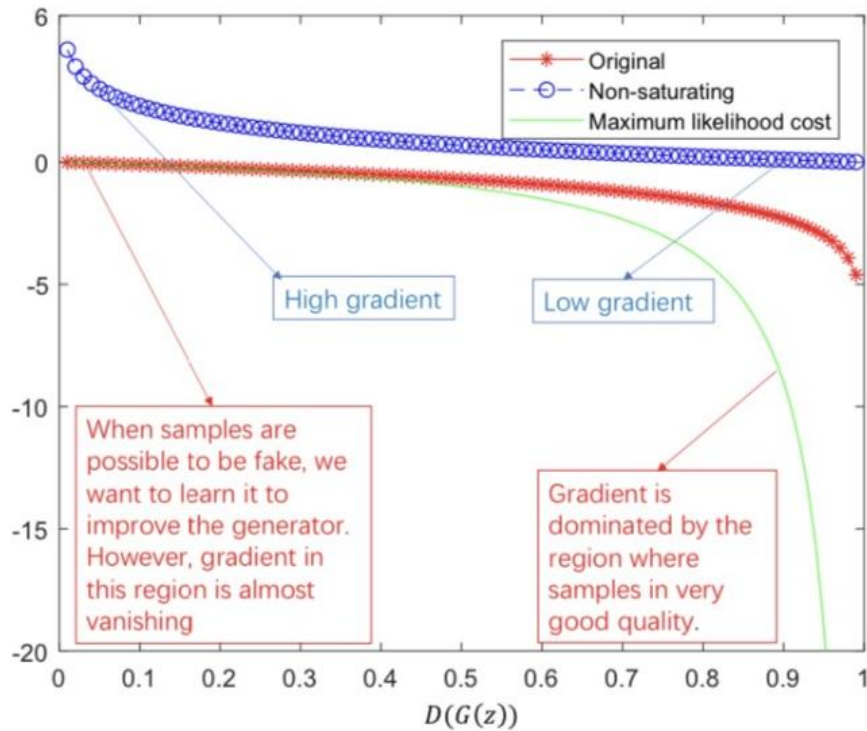


GAN의 문제점

1. 모델 훈련의 불안정성

수렴하지 못하는 경우가 발생

Generator와 Discriminator가 골고루 학습되어야 하는데, 편향적으로 학습될 수 있음



2. 성능 평가의 어려움

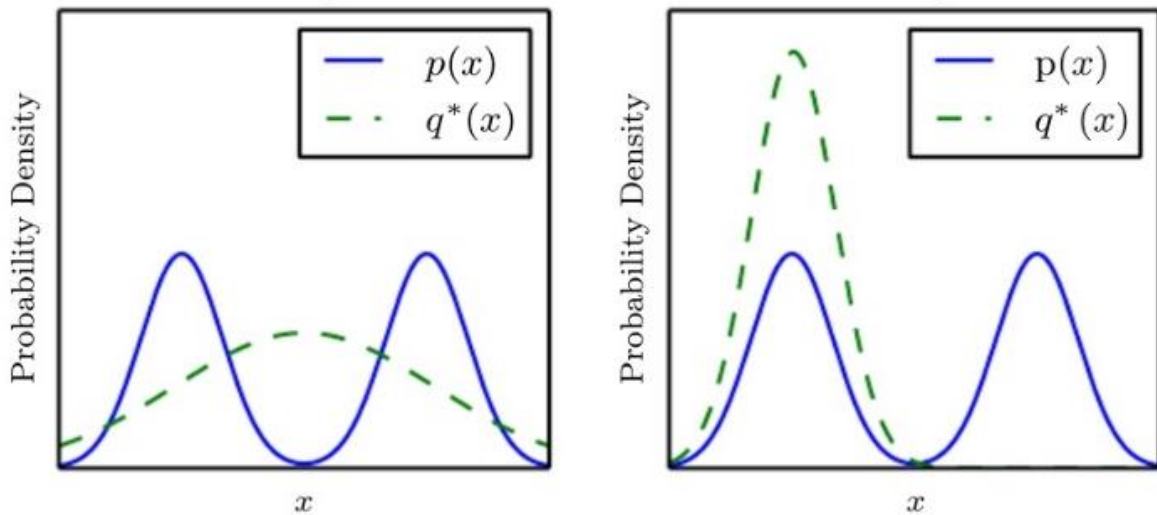
사람의 눈으로 성능을 평가해야 하는 어려운 문제가 발생

과대적합을 판단할 만한 수학적 근거가 부족

3. Mode Collapsing

클래스가 여러 개인 실제 데이터의 분포를 따라가지 않고, 특정 데이터에 강하게 몰리는 현상

모델은 손실 값만을 줄이려고 학습하기 때문에, 전체 데이터의 분포를 찾지 못하고 하나의 mode에 강하게 몰리게 되는 현상



- MNIST 데이터를 학습하는 경우 특정 숫자를 생성하지 못하는 문제
- 강아지의 제한된 색상과 특징을 생성하는 법을 학습할 수도 있음

Generator가 훈련 데이터의 분포를 충분히 모델링하지 못했기 때문에 이러한 문제가 발생

- AdaGAN, MAD-GAN은 이러한 문제를 해결하기 위하여 제안됨

GAN의 평가지표

1. Inception Score (IS)

2016년 "Improved Techniques for Training GANs" 에서 처음 제안

주로 선행 기술에 비해 개선된 점을 보여줄 때 많이 사용

GAN의 성능 평가에 널리 사용되는 지표

$$IS = \exp(\mathbb{E}_{x \sim p_{data}} D_{KL}(P(y|x) \| P(y)))$$

생성된 이미지를 평가할 때, 중요한 지표들 모두 고려하여 점수를 계산

1. Fidelity (이미지의 질)
2. Diversity (이미지의 다양성)

이미지가 학습된 클래스들과 얼마나 많이 닮았는지, 알려진 클래스에 걸쳐 얼마나 다양한 이미지세트를 캡처하는지 단일 이미지에 대한 라벨의 확률분포로 반환

GAN으로부터 만들어진 이미지를 분류기를 통해 생성된 이미지의 속성을 측정하고, 생성된 이미지들에 대한 레이블 확률 분포를 결합을 통하여 Generator의 출력의 다양성에 대해 알 수 있음

이미지가 뚜렷하게 보이는지와, 생성기의 출력에 다양성이 있는지를 하나의 점수로 결합

Label에 따라 분류가 잘 되고, 다양성을 가지는 정도에 따라 점수 계산하는데, 이때 두 확률분포가 얼마나 유사하거나 다른 지에 대한 척도를 계산할 수 있는 KL-Divergence를 사용

각각의 생성된 이미지에 대하여 고유 라벨이 있으며, 생성된 이미지의 전체 집합이 다양한 범위를 가질 때 점수가 가장 높음

1. 분류기의 훈련 데이터 세트가 없는 무언가를 생성하려는 경우, 해당 이미지가 특정 레이블로 분류되지 않기 때문에 항상 낮은 IS를 가짐
2. 분류기의 훈련 데이터와 다른 레이블 세트를 사용하여 이미지를 생성하는 경우, 다양성이 부족하다고 판단되어 IS 점수가 낮음
