Generate Synthetic Satellite Images Using GANs

Name: Udaybhan Rathore
Department: DSE

Email : udaybhan19@iiserb.ac.in

Roll Number: 19328

Name: Tanuj Singh Shekhawat
Department: DSE
Email : tanuj19@iiserb.ac.in

Roll Number: 19323

Name : Manoj Kumar
Department: DSE
Email : manoj19@iiserb.ac.in

Roll Number: 19188

Abstract --

Generative Adversarial Networks (GAN) have been used for both image generation and image style translation. For the image generation, we take advantage of the GAN training methodology, that is purposely modified to generate synthetic satellite images. The generated images that we obtained imitate closely the spectral signatures of the kind of terrain in the images, as it can be seen by comparing between synthetic and original images. Generating images from the given sample data is one of the primary applications of recent conditional generative models(cGAN). Besides testing our ability to model conditional, highly dimensional distributions, map to satellite image synthesis has many exciting and practical applications such as synthesizing photos from label maps, Turning Google Maps photos into aerial images etc. Recent progress has been made using Generative Adversarial Networks (GANs). This material starts with a gentle introduction to these topics.

Introduction --

In the field of GAN, earlier the researchers were using traditional GANs, which was not able to reach a good level of accuracy in predicting similar images. Thus, we propose the pix-2-pix model which was introduced by Isola et al. the Pix2Pix GAN is a general approach for image-to-image translation. It is based on the conditional generative adversarial network(cGAN), where a target image is generated, conditional on a given input image.

The study therefore suggests the application of GANs in satellite image resolution improvement. In simple words, offer the possibility of creating a plausible real outcome, an image, by following a set of predefined rules. Those rules can be constructed in a form of a map image in which different areas denote distinct classes. Conditional GANs can interpret these constraints to produce an image that not only meets those criteria but also is realistic.

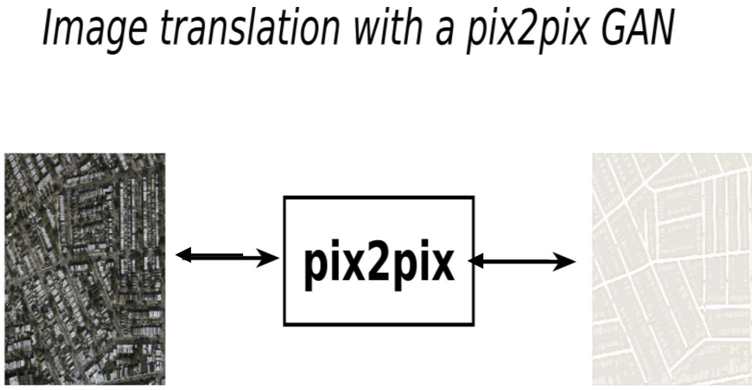


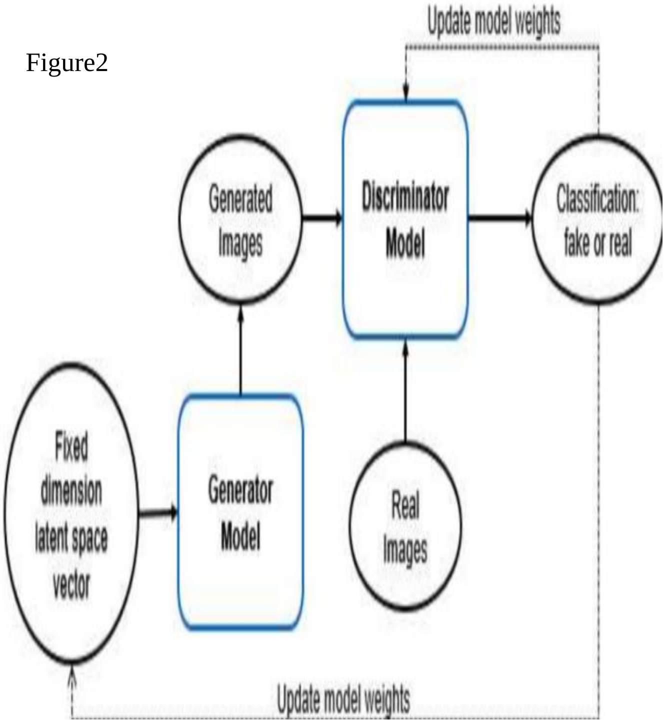
Figure1

Background --

Ever since GANs were discovered by Goodfellow et al, it has been used for several purposes such as data augmentation, image generation from scratch, style transfer, and classification among other purposes. The standard GAN architecture, in Fig.1, for image generation is made up of two convolutional neural networks: the generator is trained to produce images that are similar in distribution to the images used for training, the discriminator that classifies if the images are real (original images) or fake (generated). The two networks are trained together in a minimax fashion. The parameters of the generator and discriminator are updated iteratively and alternatively. Specifically, first, the discriminator is trained for one or more steps(epochs), then the generator is trained for one or more epochs, and the process is repeated. The generator is kept constant during the discriminator training phase. Similarly, the discriminator is kept constant during the generator training phase. As the generator improves with training, the discriminator performance gets worse because the

discriminator cannot easily tell the difference between real and fake. If the generator succeeds perfectly, then the discriminator has a 50% accuracy, reaching convergence.

