial Network (GAN)

- GAN의 의미

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
  - 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

생성적 적대적 신경망

Generative Adversarial Network (GAN)

- 생성자와 판별자 사이의 경쟁이란?
  - 생성자: 판별자가 진짜 데이터라고 잘못 판별하게 만들 가짜 데이터 생성
  - 판별자: 진짜 데이터와 가짜 데이터를 성공적으로 구분

Generative Adversarial Network

생성적 적대적 신경망

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

Generative Adversarial Network (GAN)

- GAN의 비유 - 위조 지폐 제작 과정

- 학습이 충분히 이루어지기 전의 어설픈 위조 지폐 → 위조품임이 쉽게 판별됨!

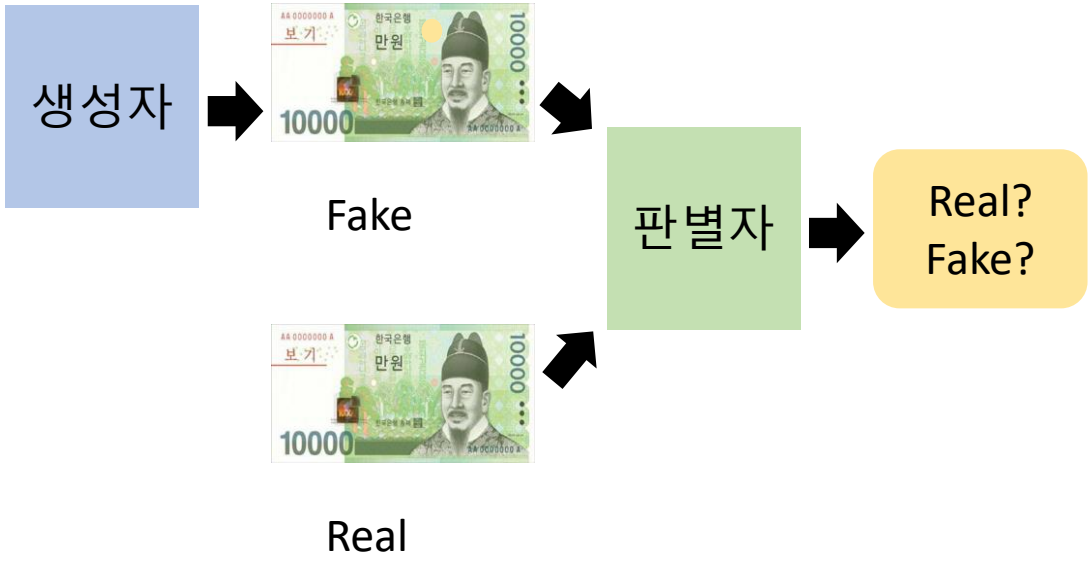
Generative Adversarial Network (GAN)

