EDIFICE DETECTION OF DOMAIN TRANSLATED SATELLITE IMAGERY USING GENERATIVE ADVERSARIAL NETWORKS

CS6611 – CREATIVE AND INNOVATIVE PROJECT

TEAM NO: 22

GOKUL. S 2018103026

SRIHARI. S 2018103601

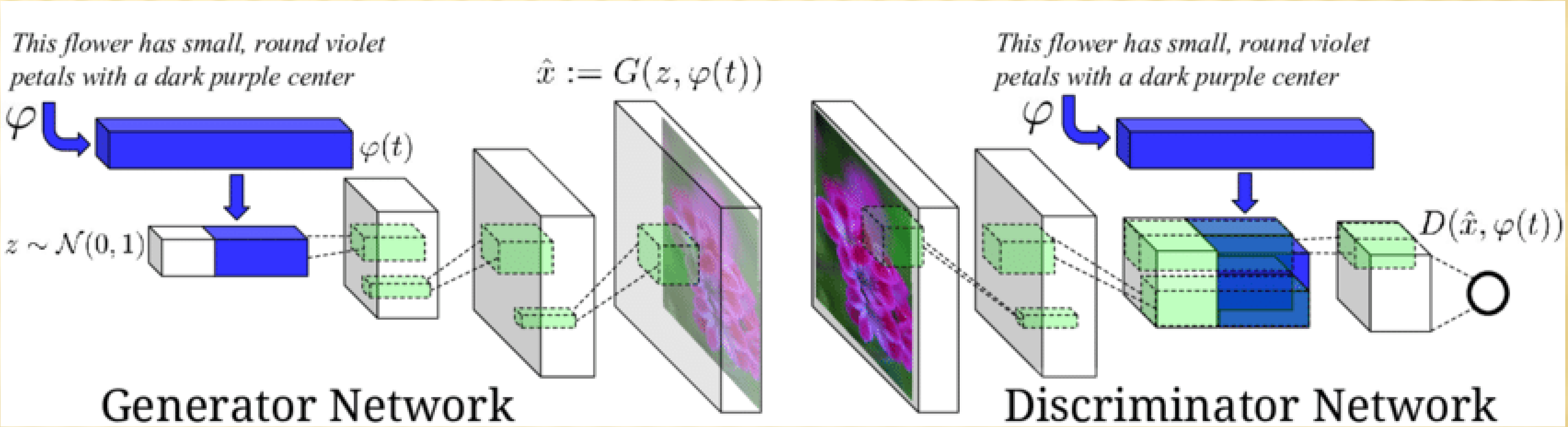
INTRODUCTION

PROBLEM SCOPE

- Image-to-image translation is to learn a mapping between images from a source domain and images from a target domain.
- After much studies, it has been found that more than about 98% of the land, environment and mineral composition/structure of the universe still remains completely unexplored to date.
- This is still an area where the rate of growth in our existing technology remains dull.
- This problem is being attempted to be solved from the perspective of image domain translation since intricate details and information would start to become evident along the way.
- In order to achieve this objective efficiently, we make use of the state-of-the-art Generative Adversarial Networks framework, in particular, the Pix2Pix GAN.
- This network also comes in handy for intelligence services and military forces to secret detect trenches and attack spots.

OBJECTIVE

- Given an input satellite image from a military satellite, the model generates a plausible aerial map version of the image using the latent sample space that it has learned during the training.
- This is further given as input to the edifice detection module which detects the hidden trenches and edifices present in the image.
- This can be passed on to the terrestrial military forces by the secret services to facilitate efficient attacks.
- This objective is achieved using the **Conditional Generative Adversarial Network** paradigm.



RELATED WORKS

Author/Publication Year	Methodology	Advantages	Limitations
Feng xiong, qianqian wang , and quanxue gao – cegan-2019	• Proposed an image-to-image translation model by combining GAN and latent space learning to arrive at Consistent Embedded Generative Adversarial Networks generating both realistic and diverse images. • It captures the full distribution of potential multiple modes of results by enforcing tight connections in both the real image space and latent space.	• To achieve realism, the discriminator in this model distinguishes the real images and fake images in the latent space to alleviate the impact of the redundancy and noise in generated images. • The ability of the model to produce more diverse and realistic results is strengthened by learning a low-dimensional latent code due to the multiple distribution in the latent space.	• Focused on generating a single result conditioned on the input and assumed a deterministic or unimodal mapping. • Failed to capture the full distribution of possible outputs. Evaluation based on AMT – Amazon Mechanical Turk Perpetual Study is subjective to human errors.
Taewon Kang, Kwang Hee Lee - Unsupervised Image-to-Image Translation with Self-Attention Networks 2020	• Adopted the self-attention network thus producing better results than the convolution-based GAN. • Verified the effectiveness of the self-attention network in unsupervised image-to-image translation tasks. • Accomplishes the transformation from a source domain to a target domain given unpaired training data.	• Long range dependency helps to not only capture strong geometric change but also generate details using cues from all feature locations. • The discriminator can accurately enforce complicated geometric constraints on the global image structure • Has overcome the limitations of CycleGAN, the most representative unsupervised image translation method, by changing the high-level semantic meaning.	• It fails to capture strong geometric changes between domains, or it produces unsatisfactory results for complex scenes, compared to local texture mapping tasks such as style transfer.

