Sketches to Realistic Images Using GANs

Shaunak Pal
shaunak18098@iiitd.ac.in

Abhinav Suresh
abhinav18003@iiitd.ac.in

Ankit Mishra
ankit18019@iiitd.ac.in

March 21, 2021

1 Introduction

Machine learning has been a boon to the IT world, from opening the doors for classification problems and prediction, they are being extensively used in the field of medicine, economics, etc. Deep learning has come a step further and with such huge amounts of data has made it possible to apply artificial intellect on much more complex problems. One such complex problem is creation of realistic images from mere sketches, which look close enough to the real ones.

In this study, we as a team have taken up this project to learn how these networks work and implement a Generative Adversarial Network which could create a real life image given the sketch. It could have many use cases, one being, a sketch of a convict when passed through the network would provide with real life matching image which could be used by law enforcing agencies.

2 Literature Survey

2.1 Generative Adversarial Nets (2014)

The concept of Deep generative networks were first introduced in this paper and talks about a deep learned networks which run in parallel to each other to compete over a single objective, like, one network would try to classify between a binary relation called discriminator vs other one will try to fool it called generator, this thief-cop play makes both the model improve there respective skill of forging and detecting, eventually providing us with a network which is very well versed in creating fake data. This paper revolutionised the field of artificial intelligence and is considered a milestone in deep learning.

2.2 Conditional Generative Adversarial Nets (2014)

This concept of conditional GANs is based upon the previous paper. In an Unconditional Generative Adversarial Network, there is no control over any of the nodes. A random noise is passed into the network, and this generates solely based on randomness. This paper tackles this issues by passing conditional data as well along with the input. This is achieved by passing any type of conditional data, like class labels, as a separate layer input into both the generators and the discriminators. This will help in directing the data generational process.

2.3 Unsupervised Representation Learning With Deep Convolutional Generative Adversarial Networks (2016)

This paper merged the two concepts of convolutional neural networks that learns features from images, and the Generative Adversarial Nets to create an unsupervised deep convolutional generative adversarial network (DCGAN). This has been a major breakthrough in the field of computer vision, as this helps in creating a competitive model for other unsupervised algorithms and also show how filters learnt by GANs have learned to draw a specific type of object.

2.4 Wasserstein GAN (2017)

This paper provides a new way of calculating loss in generative adversarial networks using the Wasserstein distance. To generate realistic data, the proability distribution of the fake data must be similar to that of the real data. Prior to this paper, the distance between these distributions in GANs were calculated using Kullback-Leibler divergence (KL). This paper shows how the Wassesrtein distance is a much better loss metric, provides a termination condition for the network, and stabilizes the learning of the model as well.

2.5 Improved Training of Wasserstein GANs (2017)

This paper builds upon the previous paper about Wasserstein GANs. the research shows that even though WGANs lead to stabilized learning, it requires the discriminator function to lie in the space of 1-Lipschitz functions which was achieved in the paper by introducing weight clipping. This paper shows how this weight clipping might result in undesirable behaviour, and so introduces gradient penalty which penalizes the norm of gradient of the discriminant with respect to its input. This helps in further stabilization of the learning to the extent that there is almost no need of any hyperparameter tuning.

2.6 Progressive Growing of GANs For Improved Quality, Stability, And Variation (2018)

This paper develops a new way to train generative adversarial networks. It uses the concept of progressively increasing the resolution of the datt by adding finer details in each layer as training progresses. This benefits in both the speed of learning as well as the stability. this results in photo-realistic images. The paper also suggests a new metric for evaluating GAN results and several implementation details that helps in discouraging unhealthy competition between the generator and discriminator.

3 Baseline Models

We used the CUHK dataset (Chinese University of Hong Kong) but since it had only 188 images we used data augmentation techniques to prevent overfitting our model.

We have built 2 baseline models at present. The generator models are passed in noise of 100 random numbers and a sketch. Then they are passed through different conv layers to finally output an RGB image. The Discriminator has a structure which is exact mirror of the generator to ensure that both the generator and discriminator have an even match and improve in tandem. But since improving the generator is much harder we have trained the discriminator 5 times per batch to give a higher loss to the Generator. We used Wasserstein loss in both the baselines. We used a gradient penalty term and initiated the weights with N(0, 0.02) distribution to limit the weights from being too large or too small and stabilize the model.

3.1 Baseline A

The discriminator model includes 5 convolution layers with each using LeakyRelu with a negative coefficient of 0.2 and InstanceNorm2d (except the first and last layers) with kernels of 4x4 and stride of 2 to reduce the dimensions of the image and increasing channels by twice each time.
The Generator model uses 5 deconvolution layers to convert noise to a 64x64 image and then concatenate it with sketch. this is then passed through 4 convolution layers where first the channel is increased by 8 times and then each time decrease it by 2 times and finally passing it though a tanh activation function.

3.2 Baseline B

This baseline differs from A by being a much deeper network. We have used Residual Net blocks in order to make the models deeper as well as to prevent vanishing gradient problem.Both the Discriminator as well as Generator have 3 residual blocks of same size added to them.

In our future models we shall try to implement Pix2Pix and PROGAN techniques and various other recent advanced techniques in the field with our own flavour into them.

3.3 References

1. Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley,Sherjil Ozair, Aaron Courville, Yoshua Bengio "Generative Adversarial Nets". Available: Link [Accessed Mar. 20, 2021]

2. Alec Radford Luke Metz, Soumith Chintala, Published: April 2016 "UNSUPERVISED REPRESENTATION LEARNING WITH DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS" Link [Accessed Mar. 20, 2021]

3. Martin Arjovsky, Soumith Chintala, and L´eon Bottou1 "Wasserstein GAN" Link [Accessed Mar. 20, 2021]

4.Ishaan Gulrajani1, Faruk Ahmed1, Martin Arjovsky, Vincent Dumoulin, Aaron Courville, "Improved Training of Wasserstein GANs" CREMA-D Deatbase [Accessed Mar. 20, 2021]

5. Mehdi Mirza, Simon Osindero Link [Accessed Mar. 20, 2021]

6. Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen Published: "PROGRESSIVE GROWING OF GANS FOR IMPROVED QUALITY, STABILITY, AND VARIATION" Link [Accessed Mar. 20, 2021]