Image to Image translation using Generative Adversarial Networks

Monica Dommaraju monica.dommaraju@sjsu.edu
Swati Ganesh Narkhede swatiganesh.narkhede@sjsu.edu
Sri Sruthi Chilukuri srisruthi.chilukuri@sjsu.edu

Abstract:
Image to image translation is a methodology that comes under the class of "Computer vision techniques". It is a process of mapping a given input image to a certain class of output images depending on the goal of the business problem. Most common examples of image to image translation are: style transfer, object transfiguration, season transfer and photo enhancement. This can be performed using a wide variety of techniques. While CNN's are the baseline networks used heavily while working with image data, a range of neural networks called generative neural networks(GAN's) have become widely popular to perform such translation tasks. We have made use of a couple of such GAN networks to perform translation which converts the given satellite image into corresponding Google maps image.

Introduction:
This paper introduces a comparative study of how Conditional Generative Adversarial networks i.e. CGAN's differ in operation from the cycle Generative adversarial networks i.e. CycleGAN's. We have built two translation models using a Pix2Pix GAN network as well as using cycleGAN to analyse how the performance varies in both of them for a given input image. The Pix2Pix GAN is also a type of GAN with conditional GAN as the baseline approach which typically contains a generator and a discriminator setup like simple GAN structures. However, in CycleGAN structures, there are a pair of generators as well as a pair of discriminators as well.

Related Work:
GAN networks have always been of keen interest to many researchers to use them as a means for performing various image transformation and translation techniques. One such implementation is to train a deep convolutional neural network to convert satellite image of a specific geographic location to Google maps image.[1] The GAN architecture consists of a generator model that outputs the google maps images that synthesized at its end to inputs the discriminator for it to perform the distinguishing task. Pix2Pix GAN uses a conditional GAN or cGAN where the generation of the output is subject to a specific condition. It has then become a common model to use upon any of the datasets to obtain noticeable translated images. The dataset was first preprocessed to a compressed numpy format and split into training and validation datasets. The converted data was then available in JPEG format where the left side of the image contained satellite data and the right side of the image was Google maps data. After the dataset was prepared and was ready to be fed into the network, discriminator and

generator were modelled to perform the training. The Pix2Pix GAN used a U-Net architecture. It essentially means that during the data traversal across the network, data is preserved and prevented from loss by applying skip connections from input to output layers.

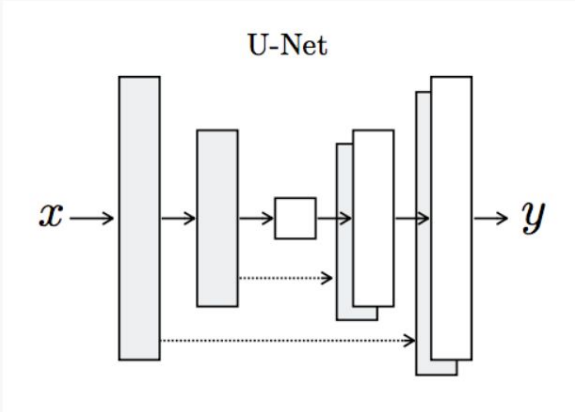

Image 1:Source: Machinelearningmastery.com

Data:

The data used for this experiment was gathered from the authorized repository of Pix2Pix architecture[2]. The obtained dataset was compressed into a numpy based npz compressed format. Then, scale of all images to match the given output image size was performed. As a next step, splitting the input image which has both source and target concatenated side by side into two seperate images (input and target) was carried out. Finally, normalizing the pixel values to [-1,1] from [0, 255] was done.

Image 2: Data preparation: Image data after preprocessing

Methodology:

Cycle GAN or cyclic GAN is a variant of GAN networks that makes use of two generators and two discriminators. It is simply composed of 2 GAN's. Given a set of different images, the discriminators check if the images computed by generators seem real or fake. Through this process, generators can better at feedback provision of their corresponding discriminators. Hence, in cycleGAN's, they get additional feedback from the opposite discriminator.The model architecture hence contains 2 GAN structures.The implementation is in Keras. The epochs for training were set to 100 with 1097 training images and batch size of 1,each iteration is done by doing 1097 training steps. So, for the whole 100 epochs the model will run for 109,700 training steps.The models are saved after every 2000 steps and the best model is chosen from them by evaluating them manually. It is later used for testing.

In the following images, Model A discriminates Google Maps images and Model B discriminates satellite images. The Model A_to_B generator generates Google Maps images for a given Satellite image and

Model B_to_A generates Maps images from Satellite images.

