Generative Artificial Intelligence Assignment-2

By:
A V Aashutosh (2021A7PS0056H)
Tushar Raghani (2021A7PS1404H)
Pavas Garg (2021A7PS2587H)

Introduction

- The paper proposes DCGANs with stable training, utilizes their discriminator networks for image classification and explores vector arithmetic properties of GAN generators.
- The concept of learning feature representations from unlabelled datasets is important because good intermediate representations learned from unlabeled data can be utilized in various supervised learning tasks.
- GANs are a type of neural network consisting of 2 networks - a generator and a discriminator that are trained , according to the authors, GANs offer an attractive alternative due to the learning process and absence of a heuristic cost function.
- Sometimes , GANs can be difficult to train and may produce bad outputs due to their instability.

Architecture of DCGAN

- They used batch normalization in both the generator and the discriminator.
- They have also removed the fully connected hidden layers
- Relu activation is used in generator for all the layers except the final , which used tanh activation function. Using a bounded activation helps the model to learn quickly and converge fast.
- LeakyRelu activation is used in the discriminator for all the layers except the final which has a sigmoid activation to output the probability.

Improving training of GANs with CNNs

- Replace traditional pooling layers with strided convolutions, which helps network learn how to upsize and downsize images on its own.
- Strided convolution is a technique for reducing the spatial dimensions of feature maps during the convolution operation by moving the convolutional filter with a certain stride.
- Downsizing, often referred to as downsampling or subsampling, is the process of reducing the spatial dimensions of an image.
- Removing fully connected layers, so as to connect the most important features directly to the input and output of generator and discriminator.
- Using batch normalization to stabilize the training by adjusting input data's mean and variance.

Generator
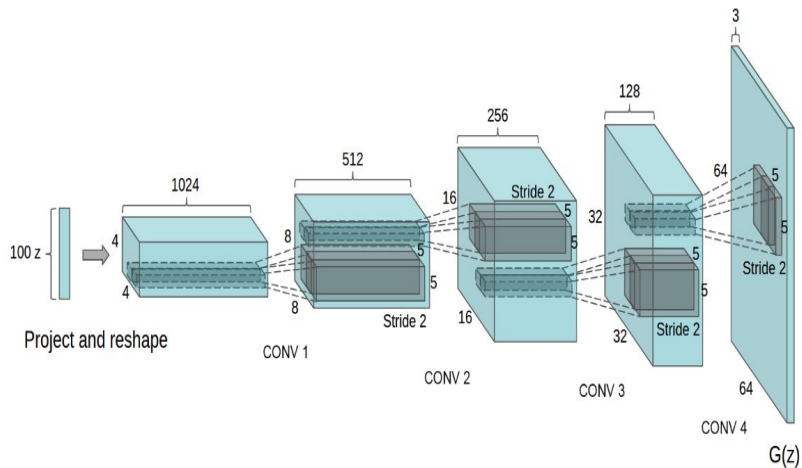
- A 100 dimensional vector z from uniform distribution (with mean 0 and standard deviation as 0.02) is fed as input to the generator.
- A series of four fractionally strided convolutions then convert this into 64*64 image.
- No fully connected or pooling layers are used.
- Learning rate of 0.0002 was used for training and beta1 value was set to 0.5

Discriminator

- A 64*64 dimensional vector which represents an image from uniform distribution is fed as input to the discriminator.
- A series of four convolutions finally output the probability of the image to be an actual image.
- No fully connected or pooling layers are used.
- Learning rate of 0.0002 was used for training and beta1 value was set to 0.5.

Training procedure

- The model was trained on 3 datasets and no preprocessing was applied to any of the training images. All the models were trained with mini batch stochastic gradient descent with 128 batch size. In the leakyRelu, the slope of leak was set to 0.2 in all models. Adam optimizer was also used.
- The results after one epoch have also been examined, addition to the ones after convergence to demonstrate that the model is not simply producing high quality images by overfitting training data
- De-Duplication was also used to ensure that the generator is not just memorising the training images.

Results

- Theoretically, the model can memorise training examples but it is experimentally unlikely as the learning rate is very low.

- There appears to be evidence of visual under-fitting as noise textures across multiple samples can be seen

Vector Arithmetics

- Through vector representations of words, a fascinating property in the representation space was demonstrated.
- By performing simple arithmetic operations, such as subtracting the word vector for Man from King and adding the vector for Woman, they obtained a resultant vector very close to the representation of Queen.
- Similar arithmetic operations also worked on the z vectors , however single example showed unstable results, but averaging the z-vector for 3 samples showed stable generations.

Conclusion

- There is some still instability in the models, there is some sort of noisy overlap in the results when vector arithmetic is applied.
- Further work would be needed to fix this instability of the models.
- Extending the framework to others domains such as audio(for speech synthesis) would be very interesting.
- There should be further research on the properties of the learnt latent space to gain more insights on it.

Results from Our Work
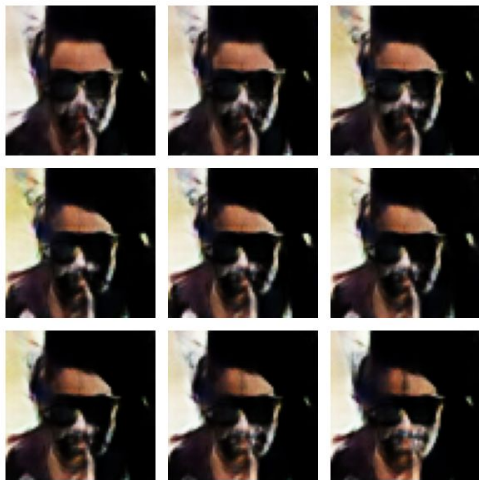
Men without glasses + People with glasses - People without glasses

DcGan Generator

Thank You