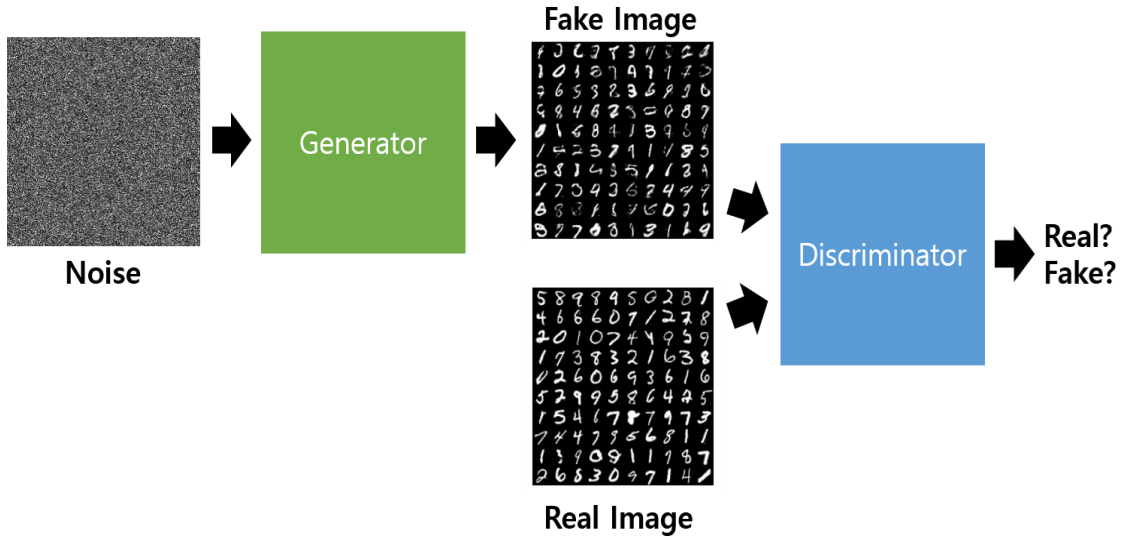
- GAN의 비유 - 위조 지폐 제작 과정

- 학습이 충분히 이루어진 뒤의 진짜 지폐와 구분이 어려운 위조 지폐 → 진품으로 잘못 판별될 여지가 생김!

Generative Adversarial Network (GAN)