Figure2



Despite their success, GAN models are often hard to train. This limitation becomes more evident when the images to be generated have high resolution since it is easier for the discriminator to differentiate between generated and real images.

A lot of studies were made to find methods that stabilize the training. One of those approaches aimed to stabilize the training is the conditional GAN (cGAN) methodology.

The cGAN architecture is considered in this paper for synthetic satellite image generation. Depending on the task for which they are employed, many adaptations to the standard GAN architecture have been considered. the image-to-image translation is initially considered conditional GANs, which is an extension of the standard GAN embedding the class label with the input to the generator and the image input to the discriminator.

The objective of image-to-image translation is to modify the content of input to have a similar style to that of a target domain. The training images are made up of pairs, one corresponding to the source domain and the second to the target domain. This is the approach adopted by pix2pix, which is the first attempt to design a GAN. This approach limits the use of GANs due to the difficulty to create a large enough training dataset of such paired images (for each image from the first domain must have a corresponding image in the second domain, i.e. image pairing).. The two components are modeled as GANs, so they are composed of a generator, which performs the mapping, and a discriminator. A so-called cycle-consistency loss is added to the main adversarial loss, with the aim of preserving key attributes between the input image and the generated image.

Material and Methods --

The Pix2Pix Generative Adversarial Network, or GAN, is an approach to training a deep convolutional neural network for image-to-image translation tasks.

Dataset:

In this work, data has been extracted from the software called SASPLANET which consists of two types of image format. The 1st one is the map image and 2nd one is the satellite image corresponding to the map image. This data is of Mumbai city and its outside area. The size of the Mumbai city image is 7424*7424 pixels and the size of the outside area of Mumbai is 3584*3584. Now both types of images of Mumbai city images are split into a total of 841 images of size 256*256 & both types of images of Mumbai city outside area images are split into 196 images of size 256*256. So there is a total of 1037 map images and 1037 satellite images corresponding to map images and the size of each image is 256*256. Now we add map images to their corresponding satellite images using coding with python. Finally, the data we have is consists of 1037 images of a size of 256*512 pixels. Then this data is split into images for training and test of our model. Total images for the training dataset and test dataset are 863 and 174 respectively.

Building model:

The Pix2Pix GAN architecture involves the careful specification of a generator model, discriminator model, and model optimization procedure.

Both the generator and discriminator models use standard Convolution-Batch Normalization-ReLU blocks of layers as is common for deep convolutional neural networks.

U-Net Generator Model:

The generator model takes an image as input, and unlike a traditional GAN model, it does not take a point from the latent space as input.

Instead, the source of randomness comes from the use of dropout layers that are used both during training and when a prediction is made.

A U-Net model architecture is used for the generator, instead of the common encoder-decoder model.

The encoder-decoder generator architecture involves taking an image as input and down sampling it over a

few layers until a bottleneck layer, where the representation is then up sampled again over a few layers before outputting the final image with the desired size.

The U-Net model architecture is very similar in that it involves down sampling to a bottleneck and up sampling again to an output image, but links or skip-connections are made between layers of the same size in the encoder and the decoder, allowing the bottleneck to be circumvented.

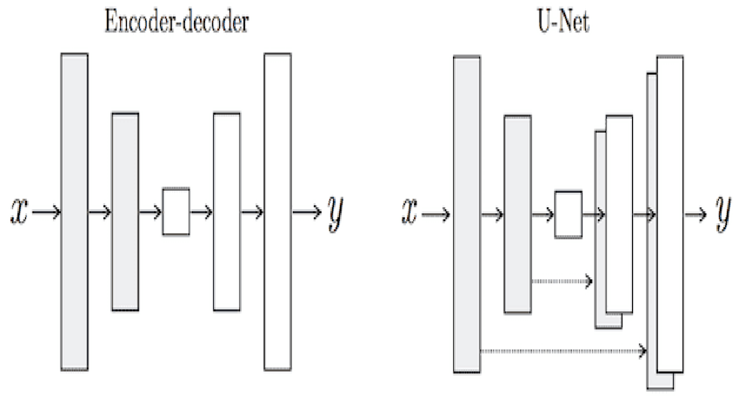


Figure3

Build the generator:

First we build the discriminator and generator , to build the generator we have to define the down sampler and up sampler .As in the figure1 , the generator take image of shape (256,256,3) and it goes to input layer , this input image passes through multiple sequential layers , after each sequential layer shape dimension of image become half ,when this size become 1*1 during this process , up sample function start , up sample function consist of two type of layer first one is sequential and the second is concatenate layers .

