

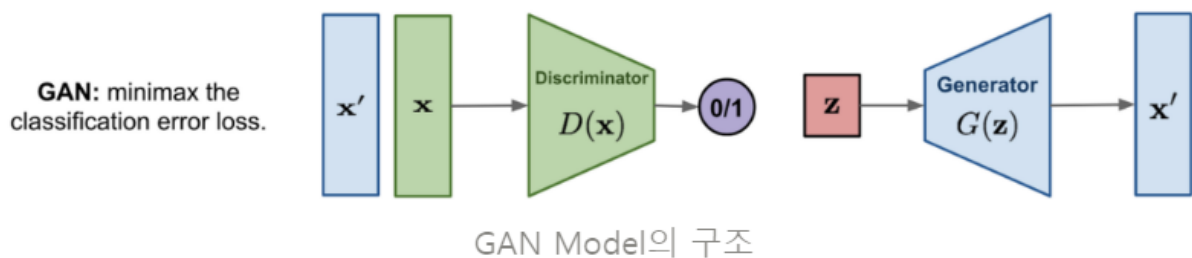
대표적인 이미지 생성 모델들

– GAN, VAE, Flow-base model, Diffusion 각각의 개념, 구조, 장단점, 차이점

[Generative Models]

1. GAN

생성자와 판별자의 minmax(adversarial) training을 통해, realistic한 이미지를 생성하는 생성자를 학습한다. 생성자와 판별자의 경쟁이 과도하게 unbalanced 될 때 mode collapse와 같은 문제를 겪을 수 있으며, 훈련이 불안정할 수 있다.

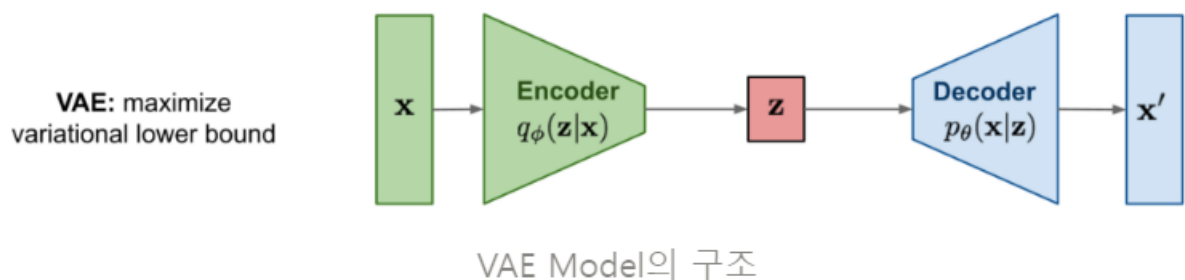


장점: 실제 사진과 유사한 일관성 있는 이미지를 생성할 수 있고, 모델 구조의 선택이 자유롭다.

단점: model의 coverage가 좁아서 사실적인 이미지를 생성하긴 하지만 다양한 이미지를 생성하지는 못하고, 훈련 과정에서의 불안정성이 존재한다.

2. VAE

Autoencoder의 구조에서 ELBO와 latent vector의 output이 Gaussian임을 가정하여 진행한다.

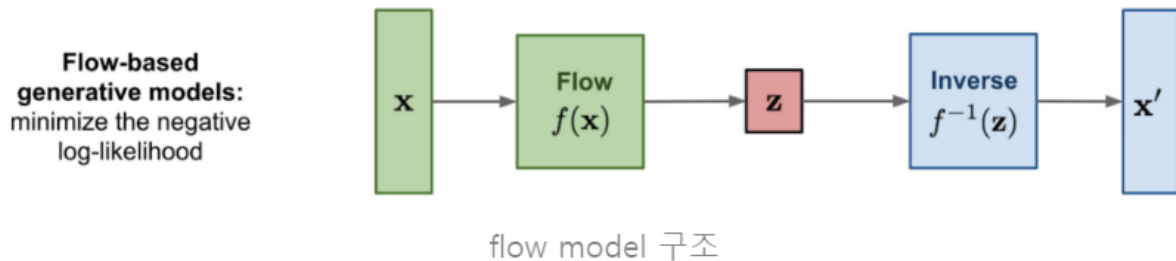


장점: 확률 모델을 기반으로 했기 때문에, 잠재 코드를 더 유연하게 계산할 수 있다.