RELATED WORKS

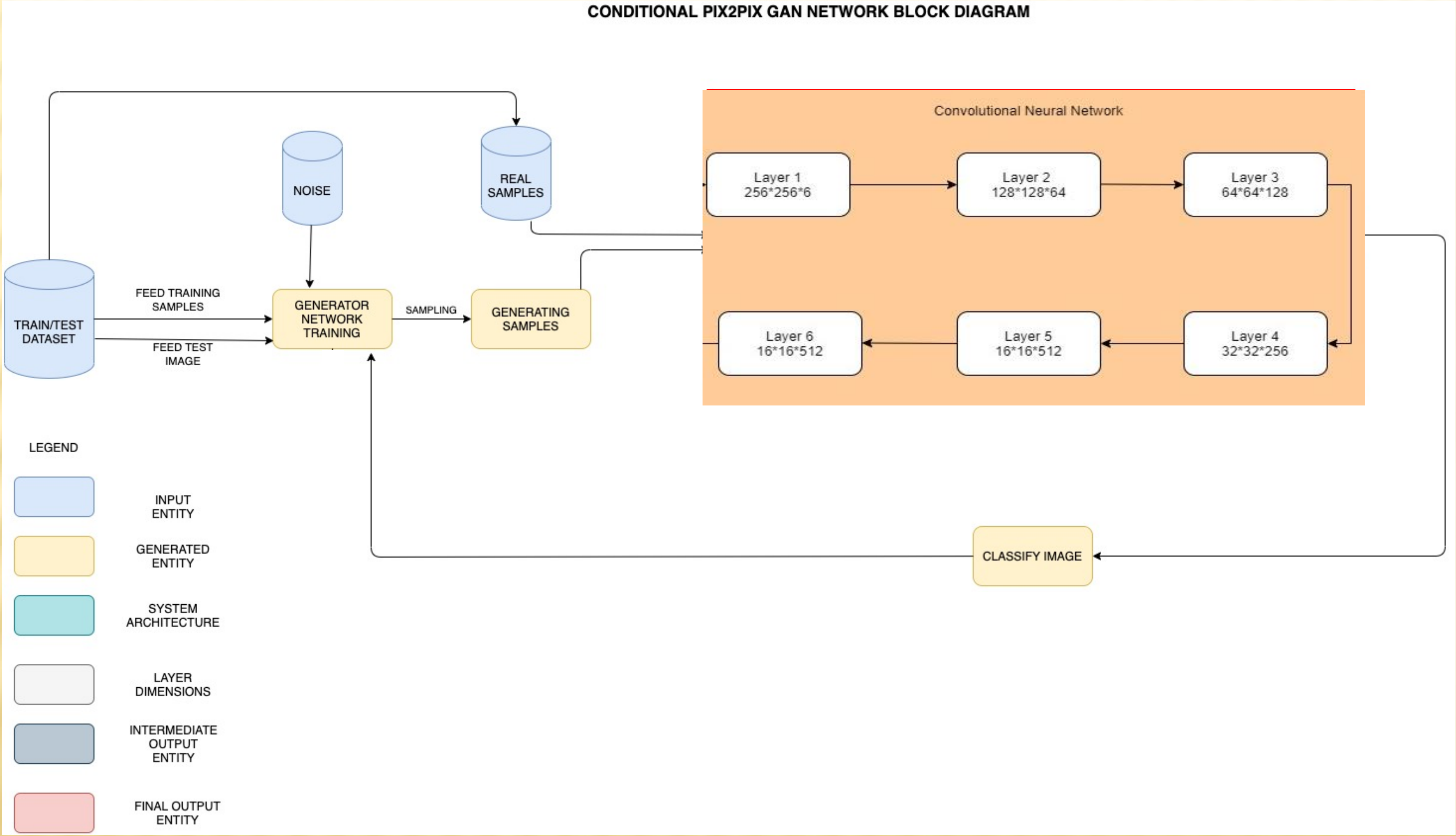
Author/Publication Year	Methodology	Advantages	Limitations
Hajar Emami, Majid Moradi Aliabadi, Ming Dong, and Ratna Babu Chinnam - SPA-GAN: Spatial Attention GAN for Image-to-Image Translation - 2019	• Incorporated the attention mechanism into image-to-image translations helping the generative network to attend to the regions of interest and produce more realistic images. • Transferred the knowledge from the discriminator to the generator to force it focus on the discriminative areas of the source and the target domains. • Attention maps are also generated which were looped back to the input of the generator.	• The discriminator network is deployed to highlight the most discriminative regions between real and fake images in addition to the classification. These discriminative regions illustrate the areas where the discriminator focusses on in order to correctly classify the input image. • Integrates the adversarial, modified cycle consistency and feature map losses to generate more realistic outputs.	• Most data is inherently non – IID and most assumptions of this model are too simplistic like the Euclidean Distance. • Struggles with modelling spatial complexities like interplay of short versus long-distance dependencies.
Zili Yi1, Hao Zhang , Ping Tan , and Minglun Gong - DualGAN: Unsupervised Dual Learning for Image-to-Image Translation – 2017	• Uses the loss format advocated by Wasserstein GAN rather than the sigmoid cross-entropy loss used in the original GAN. • L1 distance is adopted to measure the recovery error, which is added to the GAN objective to force the translated samples to obey the domain distribution. • DualGAN is constructed with identical network architecture for GA and GB. The generator is configured with equal number of downsampling (pooling) and upsampling layers.	• Performs better in terms of generator convergence and sample quality, as well as in improving the stability of the optimization. • Generator is configured with skip connections between mirrored downsampling and upsampling layers, making it a U-shaped net enabling low-level information to be shared between input and output, thus preventing significant loss of high-frequency information.	• However, DualGAN uses an unsupervised learning approach for image-to-image translation based on unpaired data, but with different loss functions. • Figuring the right combination of these loss functions is a tedious task and wastes a huge amount of time which otherwise could've been used to train the actual GAN model itself.

RELATED WORKS

Author/Publication Year	Methodology	Advantages	Limitations
H. Tang, H. Liu and N. Sebe, "Unified Generative Adversarial Networks for Controllable Image-to-Image Translation," in IEEE Transactions on Image Processing, vol. 29, pp. 8916-8929, 2020, doi: 10.1109/TIP.2020.3021789. - 2021	• The unified GAN model is used, which can be tailored for handling all kinds of problem settings of controllable structure guided image-to-image translation, including object key point guided generative tasks, human skeleton guided generative tasks and semantic map guided generative tasks, etc.	• More effective in solving specific situations mainly if the domains are arbitrarily large. • Covers an universal controllable image translation despite the very limited research.	• Since this model architecture covers a wide range of applications and considers multiple independent aspects of the image generation process, it is extremely general and its performance when it comes to specific image transformation tasks like the ones that we intend to solve isn't the best available.
X. Wang, H. Yan, C. Huo, J. Yu and C. Pant, "Enhancing Pix2Pix for Remote Sensing Image Classification," 2018 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 2332-2336, doi: 10.1109/ICPR.2018.8545870 - 2018	• Approaches the image to image translation problems by adding a controller to Pix2Pix model to improve the classification performance. • The improved Pix2Pix model is composed of three parts: generator, discriminator and controller. • The input to the controller is the remote sensing image classification map, and the output is the reconstructed version based on the classification map obtained by generator (controller is also considered as the reverse mapping of the generator).	• The aim of the controller is to adjust the classification performance by the reconstruction error, which is established on the relation between the classification performance and the smaller the reconstruction error, the better the classification performance will be. • U-net structure is used for the controller due to the advantages of the cross layer connection in preserving the high frequency characteristics of the image. • The optimization objective of the controller is to minimize the reconstruction error.	• It can only generate real images based on simple labels, but complex features are difficult to learn or the image is difficult to match well with the label. • In other words, one important drawback here is the ignorance of explicit constraints between classification performance and reconstruction error.

