Gen AI Assignment 4

Sarah Groark

October 31, 2024

Parameters

The GAN model features the following parameters:

Generator =

- 1 dense layer: 16,384 units with shape 100 noise vector as input
- 3 transpose convolution layers with 128, 64, and 32 filters, respectively
- Each transposed convolution layer features a batch normalization layer and a leaky relu activation layer
- Output layer returns images in the shape (None, 64, 64, 3)
    - Image size is 64 x 64
    - Channel number indicates presence of color images

Discriminator =

- 3 convolutional layers with 32, 128, 256 filters each, respectively
- Flatten layer reduces the size into a one dimensional vector
- Dense layer outputs a single value that represents the probability of if the image is real or fake

Batch Size = 128 and 256
Epochs = 400
Generator Learning Rate = 1e-4
Discriminator Learning Rate = 1e-4

Discussion Points

- Discuss the trade-off between generating high-quality images and maintaining diversity in the output. Are the generated images too similar, or does the model capture a wide range of bedroom styles?

In attempting to generate images, the tradeoff between quality and diversity is typically a common struggle. On one end, emphasis on the generation of high quality images may result in less variety in the generated images and more images that more closely resemble each other. On the other hand, if diversity is important, then the quality of the images might be sacrificed as a result of attempting to explore a broader range of features to use in image generation.

The generated images, although the model was trained for several hundred epochs, do not resemble the bedroom styles. However, the plot of generated images are all different and do not resemble each other.

- Consider potential enhancements to the model architecture or training process. Would experimenting with different hyperparameters or batch sizes yield better results?

Fine tuning the hyperparameters may have positive implications for the generated outputs of the model. Foundationally, increasing the amount of epochs would allow for the model to train for a longer period of time, which could likely lead to convergence between the generator and discriminator. Furthermore, the batch size could be modified based on what the model needs. If there is a need for convergence, a larger batch size could be useful; if the model is stuck in local minima and needs more noise introduced, or training is taking too long, it may be advantageous to use a smaller batch size. Another hyperparameter that could be optimized is the learning rates of the generator and discriminator. If one is overpowering the other, reducing the learning rate of the overpowering component would allow the other to train more efficiently.

- What are the practical applications for this type of generative model?

Generative Adversarial Networks are often used in image synthesis, creating realistic portrayals of art, human faces, objects, and more. Additionally, a wide variety of use cases are utilizing GANs, such as healthcare imaging, marketing and advertising, chemical solutions, and even astronomy. One specific example features the use of a GAN to convert low-resolution medical images to high-resolution images. Acquiring high-resolution imaging is costly, so the use of a model to retain high-resolution image features allows for the reduction of cost, while also pulling out complex details from images. Consequently, the physicians are able to make more informed decisions with greater confidence.

Extra Credit Research Problem

Inception score (IS) is used in machine learning to quantitatively determine how well an image is generated, namely, how realistic it is to the human eye. The inception score is based on the quality of the image as well as the diversity in the images it produces. To calculate the inception score, the utilization of a pre-trained inception model (popularly Inception-v3) is needed to pass generated images through the network to obtain the probability of its belonging to a certain class (done via the use of the softmax function). After obtaining the probabilities of all of the generated images, the results are summed and divided by the total number of classes, representing the empirical estimate of the distribution. Next, the KL divergence for each image's softmax probability is calculated against the average of the distribution as calculated in the previous step. The final inception score result is the average of the calculated KL divergence values. This returned score signifies the quality and diversity of the generated image; the higher the inception score, the better quality image.

The inception distance (also known as Fréchet inception distance) was inspired by the inception score. ID, in contrast to IS, evaluates the distribution of the generated images in comparison to the real images (the 'ground truth'). The inception distance is calculated as the distance between these two distributions of the images, taking the mean and variance from the deepest layer in the inception model (in attempts to extract the most detailed features in this layer, rather than the previous shallow layers). This metric also aims to quantitatively assess the quality and diversity of images, just with a different mathematical approach.