Katrina Cwiertniewicz
GenAI Assignment 4
11/4/2024

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?

For my output images, all images were duplicates of each other. This was likely due to limited training data for the model. I believe it was trying to fit to similar images in the data set that led to the same block of color being duplicated. For the model to capture a wide range of bedroom styles, it would need to have additional convolutional layers and a dropout rate that can increase the chances the generator is making high-quality images rather than only trying to fool the discriminator. Adding more images from the dataset would increase diversity in the model, but may lead to lower-quality images due to the generator not extracting the important features of the model if more epochs were not added.

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

I noticed that when reducing the images from 256 to 64, the generated model had more visible pixels and was more rigid. The 256-sized image looked like it would do well during training and create more realistic images. The enhancements I would make to this model are for the dataset size to be at least 5000 images to diversify the output of the images, and more convolutional layers to increase the extraction of features. Additionally, I would start with a higher amount of filters to produce images during training that would visually look more connected and adjust to feature shapes in the image better than the limited data model.

The challenge started with the addition of more layers and data as the use of TPU was not able to train the model in an efficient amount of time leading to example 1 average training epoch to be 40 minutes and example 2 first epoch not loading. The successful model trained an average of 2 minutes for each epoch and generated very little image depth.

For my model, I trained batch sizes that were 128 and 256. The model seemed to train quicker, but the model with a batch size of 128 was not able to generalize the images well. I think many factors affected the image quality, however, if I were able to successfully train a model with a batch size of 32 or 64, I would see a better representation of generated images.

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

This generative model could be used in many different settings depending on if a dataset exists that would work for its purpose. For example, a GAN could be used to generate new images for marketing images or create different fashion combinations using a dataset of clothes. It could also be used to determine counterfeits of expensive jewelry or watches to classify the images as real or fake.

Extra Credit Research Problem: Consider how you might quantitatively evaluate the quality of the generated images.
Hint: *Inception Score* (IS) or *Inception Distance* (ID) can provide measures of image quality and diversity.

Quantitative evaluations are beneficial when evaluating the quality of the generated images because it is challenging to know when the training has reached a point where the images displayed are diverse and of high quality. This can be time-consuming and subjective of what images match the criteria depending on what the model is being trained for. Some examples of quantitative evaluations are Inception Score (IS), Inception Distance (ID), and Learned Perceptual Image Patch Similarity (LPIPS).

Inception score measures the diversity and how much the generated image looks like a known class by finding the probability of the image belonging to each class to determine realistic characteristics. For this assignment, the generated images could be compared to classes of different rooms to determine if it would receive a high inception score if it could match the correct bedroom class and its overall diversity.

Inception Distance uses the Frechet distance to calculate the distance between the two distributions of real and generated images. For this assignment, the distribution of the bedroom dataset and generated images are compared to determine how diverse and interesting the images are. This measures how much effort it takes to create these generated images.

Both IS and ID are considered content variant metrics. LPIPS would be considered more similar to a content invariant metric which is used when there is only one correct answer. LPIPS evaluates image similarity by comparing the activations of two image patches. For our assignment, variant metrics make more sense as the generated image could have different objects included and doesn't have one correct answer. However, it could be helpful if we wanted to generate images that were very similar to the dataset that included a specific piece of furniture and the goal was not diversity.