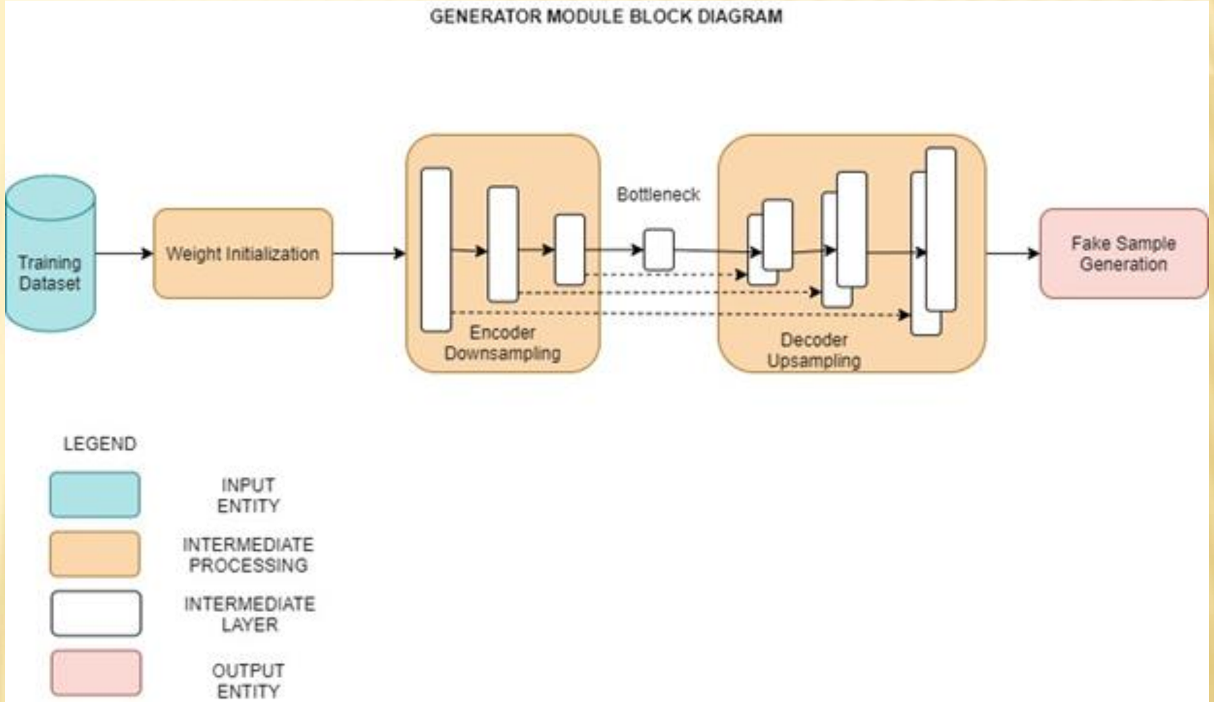
SYSTEM DESIGN

OVERALL BLOCK DIAGRAM



GENERATOR MODULE

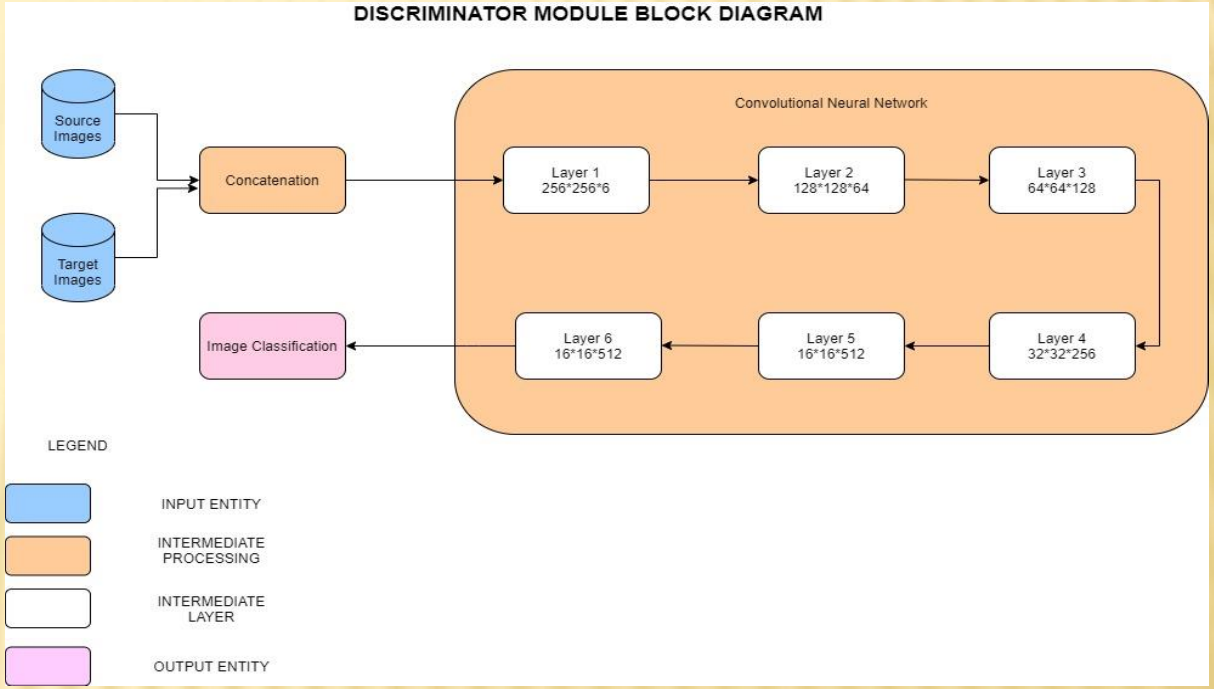
- Generator network takes a real time dataset and generates a sample of data.
- It is an encoder-decoder model using a U-Net architecture.
- Training is done via adversarial loss and the generator is updated via L1 loss measured between the generated image and the expected output.
- It does this by first downsampling or encoding the input image down to a bottleneck layer, then upsampling or decoding the bottleneck representation to the size of the output image.



Input : Training Dataset, Noise for the dataset, Discriminator Predictions
Output : Fake Samples

