

Conditional GAN(cGAN)

논문 : “Conditional Generative Adversarial Nets”, Mehdi Mirza, Simon Osindero, 2014.

Generative Adversarial Networks (GAN)

- GAN 논문
- Generator(G)와 Discriminator(D)의 상호 견제를 통하여, 기존의 생성 모델과 다르게 복잡한 확률 계산 없이 훨씬 좋은 성능을 보임
- GAN을 사용하면 학습에 사용한 이미지와 비슷한 가짜 이미지를 생성 가능
- original GAN의 한계 : 원하는 이미지를 생성하도록 제어하는 것은 불가능
- 2014년 몬트리올대의 “Mehdi Mirza”팀이 이러한 한계를 해결한 conditional GAN(cGAN) 발표
- cGAN 발표 후, 이를 바탕으로 많은 연구 발표 : 대표적인 예는 Pix2Pix, CycleGAN

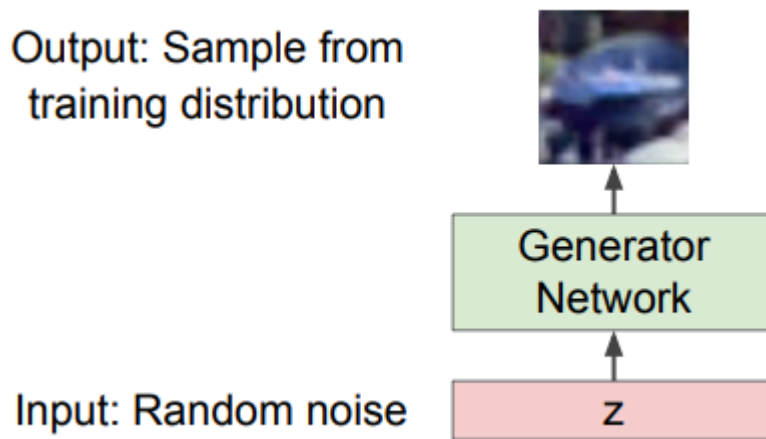
GAN의 작동 원리

- 편의 상, generator는 G , discriminator는 D 로 표기하겠음
- 2-player의 minimax 게임을 통한 학습
- 학습 과정에서 상호 협조적이지 않은 상대(G, D)가 서로 최적의 상태(Nash Equilibrium)에 도달하려고 노력하는 방식
- 위의 목표를 달성하기 위한 objective function

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

- \mathbb{E} : expectation (기댓값)
- $x \sim p_{data}(x)$: x is sampled from real data (ground truth data)
- $\log D(x)$: pobability of D (real)
- $z \sim p_z(z)$: z is sampled from $N(0, 1)$
- $\log(1 - D(G(z)))$: pobability of D (fake), $G(z)$ is fake data

Latent Variable

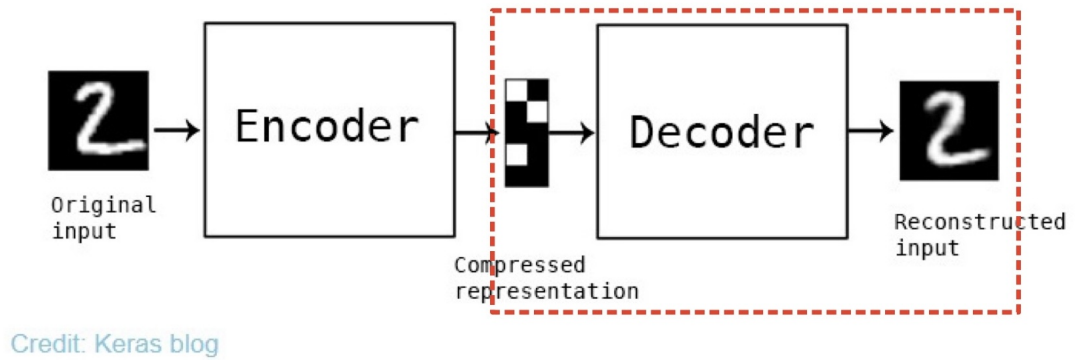


(cs231n-Lecture-13-p.103)

- G 의 input = random noise or latent variable, z
- G 의 output = 학습 영상의 분포를 반영한 실제와 비슷한 가짜 영상
- latent variable (내재 변수, 잠재 변수) = 직접적으로 관측되는 변수가 아닌, 관측이 가능한 다른 변수들로부터 추론이 가능한 변수 (wikipedia 설명 참고)
 - 예컨대, ([라온피플 블로그 참고](#)) :

“건강”이라는 추상적인 변수는 직접적으로 측정할 수 없는 “내재 변수”이지만, 몸이나 정신 상태를 표현하는 말로 자주 사용된다. 직접 계측할 방법은 없지만 혈압, 맥박, 혈당, 체온, 체중, 허리둘레 등 “(많은) 관측 가능한 외부 변수”로부터 추론 가능하며, 이런 외부 변수로부터 추상적인 건강이라는 내재 변수를 도출하려면 수학적 또는 경험적 모델이 필요하다.
 - 내재 변수의 사용 이유 :

