

[https://github.com/mrog9/Deep\\_Learning.git](https://github.com/mrog9/Deep_Learning.git)

For the DCGAN and WGAN models, the architecture was the same as given in the homework assignment. The only difference was for WGAN model, the last layer did not include a sigmoid activation function. The “critic” was allowed to take on any value where higher the value would indicate a more realistic image.

Despite similar architectures, training was different between DCGAN and WGAN models. For DCGAN, the discriminator would be trained on a sorted batch of real and generated images. The class labels, corresponding with the real images, were sorted as well (the discriminator was not trained on each class label in sequence). For each sample in the batch, if it was fake, would contribute  $-\log(1-d)$  to the loss and  $-\log(d)$  if it was real. The discriminator was trained nineteen times before the generator was trained. The generator was trained once. This process was done for five epochs. The generator would generate images and each image in the batch contributed  $\log(1-d)$ . To evaluate this model, 500 generated images were used for each class label. For each label, the ten highest sigmoid scores from the discriminator were kept. Furthermore, the discriminator output (sigmoid score) was kept for each generated image. Then, 500 real images were passed through the discriminator for each class label. The discriminator output for each of these images were kept as well. After the outputs had been collected, both the fake and real outputs were normalized on the same min-max scale. Then, the mean and standard deviation were calculated for real and fake distributions separately.

The WGAN model’s discriminator was also trained on a sorted batch of real and generated images. However, it was also trained on interpolated images of real and fake images. This interpolation batch helped determine the gradient penalty for the loss. For the sorted batch, a fake image would contribute  $d$  to the loss and for a real image would contribute  $-d$  to the loss. Then, the total loss was divided by the batch size and  $10 * \text{grad\_penalty}$  was added to it. Similar to DCGAN, the generator was trained nineteen times before the discriminator was trained and this process was repeated for five epochs. For the generator, each generated image contributed  $-d$  to the loss. A similar process to DCGAN was used on WGAN to evaluate the model. Here are the best images for each model:

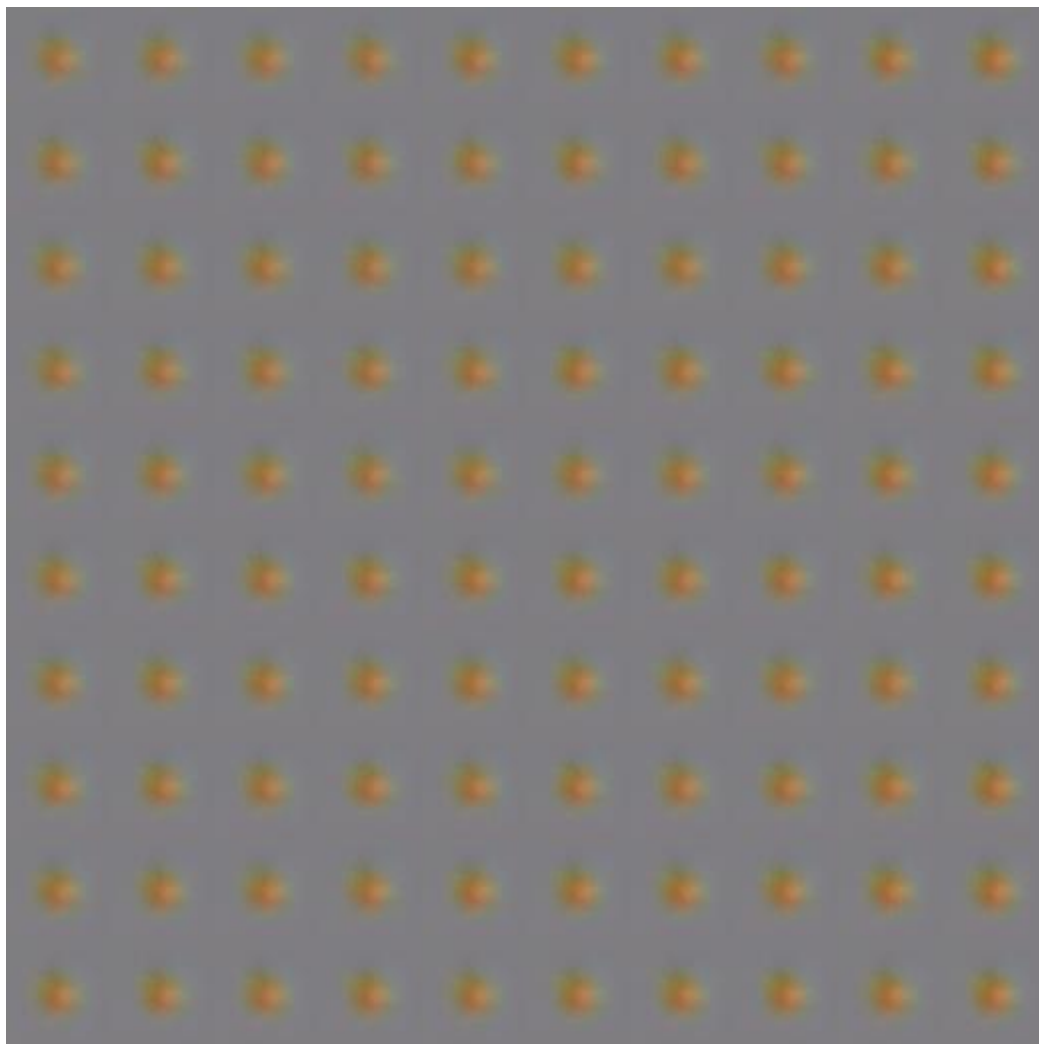


Figure1. DCGAN results of the ten best images for each class. The best images are to the right.