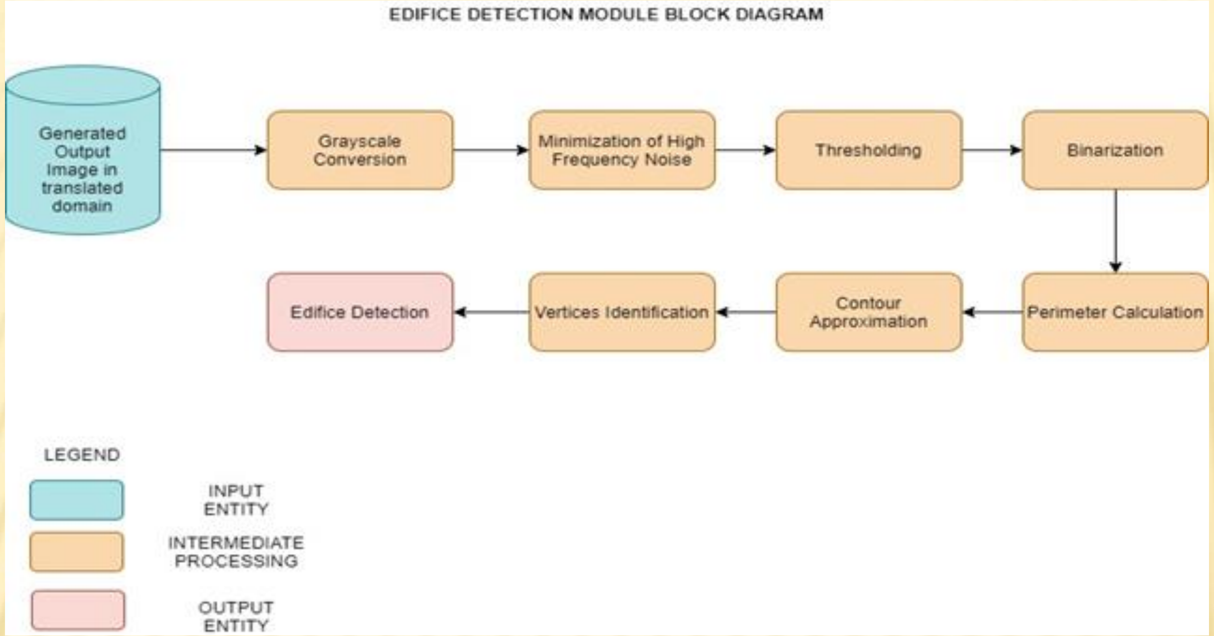
DISCRIMINATOR MODULE

- Discriminator network decides whether the data is generated or directly taken from the real sample using a binary classification problem.
- In adversarial, the model is trained in an adversarial setting and network simply means for the training of the model we use the neural networks as AI algorithms.
- It is optimized using binary cross entropy, and a weighting is used so that updates to the model have half (0.5) the usual effect.



Input : Training Dataset (Real Samples), Generated Samples (Fake Samples)
Output : Binary Classification of Images as Real or Fake

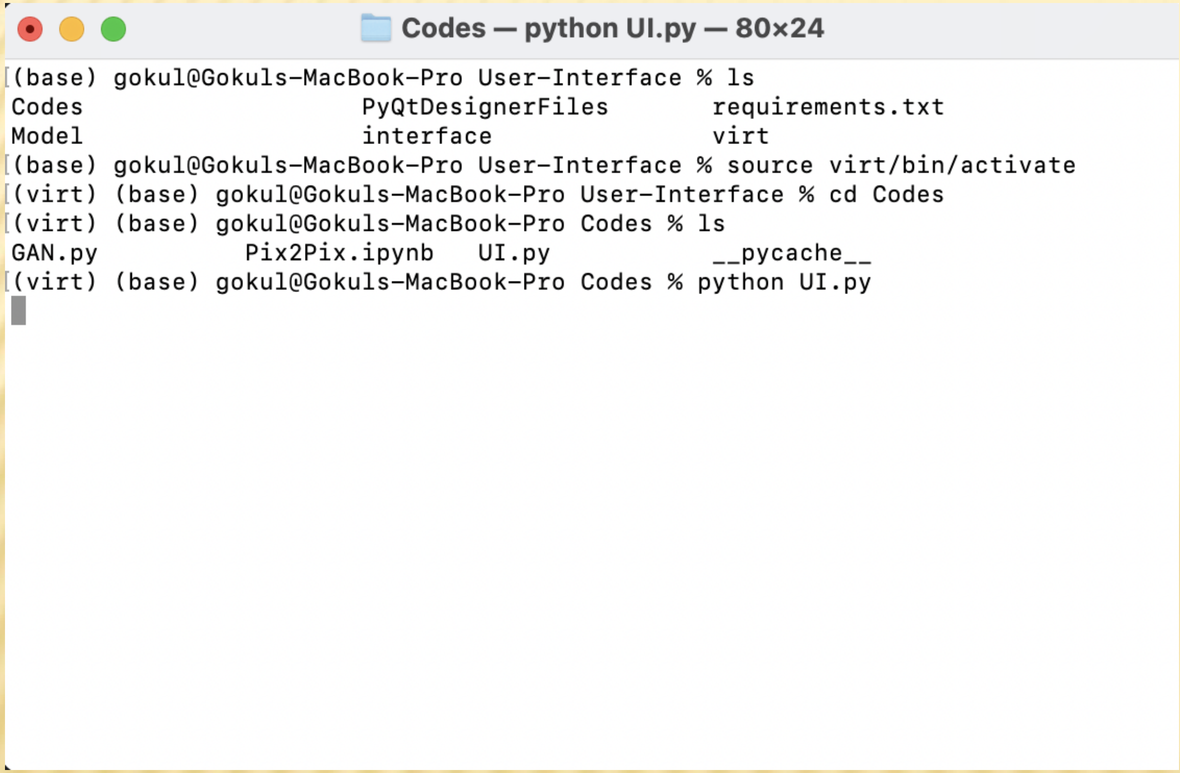
EDIFICE DETECTION MODULE



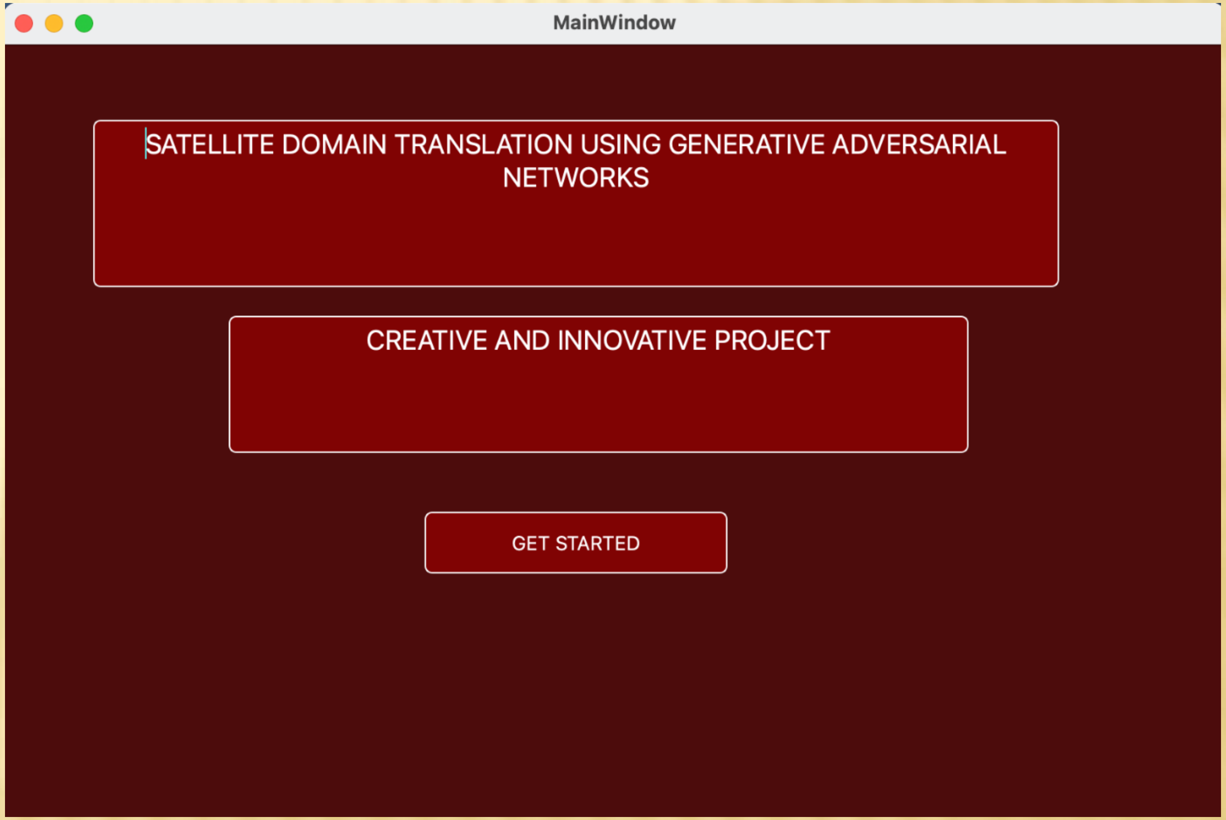
- On obtaining the generated samples, it is visualized with the aid of ApproxPolydp feature enabling us to recognize the polygons in this image.
- The putative points are figured out in the resulting image thereby enabling us to figure out the erections.
- **Input** : Generated Samples in the output domain
- **Output** :Edifice Detection