So the 1*1 size image pass through both sequential and concatenate layer. after each sequential layer size of image become double and the final layer is convolution 2d transpose. after the whole process the size of final image is equal to the input image and this is the image generated by generator . and the activation function use is ReLU

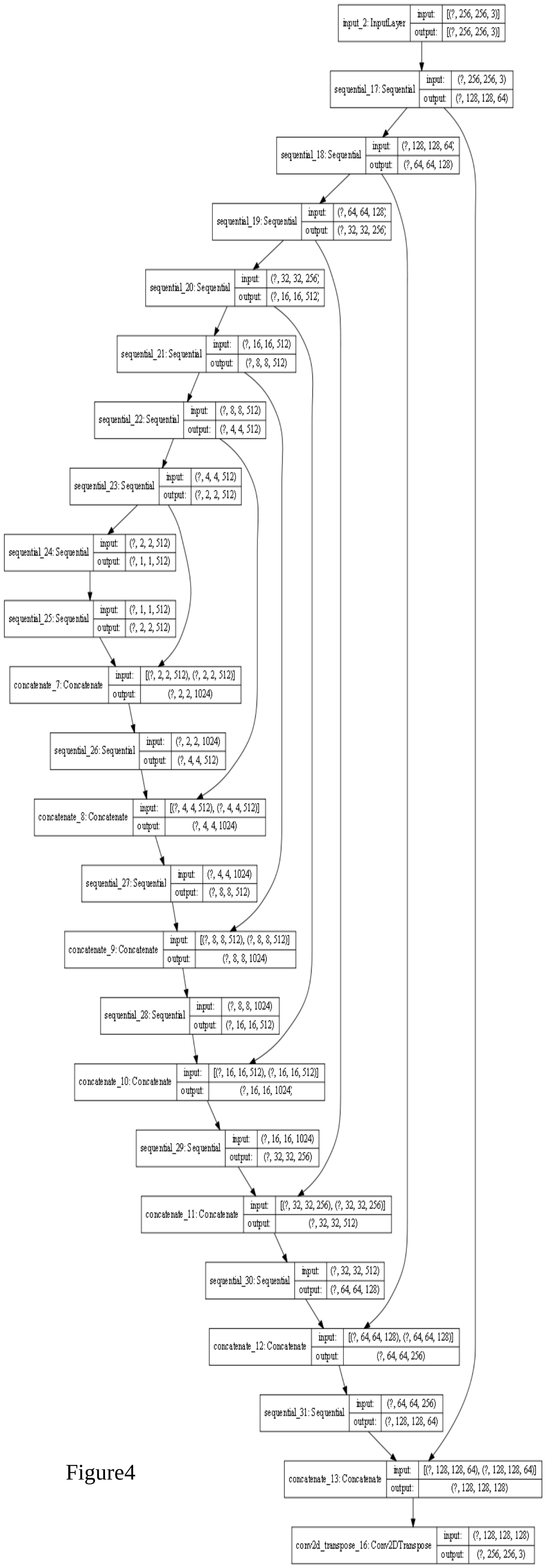


Figure4

Build the Discriminator:

ow we build the discriminator, as you can see in the figure3, discriminator take two images as input, the first one is image generated by generator and second one is target satellite image. now these two images are concatenated by the concatenate layer & the output by the concatenate layer goes to sequential layers and after each sequential layer size of each image become half.

After that, this image pass through zero padding 2D, convolution 2D, leaky ReLU , zero padding, & convolutional 2D layers and the final outcome become an image of (30,30,1) .