- GAN의 전체 구조

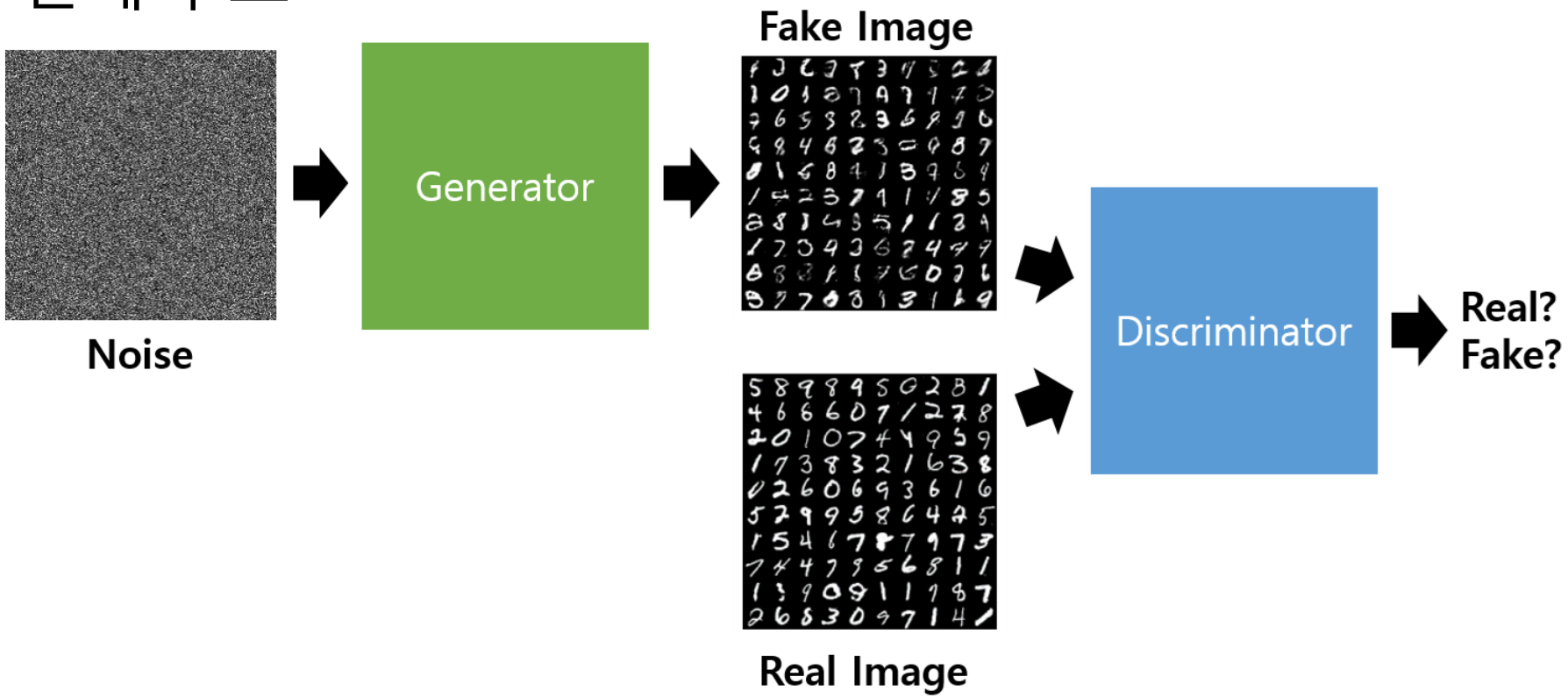




- 진짜 지폐로부터 일부분을 변형하는 과정이, Noise에서 무작위로 값을 추출하는 과정에 대응됨!

Generative Adversarial Network (GAN)

- GAN의 전체 구조



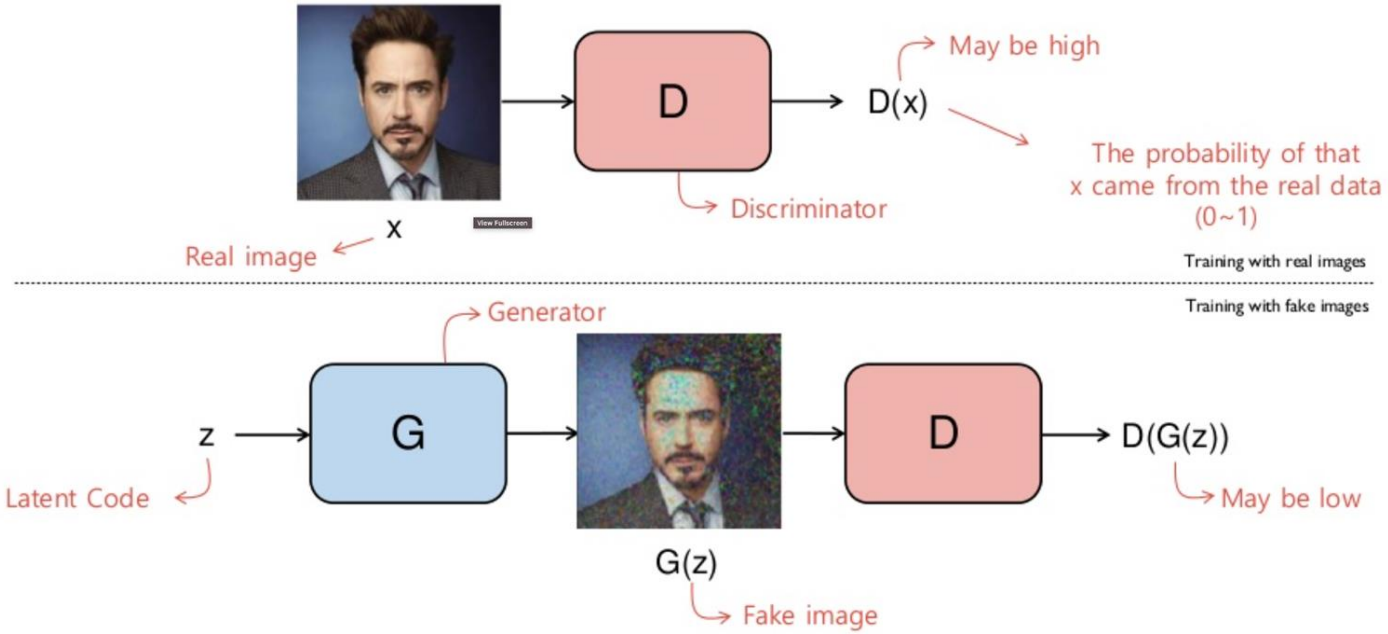
  - 생성자와 판별자 사이의 경쟁을 통해, 생성자는 실제 데이터와 유사한 데이터를 생성할 수 있게 됨!