EXECUTION SNAPSHOTS

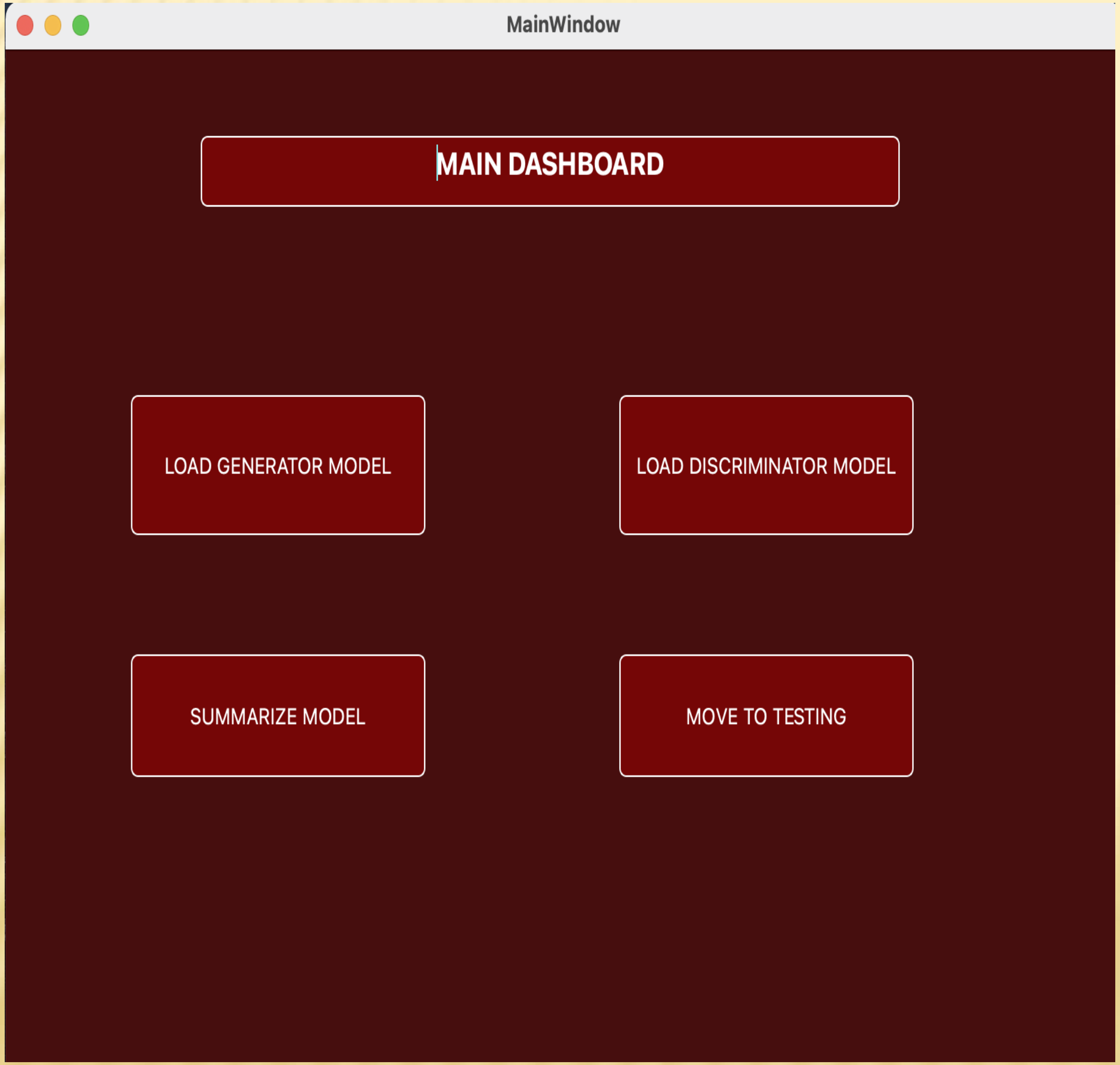
TERMINAL



GUI OPENING WINDOW

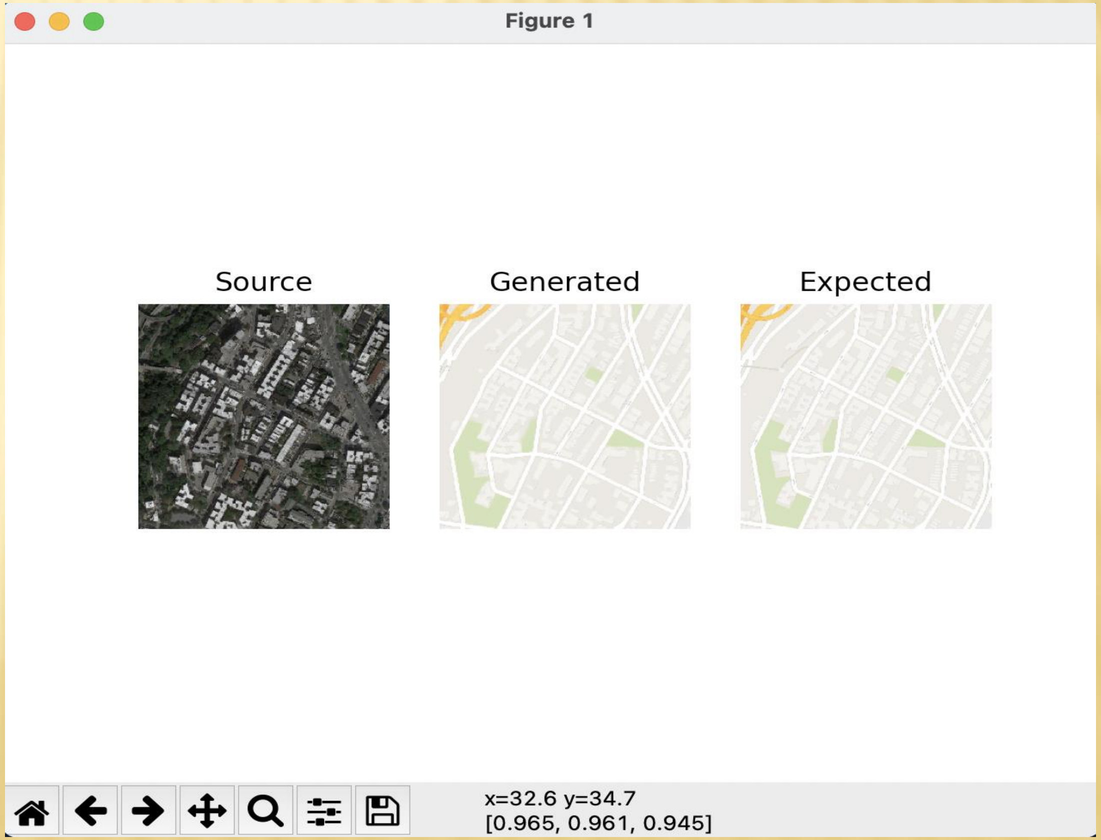


MAIN DASHBOARD

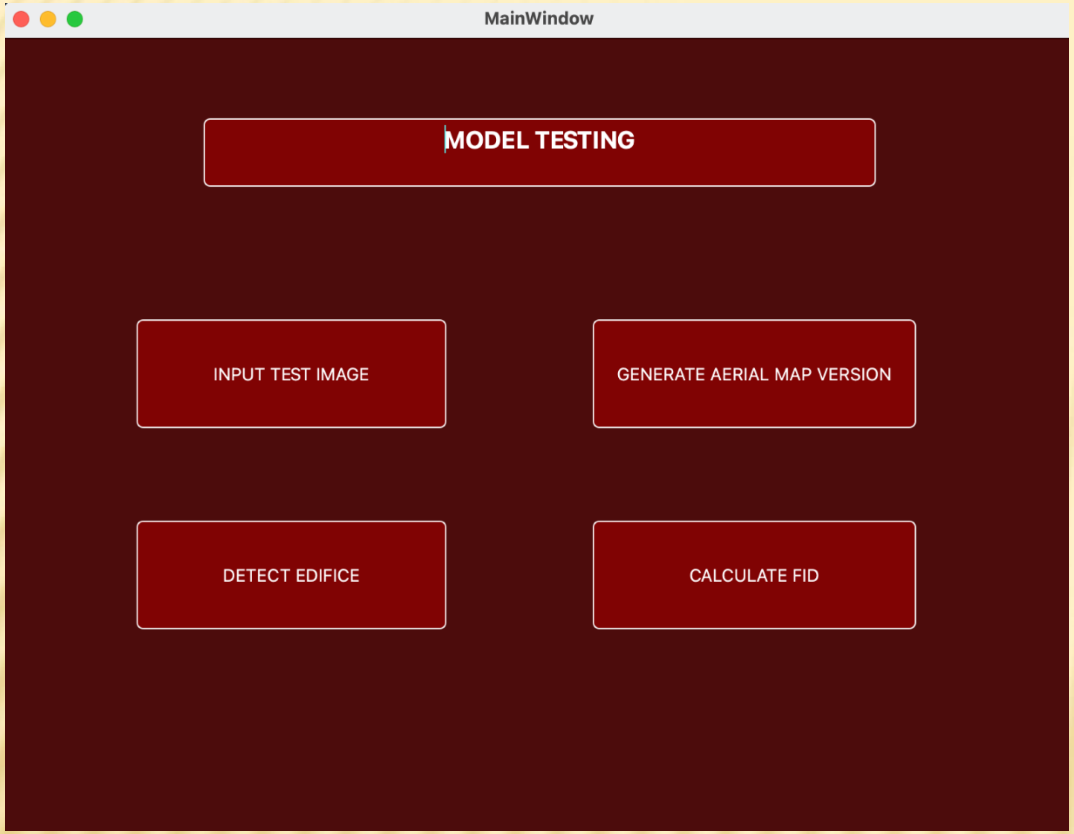


PERFORMANCE SUMMARIZED

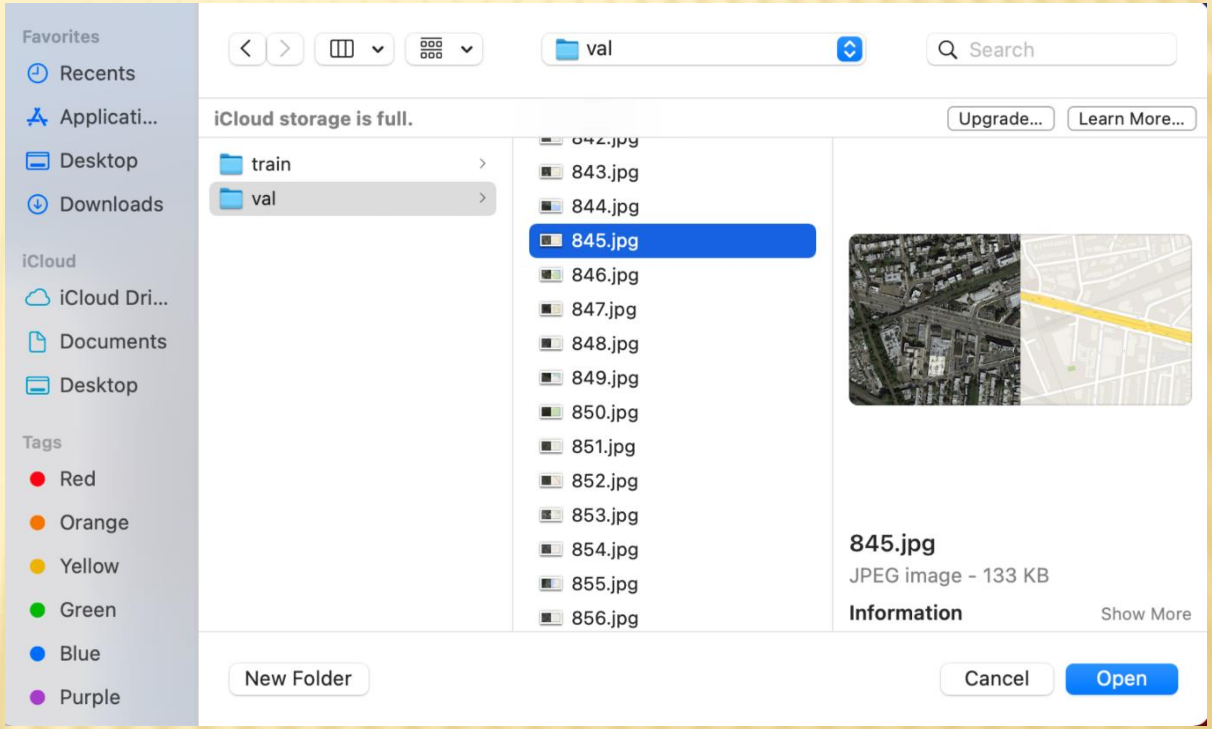
```
GENERATOR MODEL LOADED SUCCESSFULLY..!!
Input =  (None, 256, 256, 3)
Output =  (None, 256, 256, 3)
DISCRIMINATOR MODEL LOADED SUCCESSFULLY..!!
Input =  [(None, 256, 256, 3), (None, 256, 256, 3)]
Output =  (None, 16, 16, 1)
GAN MODEL CREATED SUCCESSFULLY..!!
Input =  (None, 256, 256, 3)
Output =  [(None, 16, 16, 1), (None, 256, 256, 3)]
```



