Paper Implementation Report

Wasserstein GAN

Ⅰ. Implementation of Wasserstein GAN's key contributions

The key contributions of this paper are that they provide meaningful loss metric called Earth Mover (EM) distance which is **correlates with the generator's convergence and sample quality** and they **improve the stability of training GAN,** getting rid of problems like mode collapse. To demonstrate these contributions, I conducted two major experiments. The first was **to compare Jensen-Shannon Divergence and Wasserstein-1 Distance (EM distance) to see how these metrics are related to the generated image quality using vanilla DCGAN and WGAN.** The next one was **to compare the stability of vanilla DCGAN and WGAN without batch normalization in the generator part of each model.**

I tried to **royally follow the implementation details** in this paper as far as possible. I applied weight initialization (a zero-centered normal distribution with standard deviation *0.02*) as described in the DCGAN paper and learning rate as *0.0005*, n_{critic} as 5. I also added **gradient clipping with a value of *0.005***. This is different from the value in the paper because, unlike the paper, I used **CelebA** instead of LSUN as a training dataset so that there was a difference in the data distribution and it was difficult to obtain a significant result with the original value of *0.01*. I trained the each model with *50K* steps based on *64* batch sizes.

Ⅱ. Experiment results

ⅰ. Correlation between loss metric and the generator's convergence or sample quality.

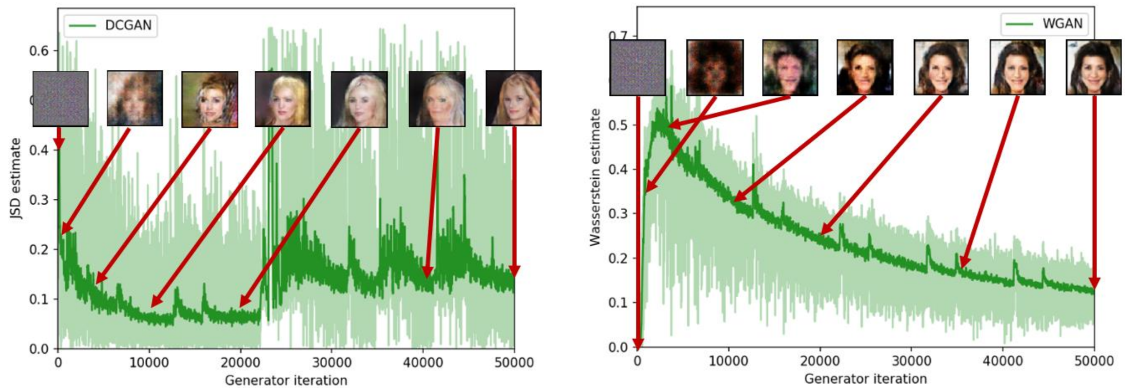

Figure 1. Compare the correlation between each loss metric and the generated image quality. (left: DCGAN, right: WGAN)

As can be seen from Figure 1, Jensen-Shannon divergence metric has no strong correlation with the quality of the generated image getting better and there is no tendency to converge. However, in the case of Wasserstein distance, it can be seen that the quality of the image is significantly improved as it decreases.

ⅱ. Comparison of the stability between DCGAN and WGAN without batch normalization.

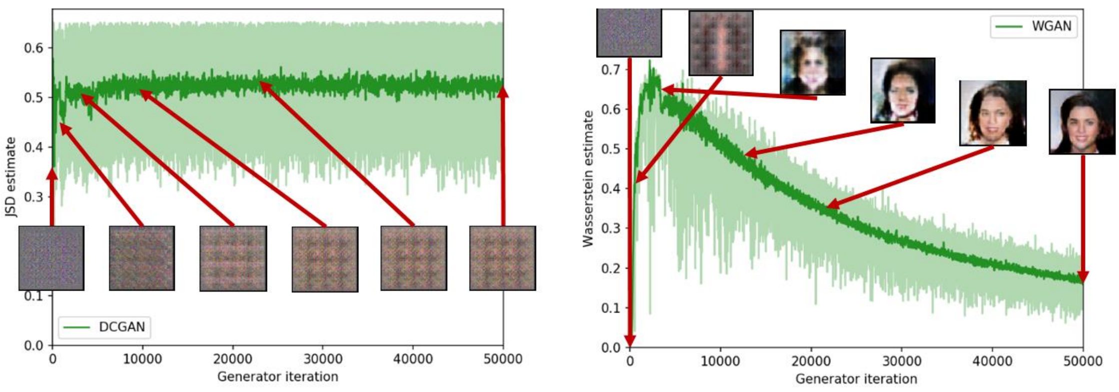

Figure 2. Compare the stability between DCGAN and WGAN without batch normalization. (left: DCGAN, right: WGAN)

As a result of training the model by removing the generator's batch norm from each model of DCGAN and WGAN, as shown in Figure 2, DCGAN could not create a meaningful image even though learning continued but WGAN was able to confirm relatively stable learning even without batch norm.

Ⅲ. Discussion

As mentioned in the paper, the above experiments confirmed the effectiveness of WGAN. However, it was not easy to find the optimal hyper-parameter for WGAN because it was CelebA, not LSUN. Also, **WGAN was only meaningful when I found the appropriate learning rate and clipping values compared to DCGAN. DCGAN also performed better when the batch norm of the discriminator as well as the generator was removed.** According to a recently published paper, the performance of certain GAN models tends to be better when the batch norm is removed. From this point of view, it is doubtful whether the experimental results on the stability of the WGAN presented by the authors were valid.