

GAN 성능 지표

1. CID index

3가지 평가지표를 곱한 지표

$$CID = Creativity \times Inheritance \times Diversity$$

1) Creativity : 실제 image와 중복이 있으면 안됨 -> 전체 생성된 image들 중에 실제와 유사한 image들이 아닌 image들의 비율

$$= \frac{|G^{rem}|}{|G|}$$

2) Inheritance : 생성된 image는 같은 스타일을 가져야함(이때 생성된 image의 분포가 실제의 분포에 가까워야됨) -> GLCM-contrast를 적용한 R 과 G^{rem} 의 각 평균 차이를 정규화한 값

$$= 1 - \frac{|gc_r - gc_g|}{\max\{gc_r, gc_g\}}$$

3) Diversity : 생성된 image들은 서로 달라야함

$$= - \sum_i^m p_i \log p_i; p_i = \frac{|C_i|}{|G^{rem}|}$$

※ 기호 정리

R	실제 image 집합
G	GAN으로 생성된 image 집합
G^{rem}	G에서 실제 image와 유사도가 80%보다 낮은 image 집합
gc_r	GLCM-contrast를 적용한 R 의 평균
gc_g	GLCM-contrast를 적용한 G^{rem} 의 평균
$C_1 \cdots C_m$	유사한 image들로 구성된 각 cluster들의 entropy

※ GLCM(Gray Level Co-occurrence Matrix) = 통계적 질감 분석 방법으로 현재 픽셀과 그 이웃하는 픽셀의 밝기 값의 관계를 대비, 상관관계, 에너지, 동질성 등과 같은 통계량으로 계산하여 표현하는 것으로 밝기를 나타낸 영상에서 정의한 변위벡터의 거리와 방향이 일치하는 픽셀 쌍의 빈도수를 표시하는 빈도수 Matrix.

2. Inception Score(IS)

IS는 ImageNet pretrained model인 inception-v3를 이용하여 GAN을 측정한다.

1) 조건부 확률 $P(y|x)$: 생성된 이미지 x 에 대해서 어떤 클래스에 속할지 예측하는 것
 -> 고품질의 이미지를 생성할수록, 하나의 클래스에만 속할 확률이 높아 엔트로피가 낮게 측정된다.

2) 주변 확률 $P(y) = \int_z P(y|G(z))$: 여러 noise vector에 대해서 예측된 class들을 의미
 -> 만약 GAN이 diverse한 이미지를 생성한다면 엔트로피가 $P(y)$ 가 uniform 하게 나올 것이며, 높은 엔트로피를 갖게 된다.

=> IS를 계산하기 위해서 조건부 확률과 주변 확률의 KL-divergence를 계산해준다.

$$: IS = \exp(E_{x \sim p_{data}} D_{KL}(P(y|x) \parallel P(y)))$$

3. Frechet Inception Distance(FID)

FID를 측정하기 위해 ImageNet으로 pretrain된 Inception-v3모델이 필요하다.

Inception-v3모델을 feature extractor로 사용하여 실제 이미지와 생성된 이미지 사이의 activation map을 추출한다. Activation map을 통해 multivariate gaussian distribution을 구할 수 있고 실제 이미지의 확률 분포와 생성된 이미지의 확률 분포 사이의 Wassertein-2 distance를 측정하여 FID를 구할 수 있다.

$$: FID = d^2 = \|\mu_1 - \mu_2\|_2^2 - \text{Tr}(\Sigma_1 + \Sigma_2 - 2\Sigma_1\Sigma_2),$$

(μ = mean, Σ = covariance, Tr = matrix에서의 대각합)

FID가 낮게 나온 생성된 이미지는 실제 이미지와 확률 분포적으로 유사하다고 볼 수 있다.
 따라서 GAN의 FID가 낮게 측정될수록 high-quality의 이미지를 생성한다고 평가된다.

1. Learned Perceptual Image Patch Similarity (LPIPS)

두 이미지 x, x_0 가 주어졌을 때, ImageNet으로 pretrain된 network의 몇몇 layer l 에서의 activation map $\hat{y}_{hw}^l, \hat{y}_{0hw}^l$ 을 얻어 Euclidean distance 계산한 후 w^l 로 scaling한 다음 channel-wise averaging을 한 값을 l 에 대해 평균을 내면 LPIPS가 얻어진다.

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

낮은 LPIPS는 두 이미지가 perceptually similar하다고 볼 수 있다.