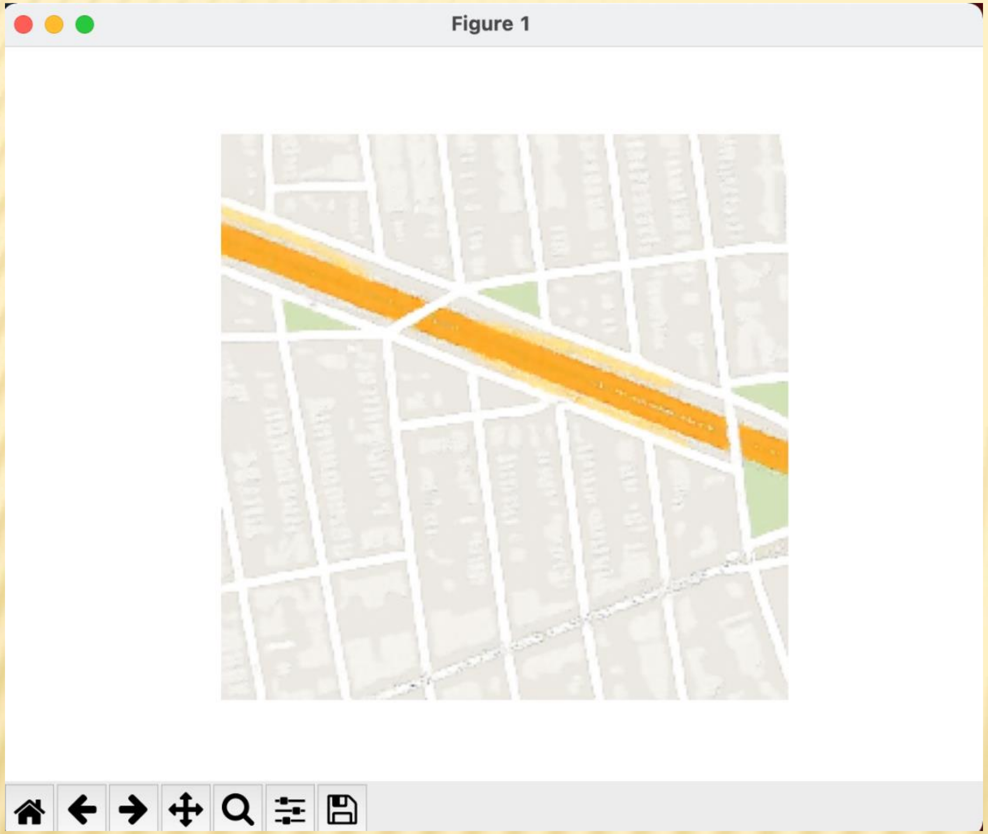
TESTING DASHBOARD



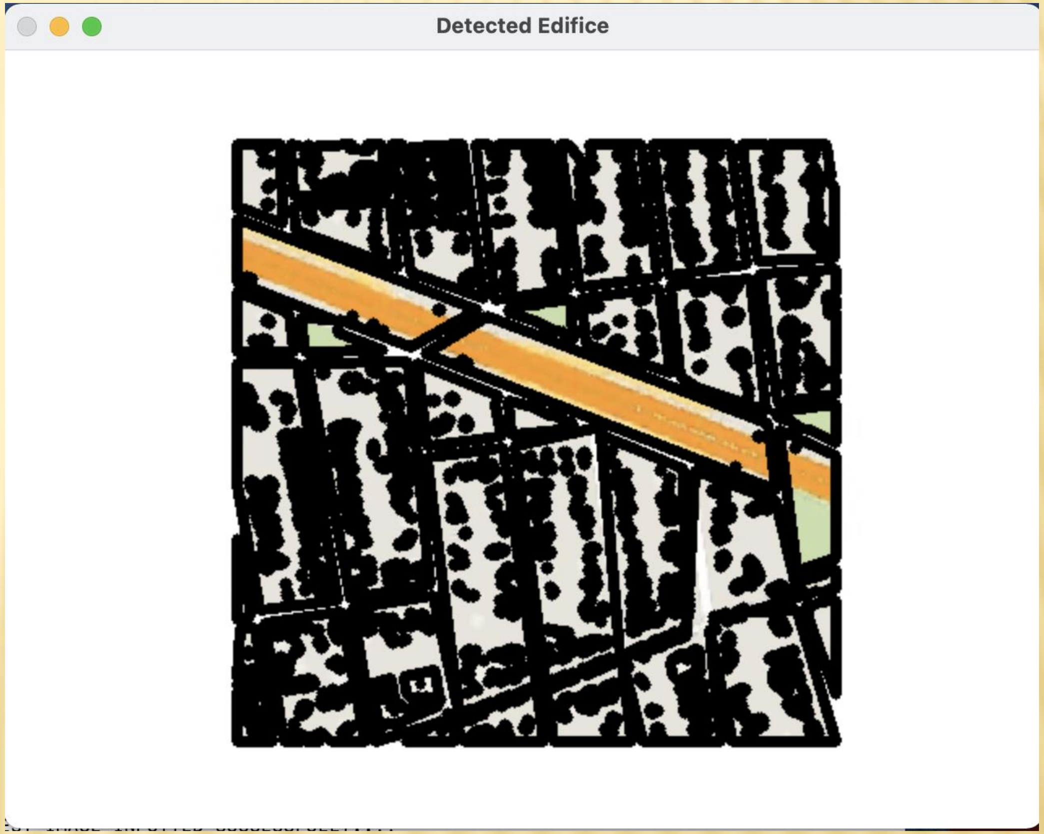
FILE BROWSER



GENERATED AERIAL MAP IMAGE



EDIFICES DETECTED



EXPERIMENTAL RESULTS

LIST OF METRICS USED:

1. **FRECHET INCEPTION DISTANCE**

   - The FID metric is the Wasserstein metric between two multidimensional Gaussian distributions: the distribution of some neural network features of the images generated by the GAN and the distribution of the same neural network features from the "world" or real images used to train the GAN.

$$\mathrm{FID} = |\mu - \mu_w|^2 + \mathrm{tr}(\Sigma + \Sigma_w - 2(\Sigma\Sigma_w)^{1/2})$$

2. **MANHATTAN NORM**

   - The set of vectors whose 1-norm is a given constant forms the surface of a cross polytope of dimension equivalent to that of the norm minus 1. The Manhattan norm is also called the L1 norm. The distance derived from this norm is called the Manhattan distance .

$$d_1(\mathbf{p}, \mathbf{q}) = \|\mathbf{p} - \mathbf{q}\|_1 = \sum_{i=1}^{n} |p_i - q_i|$$

LIST OF METRICS USED:

3. **MEAN SQUARE ERROR**

- The mean squared error function takes the target image and the generated aerial map image and compares the pixel intensities. A value of 0 for MSE indicates perfect similarity. A value greater than one implies less similarity and will continue to grow as the average difference between pixel intensities increases.

$$MSE = \frac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2$$

4. **STRUCTURAL SIMILARITY**

- SSIM attempts to model the perceived change in the structural information of the image.
- It takes as parameters the (x, y) location of the N x N window in each image, the mean of the pixel intensities in the x and y direction, the variance of intensities in the x and y direction, along with the covariance. The SSIM value can vary between -1 and 1, where 1 indicates perfect similarity.