3. 분류기 네트워크가 이미지 품질 개념과 관련된 특징을 감지할 수 없는 경우, 예를 들어 이미지 분류 네트워크는 Texture에 포커스를 맞추기 때문에 인간의 눈에는 좋은 결과물이 아니더라도 높은 IS 점수를 받을 수 있음
4. 생성기가 이미지의 클래스 당 하나의 이미지만 생성하고 각 이미지를 여러 번 반복하는 경우, IS 클래스 내의 다양성을 측정할 수 없음
5. 생성기가 훈련 데이터를 그대로 기억하고 복제하는 경우, 높은 IS 조건에 충족하지만, GAN의 목적인 실제 같은 가짜 이미지를 생성하는 것은 아니기 때문에 한계점

2. Frechet Inception Distance (FID)

Frechet Inception Distance(FID)는 생성된 이미지의 분포와 원래 이미지의 분포가 어느정도 비슷한지 측정하는 지표

$$FID = d^2 = \|\mu_1 - \mu_2\|_2^2 - Tr(\Sigma_1 + \Sigma_2 - 2\Sigma_1\Sigma_2)$$

이미지는 고차원 공간이므로 간단하게 분포의 거리를 계측하기 어려움

따라서, 최근 인간을 초월한 이미지 인식 정밀도를 가진 모델을 사용하여 이미지를 저차원 공간으로 변경한 뒤, 그 공간에서 분포의 거리를 구함

Inception V3이라는 네트워크로 저차원 공간으로 만든 벡터로 Wasserstein-2 거리를 계산

m, c 는 매립된 공간상에 평균 벡터, 공분산행렬로, 첨자로 붙은 w 는 생성 이미지, 첨자가 붙지 않은 것은 실제 이미지

분산의 거리이므로, 값은 작을 수록 실제 이미지에 가까운 가짜 이미지가 생성

즉 Generator 성능이 좋음

3. Learned Perceptual Image Patch Similarity (LPIPS)

2018 "The Unreasonable Effectiveness of Deep Features as a Perceptual Metric"

기존의 IS 나 FID와는 다르게 유사도를 사람의 인식에 기반하여 측정하려 시도

AlexNet, VGG, SqueezeNet 의 feature map 이 사람의 인식과 유사하기 때문에 이를 활용

$$LPIPS = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \|w^l \odot (\hat{y}_{hw}^l - \hat{y}_{0hw}^l)\|_2^2$$

Classification Task를 Supervised, Self-supervised, Unsupervised 딥러닝 모델로 학습하고,

비교할 이미지 2 개를 각각 학습된 Network를 사용해 deep feature(Activation 결과값)를

추출하고, 이를 비교하여 유사도를 평가

기존 방법론 보다 좀더 perceptual 하게 이미지들의 유사도를 평가