단점: Density를 직접적으로 구한 것이 아니기 때문에 Pixel RNN/CNN과 같이 직접적으로 Density를 구한 모델보다는 성능이 떨어진다.

3. Flow-based model

Flow model은 딥러닝 모델에 대한 역변환을 통해 단순한 분포에서 시작하여 복잡한 확률 분포를 모델링하는 방법이다. 확률론에서 자주 언급되는 change of variable 이론에 기반한 방법이며, log likelihood objective로 학습이 이루어진다.

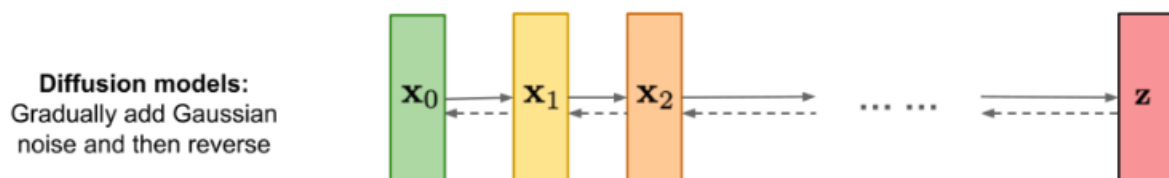


장점: 데이터의 확률 분포를 모델링하기 위해 사용되기에 확률적인 샘플링, 확률 밀도 추정 및 생성 작업에 유용하다.

단점: 훈련과 추론 과정에서 계산 비용이 상대적으로 높을 수 있고, 이로 인해 대규모 데이터셋에서의 훈련 및 추론이 느릴 수 있다.

4. Diffusion

Diffusion model은 데이터를 만들어내는 deep generative model 중 하나로, data로부터 noise를 조금씩 더해가면서 data를 완전한 noise로 만드는 forward process와 이와 반대로 noise로부터 조금씩 보원해가면서 data를 만들어내는 reverse process를 활용한다. 즉, noise를 확산시킨 다음 noise를 제거하는 방식으로 샘플링을 진행한다.



장점: Distribution coverage가 높아서 다양한 이미지 생성이 가능하다.

단점: 순차적인 inverse process를 통해 이미지가 생성되므로 생성 속도가 비교적 느리다.

[Differences]

- VAE는 image를 encoding하는 network와 latent code를 바탕으로 image를 decoding하는 network 모두를 학습하는 반면, Diffusion model은 이미지를 encoding하는 forward process는 fix 된 채 image를 decoding하는 reverse process single network만을 학습한다.

- VAE는 하나의 latent variable을 활용하는 반면 Diffusion 모델은 sequential한 여러 개의 latent variable을 활용한다.

- VAE는 이미지 생성에 대한 확률적 접근을 사용하므로 생성된 이미지의 품질이 GAN보다 낮다.
- VAE는 판별자가 없기 때문에 생성된 이미지를 직접적으로 평가하거나 개선하기 어려운 반면, GAN은 판별자를 통해 생성된 이미지의 질을 지속적으로 향상시킬 수 있다.
- VAE & GAN은 매우 복잡한 딥러닝 신경망 아키텍처를 사용하여 데이터의 고차원 구조를 학습하기에 이미지 다양성을 높일 수 있지만, Flow-base model은 변환 함수의 제약으로 인해 모델의 복잡성이 제한되어서 이미지 다양성 측면에서 한계가 있을 수 있다.
- Diffusion 모델은 안정적인 훈련 과정과 확률적인 요소를 활용하여 GAN의 mode collapse 문제를 피할 수 있다.
- Diffusion 모델은 GAN과는 달리 stationary training objective를 사용한다.
- Diffusion 모델은 상대적으로 노이즈가 많을 수 있는 샘플링 과정, 세부 정보를 재구성하기 어려울 수 있는 작업 특성 등으로 인해 GAN에 비해 fidelity가 낮다.
- Flow-base model은 훈련이 비교적 안정적이며 수렴하기 쉬운 경향이 있어서 GAN과 같은 경쟁적인 모델보다 훈련에 대한 제어가 더 쉽다.

[References]

- <https://jang-inspiration.com/generative-model#eb2149ae4042d09f657f7a15159cf4>
- <https://velog.io/@reversesky/DDPM-Understanding-Diffusion-models-A-Unified-Perspective>