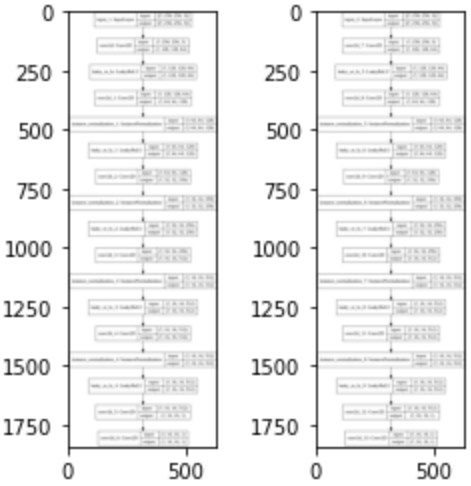

Image 3: Discriminators for Model A and Model B

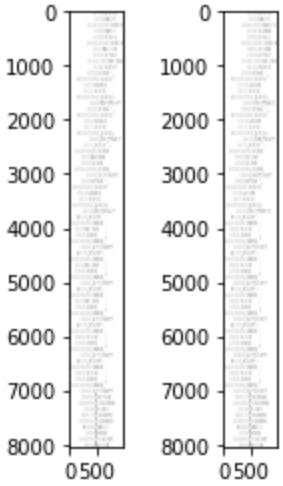

Image 4: Generators for Model A_to_B and Model B_to_A

TensorBoard visualisations: tensorboard has been set up to visualise the loss function during the training period for all the 109,700 steps.

Image 5: d_loss1 for all steps

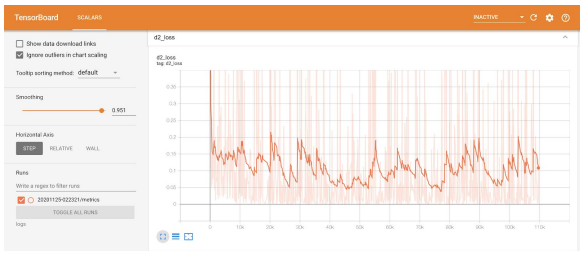

Image 6: d_loss2 for all steps

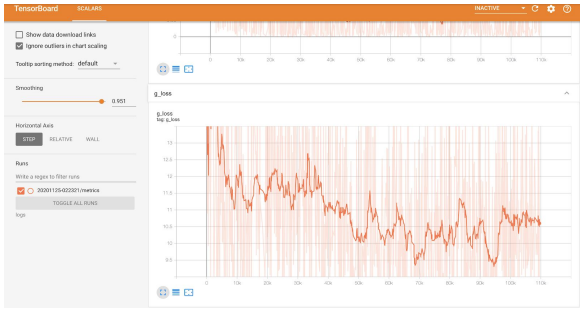

Image 7: g_loss

Experiments:

The architectural setup and training were done using keras. The hyperparameters like learning rate, number of epochs, generator loss, discriminator loss, composite loss have been fine tuned to various values to understand the optimal thresholds to obtain best results.

Exp No.	No. of Epochs	Learning Rate	Generator Loss	Discriminator Loss	Composite Loss
1	100	0.0002	0.002	0.009	8.06
2	100	0.001	0.02	0.013	13.512
3	150	0.0002	0.0002	0.070	11.024

Image 8: Hyperparameter tuning and experiments

Deployment:
A User Interface has been created to generate the translated images based on our trained model. The trained model has been deployed on Amazon EC2. Also, the best performing models of both the Pix2pix and cycleGAN have been saved onto the Amazon S3 bucket. Alongside, DockerHub and git have been part of the deployment pipeline.

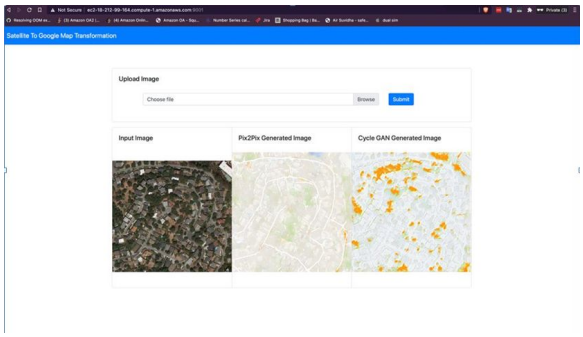

Image 9: User Interface of the application

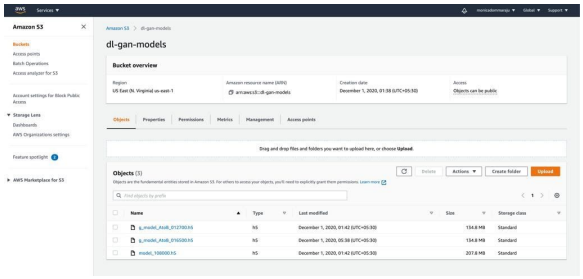

Image 10: Saving best model on Amazon S3

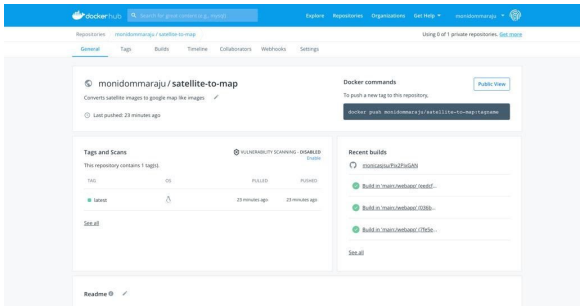

Image 11: DockerHub pipeline

Conclusion: Hence,Both the cycleGAN and Pix2Pix GAN methodologies have been explored to perform image to image translations. As an end result it was understood that the cycleGAN is computationally more expensive than generic GAN networks.The advantage of cycle GAN's is that the images can be interconverted i.e. from satellite images to Google maps images and vice versa. More training time with increased dataset size can possibly generate more accurate results.

References:
[1]Image-to-Image Translation with Conditional Adversarial Networks
[2] Datasource: https://phillipi.github.io/pix2pix/