GAN의 학습 - Minimax 알고리즘

- 서로 **경쟁하는 두 대상** 사이에서 각자 채택할 수 있는 전략 중 하나
  - 자기 자신(생성자)은 나의 점수를 최대화하는 선택 → 판별자가 위조품을 진품으로 판별하도록
  - 상대(판별자)는 나의 점수를 최소화하는 선택을 하는 원리의 알고리즘 → 판별자가 잘 판별하도록

GAN의 학습 - Minimax 알고리즘

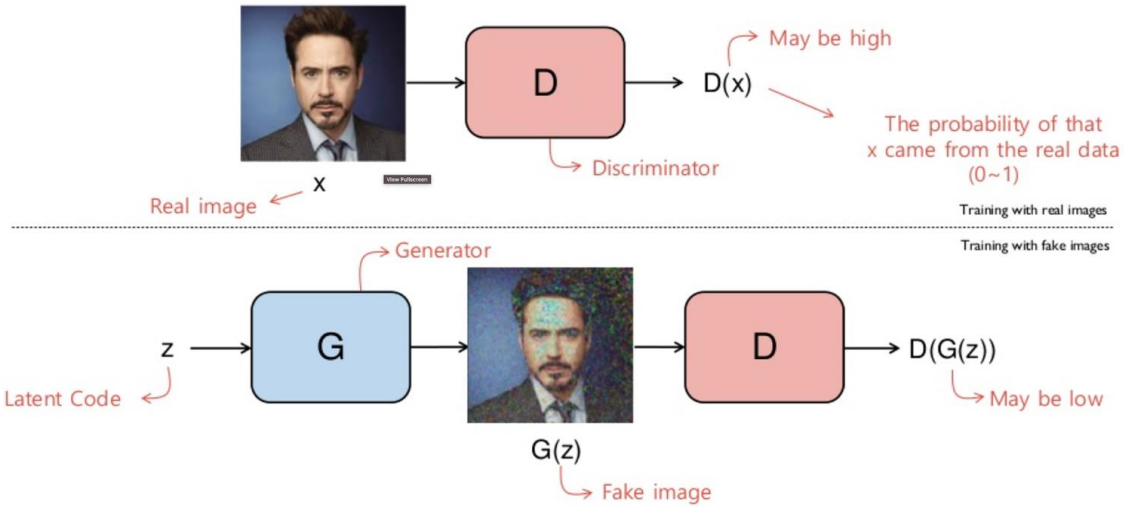
- Minimax 알고리즘과 GAN의 학습 원리
  - 판별자의 출력 값은 가짜 데이터에 가까울수록 0에 가까워져야 하고, 진짜 데이터에 가까울수록 1에 가까워져야 함(보통의 classifier)
  - 생성자의 경우, 판별자의 출력이 1이 되게 하는 가짜 데이터를 생성하는 것이 목표!



https://wikidocs.net/22893

GAN의 Loss Function

- GAN의 loss function은 Minimax 알고리즘에 의존적!



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

Real data　　Fake data

https://wikidocs.net/22893

GAN의 Loss Function

- 생성자는 Loss를 최소화하려는 반면, 판별자는 Loss를 최대화하려고 함!

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

Real data

Fake data

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

https://wikidocs.net/22893

GAN의 Loss Function

- Generator Loss & Discriminator Loss

$$\min_{\theta_g}\max_{\theta_d}\left[\mathbb{E}_{\boldsymbol{x}\sim\boldsymbol{p}_{\text{data}}}\log \boldsymbol{D}_{\theta_d}(\boldsymbol{x}) + \mathbb{E}_{\boldsymbol{z}\sim\boldsymbol{p}(\mathbf{z})}\log\left(1-\boldsymbol{D}_{\theta_d}\left(\boldsymbol{G}_{\boldsymbol{\theta}_g}(\mathbf{z})\right)\right)\right]$$