위의 예에서 들었던 각각의 변수를 따로 떼어내 해석하면 몸 상태를 파악하기 어렵지만, 이것들을 건강이라는 내재 변수(추상적 개념)로 엮으면 해석이 쉬워지기 때문
- 내재 변수 : auto-encoder에서 입력 데이터를 받은 후, encoding 과정을 거쳐 만들어진 압축된 결과(compressed representation)
 - (예) 채널 3인 픽셀들의 배열로 이루어진 이미지의 경우, 인코더를 통해 압축한 결과인 z 를 얻고, 이 z 를 다시 디코더에 넣고 통과시켜 복원된 데이터를 얻음



- 위 그림의 디코더 부분은 G 의 과정과 매우 유사
- GAN의 input vector z 에 대한 서로 다른 표현 :
 - noise : 사전에 정의할 수 없음을 의미하는 표현
 - latent variable : auto-encoder처럼 이 벡터로부터 출력 영상을 만들 수 있다는 의미를 나타내는 표현
- conditional GAN의 아이디어
 - 내재 변수에 이미지 생성과 관련된 조건(condition)을 추가하고, 이를 통해 이미지 생성을 제어할 수 있다면, 원하는 이미지를 생성할 수 있고 다양한 분야에도 응용이 가능할 것이다!
 - 실제로 cGAN은 다양한 시도를 해볼 수 있는 단초가 됨

Conditional GAN (cGAN)

- original GAN에 조건을 추가하는 방법 : G 와 D 에 특정 조건을 나타내는 정보인 y 를 input
- 조건 y 의 형태 : 특별히 정해진 것이 아닌 다양한 형태 가능 (아래의 예 참고)
 - 숫자 필기체를 인식하는 MNIST에서 원하는 숫자를 생성하고 싶다면, 숫자의 class에 해당하는 label(논문에서는 one-hot encoding 사용)을 추가로 input
 - 만약 MNIST를 one-hot encoding 하면 10-bit가 필요하기 때문에, y 는 10비트가 됨
 - 이런 class의 라벨 뿐 아니라, 다양한 형태(multi-modal)도 가능
 - 아래의 그림처럼 G 와 D 의 기존 latent variable에 y 를 더해주면 됨 (original GAN과 이것만 다름)

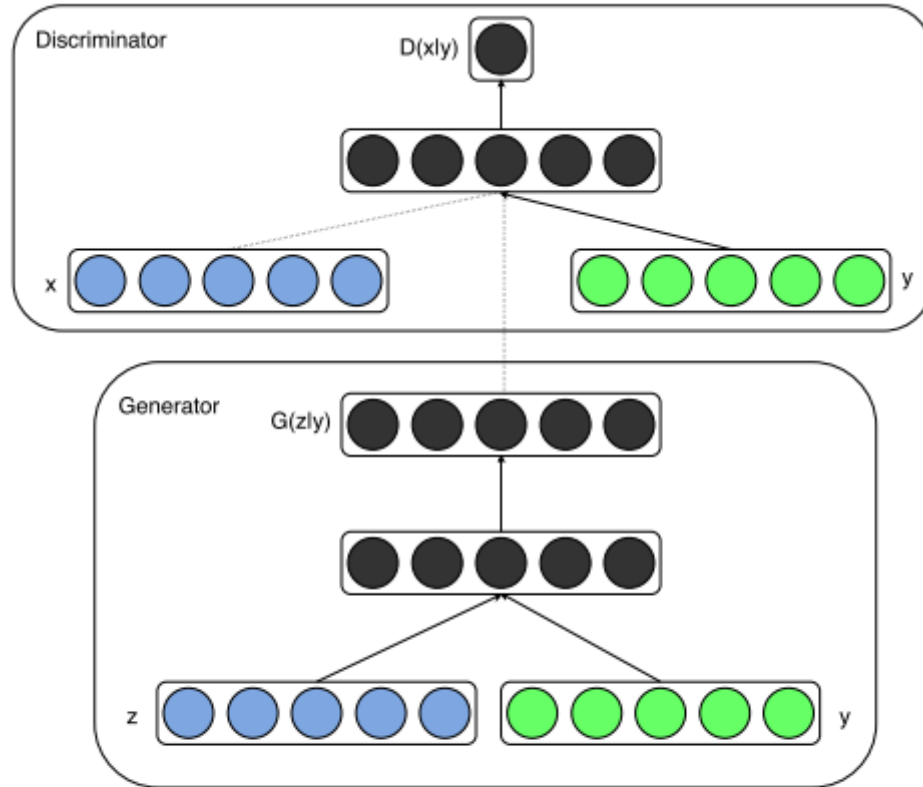


Figure 1: Conditional adversarial net

- cGAN을 풀기 위한 objective function (original GAN에 조건 y 만 추가된 형태)

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

cGAN의 결과

- (논문에서는) MNIST 데이터 사용
- 조건 y 를 one-hot encoding 시킨 class label 사용하여, 아래의 그림처럼 추가한 조건 y 에 맞게 원하는 숫자를 생성하는 것을 확인



Figure 2: Generated MNIST digits, each row conditioned on one label

이후 연구 결과들

1. pix2pix
2. cycle GAN