

9주차 Gen Basic 과제2

How to evaluate GANs ? Search for some metrics and review them as well.

GAN은 다음 두 가지 관점에서 평가된다.

1. **Quality** : 실제 이미지와 유사한 이미지가 생성되는가
2. **Diversity** : 여러 noise vector에 걸쳐 다양한 이미지가 생성되는가

위 두 가지 관점을 바탕으로 GAN을 평가하기 위해 사용되는 다음 3가지의 metric들을 소개하고자 한다.

Inception Score (IS)

ImageNet pretrained model인 inception-v3를 이용해서 GAN을 측정한다. IS를 계산하기 위해 두 가지 확률을 필요로 한다.

1. 조건부 확률 $P(y|x)$: 이 조건부 확률로 생성된 이미지 x 에 대해서 어떤 클래스에 속할지 예측하는 것이 필요하다. 고품질의 이미지를 생성할 수록, 하나의 클래스에만 속할 확률이 높아 엔트로피가 낮게 측정될 것이다.
2. 주변 확률 $P(y) = \int_z P(y|G(z))$ 로 여러 noise vector에 대해서 예측된 class들을 의미한다.

만약 GAN이 diverse한 이미지를 생성한다고 하면 엔트로피가 $P(y)$ 가 uniform하게 나오게 될 것이고, 높은 엔트로피를 갖게 된다. IS를 계산하기 위해서 조건부 확률과 주변 확률의 KL-divergence를 계산하면 아래와 같다.

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

Frechet Inception Distance (FID)

FID를 측정하기 위해 ImageNet으로 pretrain된 Inception-v3모델이 필요하다. Inception-v3 모델을 feature extractor로 사용하여 실제 이미지와 생성된 이미지 사이의 activation

map을 추출한다. Activation map을 통해 multivariate gaussian distribution을 구할 수 있다. 그럼 이제 실제 이미지의 확률 분포와와 생성된 이미지의 확률 분포 사이의 Wassertein-2 distance를 측정하여 FID를 구할 수 있다.

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

- 여기서 μ 와 Σ 는 각각 평균, 공분산 행렬을 의미하고, Tr 은 행렬에서 주 대각선의 합을 의미한다.
- FID가 낮게 나온 생성된 이미지는 실제 이미지와 확률 분포적으로 유사하다고 볼 수 있다.
- 따라서 GAN의 FID가 낮게 측정될 수록 high-quality의 이미지를 생성한다고 평가된다.

Learned Perceptual Image Patch Similarity (LPIPS)

LPIPS는 비교적 초기의 ImageNet classification 모델인 AlexNet, VGG, SqueezeNet을 사용한다. 기존의 IS나 FID와 달리 유사도를 사람의 인식에 기반하여 측정하려 시도한다. 그 과정에서 AlexNet, VGG, SqueezeNet의 feature map이 사람의 인식과 유사하기 때문에 이를 활용한다.

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

낮은 LPIPS는 두 이미지가 perceptually similar하다고 볼 수 있다.

▼ 참고자료

<https://jjuon.tistory.com/33>

<https://wikidocs.net/149481>