Learning to generate Chest X-Ray Images from sample data

Umesh Singla
usingla@ucsd.edu

Abstract

We use an existing implementation of deep convolutional GAN in pytorch to learn and generate chest X-Ray images. The dataset includes combined 5216 images of healthy people chest X-Rays and that of people with pneumonia (Figure 1).

Note: 1 grace day used bringing total to 4.

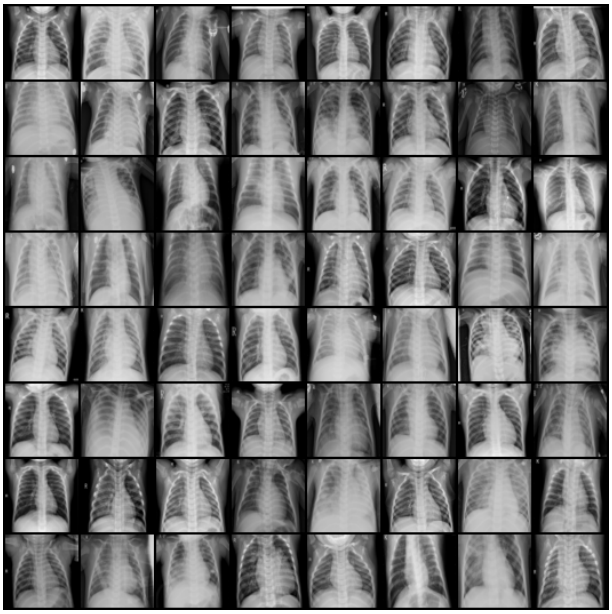

Figure 1: Sample images from the Chest XRay dataset

1 Formulation

The goal of training the discriminator is to maximize the probability of correctly classifying a given input as real or fake. We want to update the discriminator by ascending its stochastic gradient. We maximize $\log(D(x)) + \log(1 - D(G(z)))$.

Discriminator loss calculated as the sum of losses for all real and all fake batches. We want to train the Generator by minimizing $log(1 - D(G(z)))$ in an effort to generate better fakes and we do this by classifying the Generator output from Part 1 with the Discriminator, computing G's loss using real labels as GT and instead try to maximize $log(D(G(z)))$ based on various recommendations in the field.

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

The loss formulation uses gradient descent optimization class method (Adam optimizer) to learn parameters. Learning Rate is set to 0.0002 and beta hyperparameter is set to 0.5. We study the loss, FID and IS scores as the training progresses.

The number of latent dimensions is kept 100.

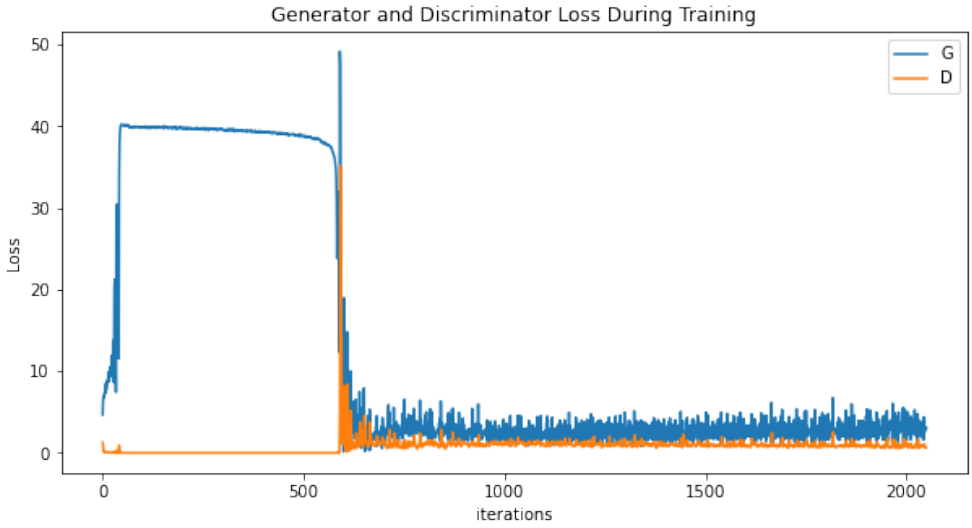

Figure 2: Plot of D and G's losses versus training iterations.

2 Generated Images

After each epoch of training the model, we generated an image feeding only random inputs of size 100x1 to the latent space. Figure 3 shows an example of how generated images evolved during the training.

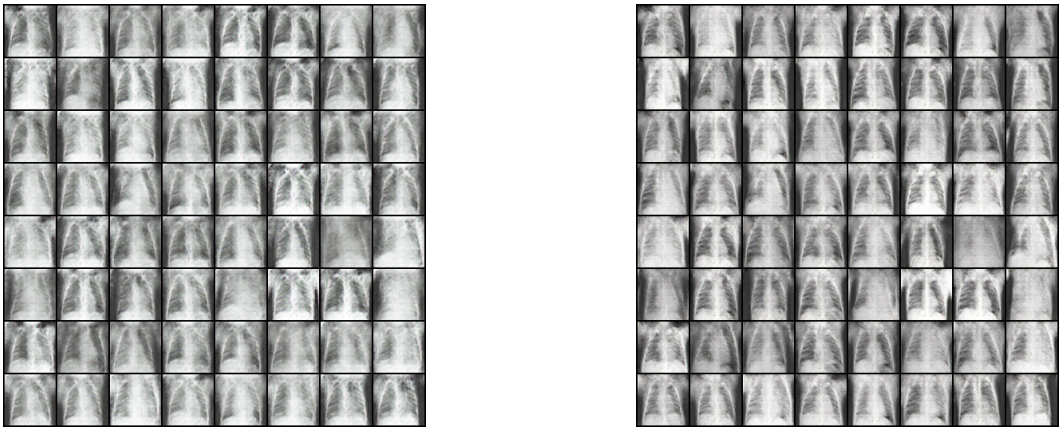

Figure 3: Images generated by the model at halfway (left) and towards the end (right) of the training using only random inputs.

3 IS score

The Inception Score (IS) is an objective metric for evaluating the quality of generated images. The inception score involves using a pre-trained deep learning NN model for image classification to classify the generated images. The score tries to capture quality and diversity in generated images. The inception score has a lowest value of 1.0. (Table **??**)

We generated a total of 4096 fake images from noise using the generator at the end of training and got an inception score of 1.14. More epochs of training and more hyperparameter tuning can help the IS score but given the inherent low variety of our images, a good inception score is hard to achieve.

4 FID score

The Fréchet inception distance (FID) is another metric used to assess the quality of images created by a generative model. FID compares the distribution of generated images with the distribution of real images that were used to train the generator. Rather than directly comparing images pixel by pixel (as in L2 norm), the FID compares the mean and standard deviation of one of the deeper layers in Inception v3. Lower scores indicate the two groups of images are more similar with a perfect score being 0.0 indicating that the two groups of images are identical. (Table 1)

Iterations	FID
500	7.0850
1000	3.8293
2000	3.3745

Table 1: FID scores as the training progressed.