Hong Pengfei 1002949

1. What are the differences between this work (Self-supervised GAN) and the original unconditional GAN as discussed in 50.035?

The first difference is during training it add an extra loss function – rotational loss for both discriminator and generator that can force the discriminator to learn high level features and stables training.

The second difference is that it not only predicts which image is generated also predicts the self-generated label of the image which partly belongs to the conditional Generation process.

1. In Section 2, the authors discuss discriminator forgetting. Explain what discriminator is forgetting. Why does it happen in training a GAN?

The training process for GAN is an iterative of 2 processes, 1. Fix the generator G and update the Discriminator D. 2. Fix Discriminant D and update Generator. For GANs, varying the level of details maybe different learning problems for GANs, for example, learning the global structure representations in the image and its details such as resolution and texture can be different problems. While the generator learning generates the global structure, the discriminator will most likely be classify based on the general structure of the image. But after the generator finished learning the general structure of the image and can predict perfectly with the structure of the image, but cannot perform well on the details of the image. Because it is not explicitly required to maintain a general data representation, the discriminator will focus more on distinguished the details and forget about the **generalizable representation of the images it has learnt during the previous stage**.

1. What are the results in Figure 3? What is the motivation of this experiment? There are repetitive dips in Figure 3(a) and 3(b). What causes these dips? Why are some dips more significant than the others?

The results in Figure 3a shows how the classification loss on CIFAR 10 dataset varies with respect to the training iterations. Specifically, they changed the classification task to be a 1-vs-all task, so they only predict one label for 1k steps and then they switch to the next stage.

The motivation of this experiment is they want to show when a classifier is being trained on non-stationary environment, it can forget its previous learnings and not able to learn a generalization representation. Since the performance for figure 3a did not increase when it starts a new cycle. But when the adding the self-supervised loss in this case rotation loss, the model is able to progressively learn the generalizable features between images showed in Figure 3b. This is a similar case for GAN during training since the discriminator in GAN is also in a non-stationary environment, therefore **proving after the adding the rotation loss for GANs it may help discriminator to learn global representations in the image and solve the Discriminator Forgetting problem.**

The dips are because every 1k iterations, they change the task, so the label for the dataset is all changed, the weight in the model is able to work on predicting the previous label but not working well on predicting this label, since they are different. Therefore, it caused dips in performance every time it changes its task.

Some dips are more significant than others because the weights after training for one class maybe a bad initialization for the other class. An example can be made like this: if we have trained a classifier to predict dogs, after switching the label to another class like airplane, since dogs are very not alike airplane, it is a bad initialization for airplane and it will almost predict all the airplanes wrongly. But if we have trained a class like automobile, then we switch to predict truck which is somehow similar to automobiles, it may be considered as a good initialization weight for predicting truck, and the weights can be trained easily.

1. What are the purposes to perform rotation degree classification in Figure 1? How is Figure 1 related to the equations in Section 3?

The purpose to perform a rotation degree classification is to prevent the discriminator forgetting problem mentioned before because it can force the model to learn a generalization of the model representation, more stable representation of the image and high-level abstraction of the image without having to label the dataset.

A close up of an animal

Description automatically generated

This part refers to the discriminator loss in V(G, D) that the discriminator wants to minimize, it distinguish which photo is generated and which is not.

A picture containing clock

Description automatically generated

This part refer to the rotation loss in the equation in the image. ﻿αEx∼PgEr∼R [logQD(R = r | xr)] refer to the discriminator classify and its loss on the generated image (the red line) which affects generator. And ﻿﻿βEx∼PdataEr∼R [logQD(R = r | xr)] refer to the discriminator loss on the real images (blue line).

1. In Section 3, in the equation for *LG*, why *Pdata* is not used in the rotation-based loss?
2. In order for the discriminator to predict the correct angle for by the generation, the Generator must generate images that have a generalizable global structure that works on all the images, therefore by adding this loss, this can enforce the Generator to generate pictures with generalizable representation.
3. As the Pdata is independent on the Generator, using Pdata in the rotation-based loss will not affect Generator as the gradient will only flow back to it.