An enhanced GAN for image generation

For Fréchet Inception Distance (FID) [1], firstly, we use a Inception V3 model to extract image features. Secondly, we use Eq. (1) to compute difference between obtained features from generated and original images as FID values. The FID value is smaller, generated images are closer to real images. Mentioned Eq. (1) can be shown as follows.

$$FID = \|\mu_{\mathrm{g}} - \mu_r\|_2^2 + T_r(\sum\nolimits_g + \sum\nolimits_r - 2(\sum\nolimits_g \sum\nolimits_r)^{\frac{1}{2}}) \tag{1}$$

where $(\mu_{\mathrm{r}}, \sum_r), (\mu_{\mathrm{g}}, \sum_g)$ are the mean and variance of given reference and generated images.

Learned Perceptual Image Patch Similarity (LPIPS) [2] is used to measure the difference between given reference and generated images. When LPIPS value is lower, corresponding method has poorer performance of image generation. Conversely, corresponding method has better performance in image generation. Mentioned process can be summaried as the following equation.

$$LPIPS = d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{h,w} \left\| w_l \odot (\hat{y}^l_{hw} - \hat{y}^l_{0hw}) \right\|_2^2 \tag{2}$$

where (x, x_0) are real image (given reference image) and generated image, l represents the index of the layer, H_l and W_l represent the height and width of the feature map in layer l, h and w represent the coordinates, w_l is the weight of layer l, and $\hat{y}^l_{hw}$ and $\hat{y}^l_{0hw}$ represent the features of the image in layer l with coordinates (h, w).

Multi-Scale Structural Similarity Index Measure (MS-SSIM) [3] is used to compute structural similarity (SSIM) between different images at different scales. When MS-SSIM value is lower, corresponding method has better performance for image generation. Mentioned illustrations can be summarised as the following equations.

$$MS - SSIM(x, y) = [l_M(x, y)]^{\alpha M} \cdot \prod_{j=1}^{M} [c_j(x, y)]^{\beta_j} [s_j(x, y)]^{\gamma_j} \tag{3}$$

$$l(x, y) = \frac{2\mu_x \mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \tag{4}$$

$$c(x, y) = \frac{2\sigma_x \sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \tag{5}$$

$$s(x,y) = \frac{2\sigma_{xy} + C_3}{\sigma_x \sigma_y + C_3} \tag{6}$$

where M is the scale, l, c, s are the luminance, contrast and structure comparison measures, x, y are the real image and the generated image, the exponents $\alpha M, \beta_j$ and γ_j are used to adjust the relative importance of different components, C_1, C_2 and C_3 are small constants given by $C_1 = (K_1 L)^2, C_2 = (K_2 L)^2, C_3 = \frac{C_2}{2}$.

Kernel Inception Distance (KID) [4] is used to evaluate the quality and diversity of generated models, which measures the difference between generated and real images. It can be expressed as the following equation.

$$KID = \left\| \frac{1}{N} \sum\nolimits_{i=1}^{N} x_i x_i^T - \frac{1}{M} \sum\nolimits_{i=1}^{M} y_i y_i^T \right\| F \tag{7}$$

where x_i, y_i are the feature of generated image and real image, $\|\cdot\| F$ represents the Frobenius norm, which is the square root of the sum of squares of the elements of a matrix.

Number of statistically-Different Bins (NDB) [5] is a metric to test performance of a generator. Its basic idea is that if there are two sets of samples that should represent the same distribution, then apart from the influence of sampling noise, the number of samples falling into a given box should be the same. It can be expressed as the following equation.

$$\frac{1}{N_p} \sum\nolimits_i I_B(s_i^p) \approx \frac{1}{N_q} \sum\nolimits_j I_B(s_i^q) \tag{8}$$

where $I_B(s)$ is the indicator function for box B,(s_i^p) is N_p samples from distribution p and (s_i^q) is N_q samples from distribution q.

Inception Score (IS) [6] can be used to measure the diversity and authenticity of image generation models and its value can be obtained as follows.

$$IS = \exp(E_{x \sim p_g} D_{KL}(p(y\,|x\,) \,\|p(y)\,)) \tag{9}$$

where $x \sim p_g$ represents result x, which is an image sample generated by p_g, and $D_{KL}(p\,\|q)$ represents the KL divergence of the distributions p and q, $p(y\,|x\,)$ represents the probability of being classified as y under a given image x, and $p(y)$ represents the edge distribution of the category.

References

[1] Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. Advances in neural information processing systems **30** (2017)

[2] Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 586–595 (2018)

[3] Wang, Z., Simoncelli, E.P., Bovik, A.C.: Multiscale structural similarity for image quality assessment. In: The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, vol. 2, pp. 1398–1402 (2003). Ieee

[4] Bińkowski, M., Sutherland, D.J., Arbel, M., Gretton, A.: Demystifying mmd gans. arXiv preprint arXiv:1801.01401 (2018)

[5] Richardson, E., Weiss, Y.: On gans and gmms. Advances in neural information processing systems **31** (2018)

[6] Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. Advances in neural information processing systems **29** (2016)