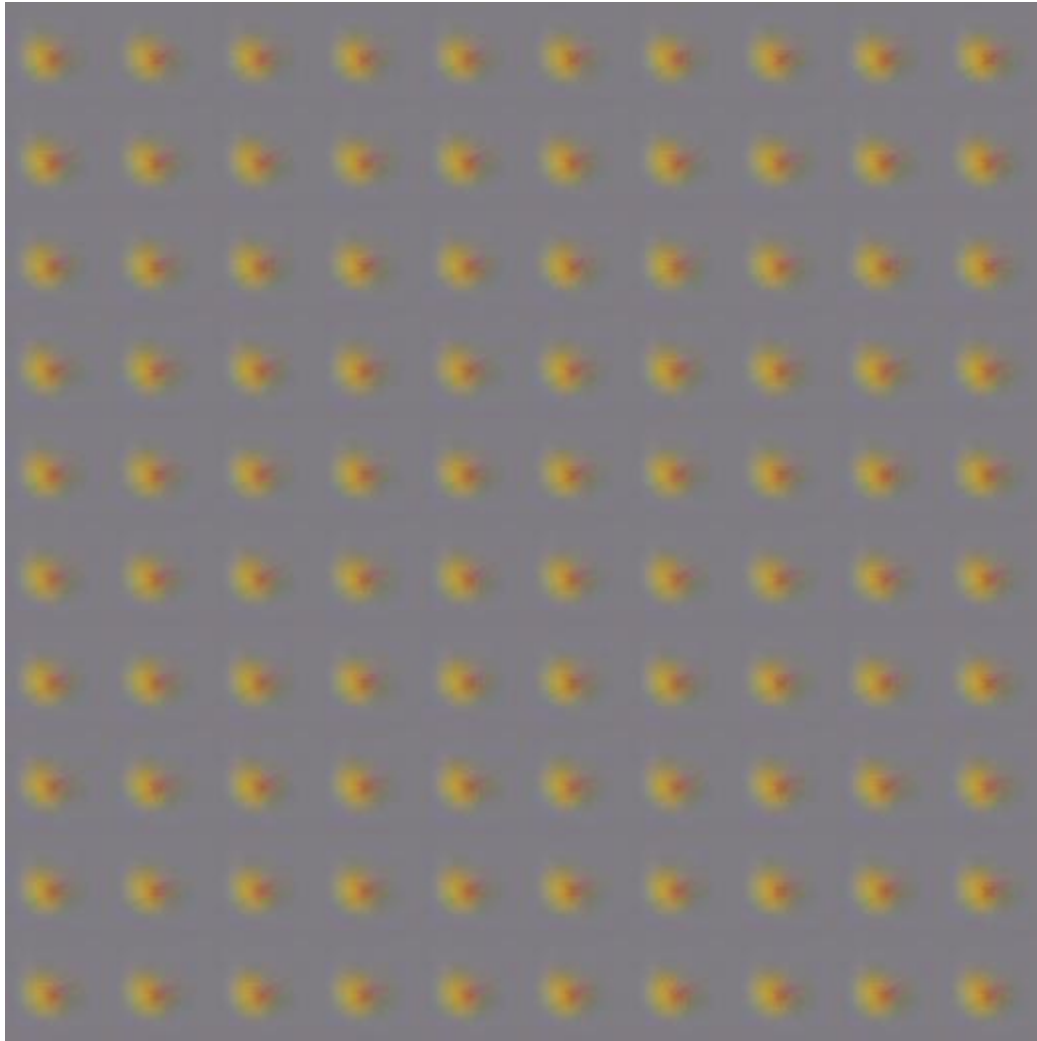


Figure2. WGAN results of the ten best images for each of the ten classes. The best images are to the right.

Qualitatively it is difficult to compare these model's images. The WGAN images could be slightly sharper than the images from DCGAN. In order to better compare the performance of these models, two scores are going to be created. The first score is going to be "confidence" which is defined as the confidence of the model's discriminator in labeling all of the images as real or fake. This score is going to be equal to  $1/\text{s.d.}$  (inverse of standard deviation). The second score is going to be "confusion" which will be defined as how "confused" the discriminator was in labeling real and generated images. This metric will be equal to  $\text{confidence}(d(\text{fake}))/\text{confidence}(d(\text{real}))$ . The mean will be ignored.

For DCGAN model, the standard deviation of the discriminator outputs for the generated images was 0.3513. The standard deviation for the real images was  $1.6858\text{e-}9$ . This means that the discriminator's confidence in that the generated image was real is 2.84 whereas for the real images it is  $\sim 600,000$ . Using these values this means the confusion score was  $\sim 4\text{e-}6$ .

For WGAN model, the standard deviation for the discriminator outputs for the generated images was 0.0045 and 0.1570 for the real images. These standard deviations mean that the confidence of the discriminator for classifying the generated images as real is  $\sim 222$  and  $\sim 6$  classifying the real images as real. The confusion score would then be  $\sim 37$ .

These confusion and confidence metrics were improvised, but illustrate an important observation between these two models. The DCGAN's discriminator was "highly confident" or significantly consistent in labeling real images (as real) and was much less consistent labeling the generated images as real. The WGAN's discriminator, instead, was more consistent in labeling the generated images as real than the real images. It could be perceived that the WGAN's discriminator was more "confused" than the DCGAN's discriminator. However, this observation is not sufficient in classifying the WGAN generator as better because the differences in how these two models were trained could have made the WGAN's discriminator more "generous" in its critique of real or fake classification.