Real data

Fake data

  - Generator Loss

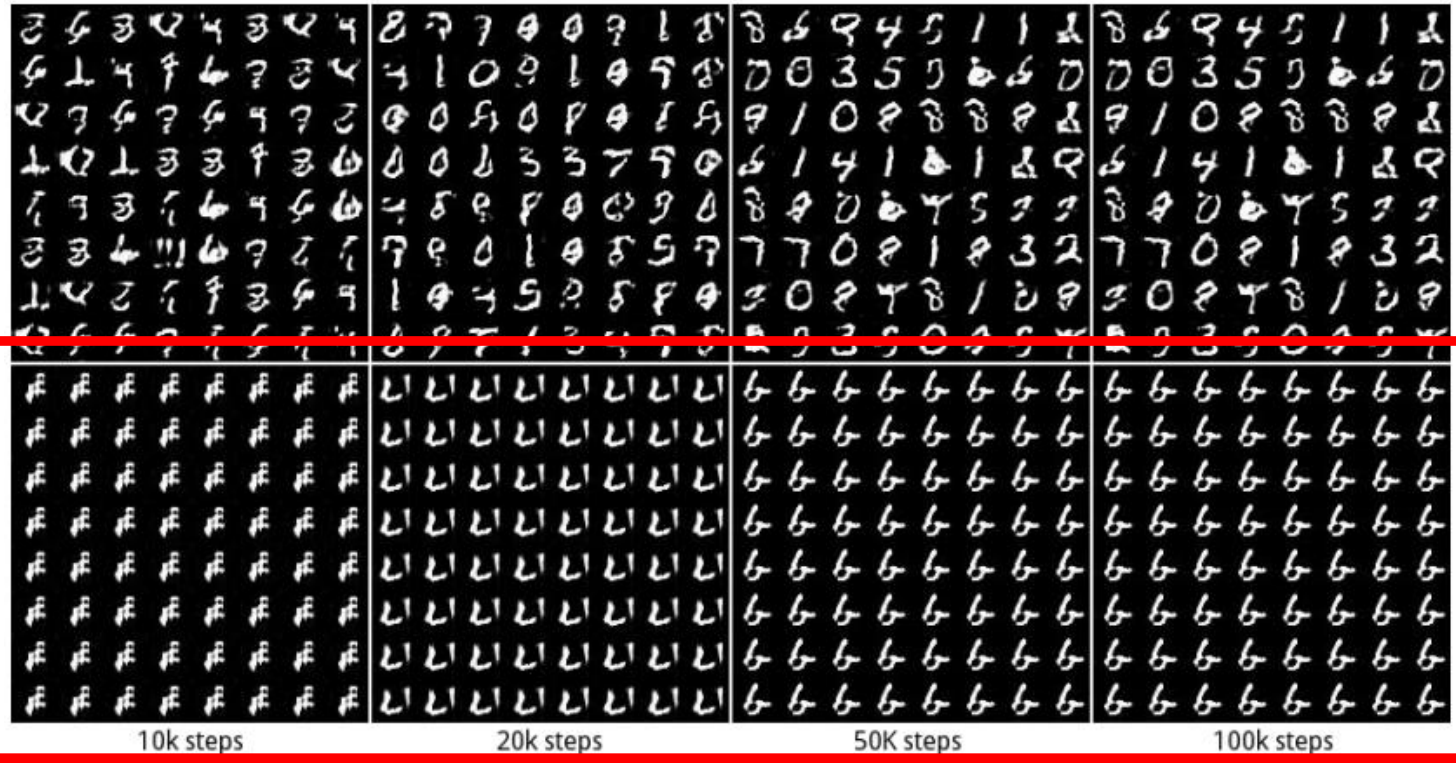
$$\mathcal{L}_{\mathrm{G}}^{\mathrm{GAN}} = \mathbb{E}_{\hat{x}\sim p_g}[\log(1-D(\hat{x}))]$$

  - Discriminator Loss

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

GAN의 단점

- Mode Collapse
  - 동일한 데이터가 반복적으로 생성되는 현상
  - 생성자와 판별자 사이의 경쟁 관계가 무너지고, 한 쪽의 학습만 잘 이루어지는 경우 발생



GAN의 응용

- 이미지 모방 (Practice)

Practice

- MNIST 데이터셋의 숫자 손 글씨들을 모방하는 GAN 모델의 구현