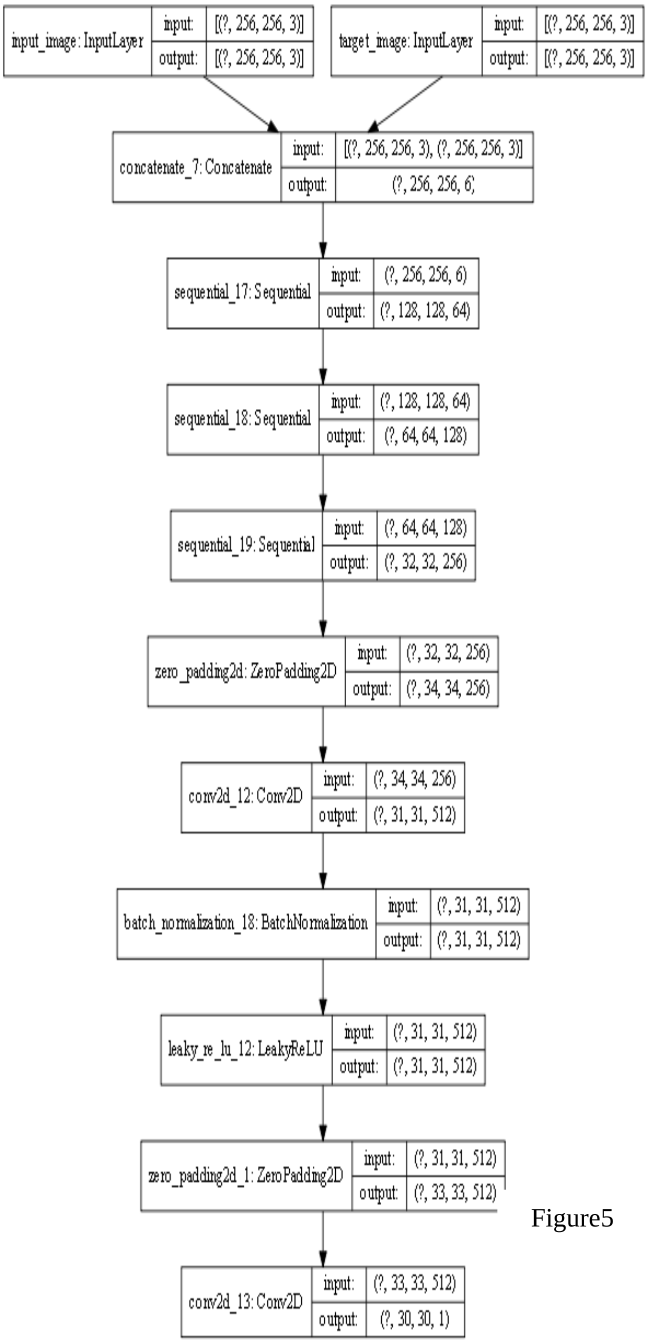


Figure5

Discriminator network decides whether the data is generated or taken from the real sample using a binary classification problem with the help of a sigmoid function that gives the output in the form or the range o & 1. Now we are going to pass reconstructed data and original data to our discriminator and this will provide us a single number and the single number will tell the probability of the input belonging to the original data. Discriminator will try to maximize the chances of predicting correct images but Generator will try to fool Discriminator. Here we use epochs to train our Generator perfectly so that error by Discriminator will be maximum and efficiency of Generator become high.

When Discriminator finds it difficult to detect fake satellite images from the real images then will get high resolution image generated by Generator. The training continues until the Discriminator finds it difficult to detect fake satellite images from the real images.

Finally, model is ready to train and we train the model with 40000 steps instead of epochs because the data that has been pass through the model is very large.

GAN LOSS:

This version of GAN is used to learn a multimodal model. It basically generates descriptive labels which are the attributes associated with the particular image that was not part of the original training data.

CGANs are mainly employed in image labelling, where both the generator and the discriminator are fed with some extra information y which works as an auxiliary information, such as class labels from or data associated with different modalities.

The conditioning is usually done by feeding the information y into both the discriminator and the generator, as an additional input layer to it.

The following modified loss function plays the same min-max game as in the Standard GAN Loss function. The only difference between them is that a conditional probability is used for both the generator and the discriminator, instead of the regular one.

$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x}|\boldsymbol{y})] + \mathbb{E}_{\boldsymbol{z}\sim p_{z}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z}|\boldsymbol{y})))]$$

Why conditional probability? Because we are feeding in some auxiliary information (the green points), which helps in making it a multimodal model, as shown in the diagram below:

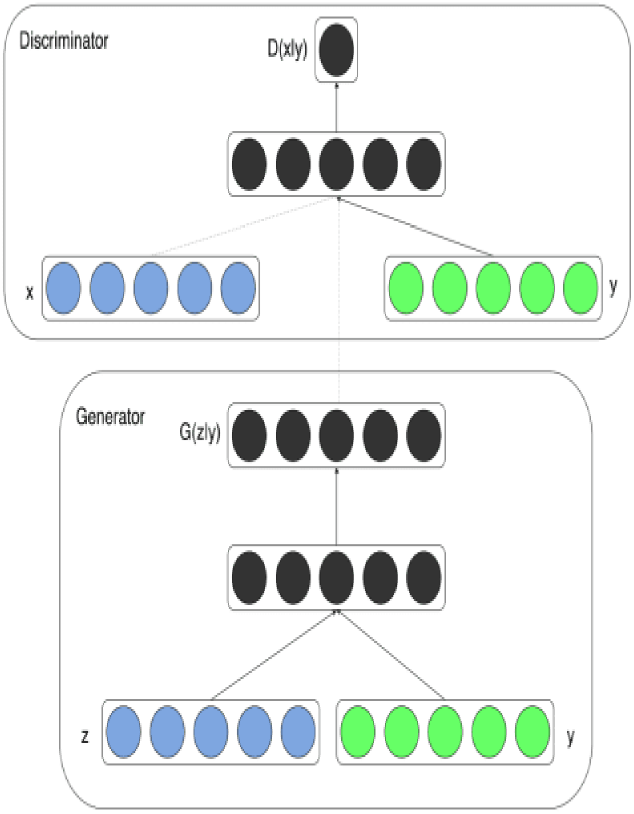


Figure6

There are two different types of losses:

1. Total_ generator _loss: gan_loss + lambda * l1_loss,

where LAMBDA = 100

l1_loss = Least Absolute Deviations

2. Total_disc_loss = real_loss + generated_loss

Result—

The results of our study focus on satellite image resolution improvement
When we increase the steps (40,000) of the model, the efficiency of the Generator will increase
In the result, we have provided sample of two images which consist of input image, ground truth & predicted image respectively.
Input image is a map image of data , ground truth image is satellite image of the data & predicted image is the synthetic image generated by our model
Thus after training , the predicted image of satellite image of Mumbai and its outside area , generated by our model is shown in figure no.

DISCUSSION:

[5] Result of our model is not same every time we run our model due to give stochastic nature of the algorithm or evaluation procedure, or differences in numerical precision. Consider running the example a few times and compare the average outcome.
As before, the loss of the model is reported each training iteration. If loss for the discriminator goes to zero and stays there for a long time, consider re-starting the training run as it is an example of a training failure.

Future recommendations:

In future, data from a new generation of the Geostationary and polar orbiting satellites will become available. To prepare for these data, representative imagery of these satellites is desirable. Two methods have been developed to create imagery from future satellites.
One method uses simulated imagery i.e. data from current operational and experimental satellites. Another method uses synthetic imagery (generated by GANs) which is generated by using numerical models.

The combined usage of both the systems together would help us overcome the shortcomings of each system when used individually.

Contribution:

A strategy for using this imagery in the forecasting of severe convective weather is presented. As computing power continues to increase, it is now possible to generate more computationally expensive model-derived fields, such as infrared brightness temperatures and radar reflectivity, and to display these data in a manner consistent with typical satellite and radar visualization methods.

Two important factors responsible for this are availability of aerial photographs and ever improving spatial resolution of satellite images and this model will help in overcoming these problems and gives us the data to analysis the satellite images accurately.

Conclusion --

In this work, we proposed a GAN model for map-to-satellite image conversion.

Considering the challenge that some objects might not be easily identified visually from the map image, we integrate the external geographic data into a GAN structure to guild the conversion. Satellite imaging of the Earth surface is of sufficient public utility that many countries maintain satellite imaging programs. But in the field of GANs in India, We don't have proper data to train over model to get better predicted images than other countries as show below slide.

Therefore, we are not able to predicted satellites images in high resolution than other countries.

```
In [*]: fit(train_dataset, test_dataset, steps=40000)

Time taken for 1000 steps: 293.90 sec
```


```
Step: 27k
................................
```

```
In [ ]: generator_loss(disc_generated_output, gen_output, target)
        discriminator_loss(disc_real_output, disc_generated_output)
```

I. REFERENCES

[1]: Ian J. Goodfellow, J. Pouget-Abadie, M. Mirza, and Others, Generative Adversarial Networks (2014).
[2]Kodali, N., Abernethy, J., Hays, J., Kir Z., (2017), "On Convergence and Stability of GANs", abs/1705.07215
[3] Radford A., Metz L., Chintala S., "Unsupervised representation learning with deep convolutional generative adversarial networks", CoRR, 1511.06434, 2015.
[4]: J. Langr, V.Bok, GANs in Action: Deep learning with Generative Adversarial Networks (2019), Manning Publications

[5] ***Jason Brownlee*** **PhD, https://machinelearningmastery.com/different-results-each-time-in-machine-learning/**