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

EXISTING VS PROPOSED SYSTEM

Evaluation Metric	Existing System	Proposed System
Frechet-Inception Distance	625.25	560.961
Mean Square Error	0.0026	0.0020
Structural Similarity	0.86	0.87
Manhattan Norm	33.68	20.73

ANALYTICAL ABLATION STUDY

EPOCH WISE EVALUATION METRIC VALUES

Epochs	Evaluation Metric	Proposed System
50	Frechet-Inception Distance	625.173
	Mean Square Error	0.0025
	Structural Similarity	0.83
	Manhattan Norm	35.96

Epochs	Evaluation Metric	Proposed System
100	Frechet-Inception Distance	601.253
	Mean Square Error	0.0023
	Structural Similarity	0.84
	Manhattan Norm	30.42

Epochs	Evaluation Metric	Proposed System
150	Frechet-Inception Distance	580.652
	Mean Square Error	0.0022
	Structural Similarity	0.85
	Manhattan Norm	23.83

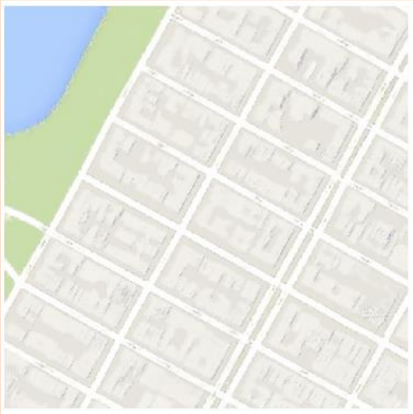
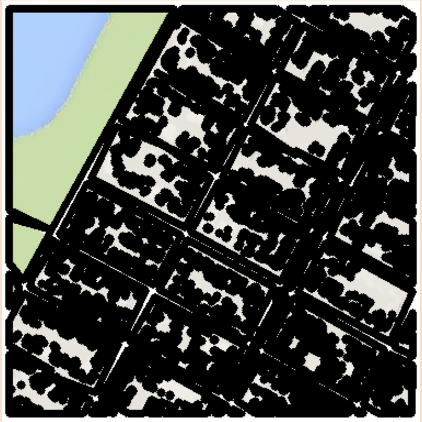
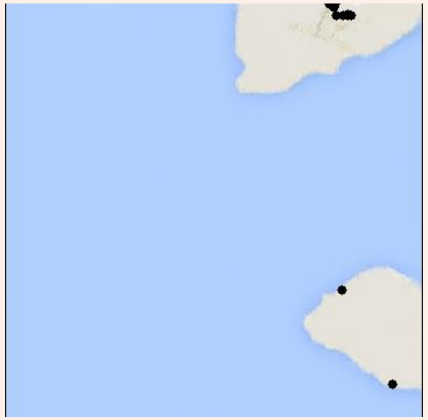
Epochs	Evaluation Metric	Proposed System
200	Frechet-Inception Distance	560.961
	Mean Square Error	0.0020
	Structural Similarity	0.87
	Manhattan Norm	20.73

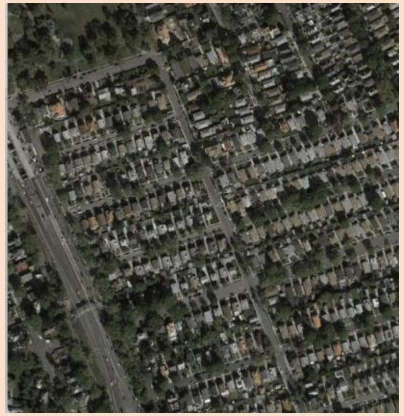
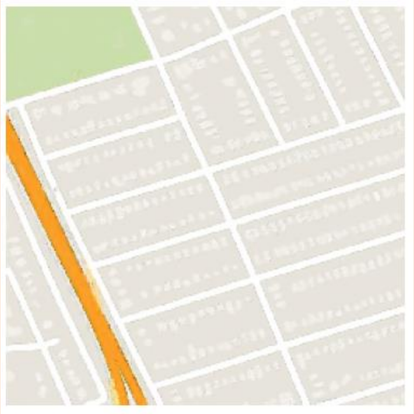
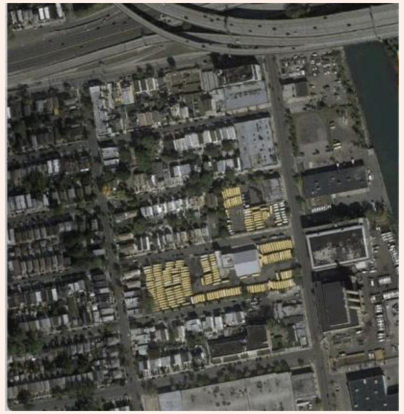
OUTPUT SNAPSHOTS OF METRICS

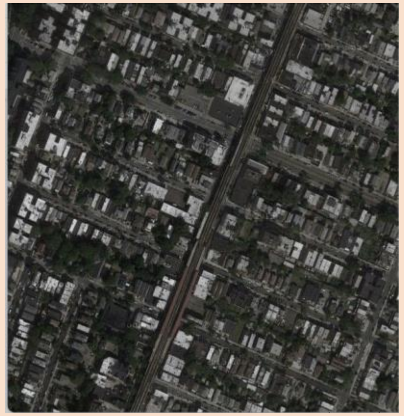
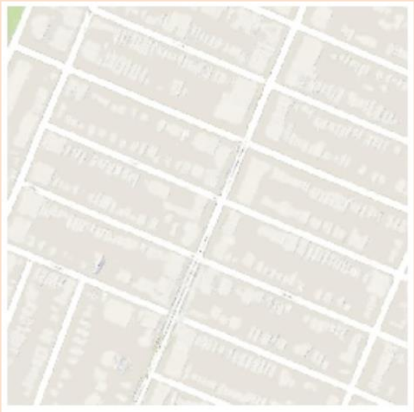
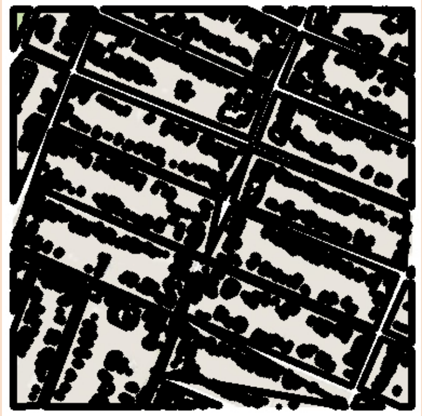
```
[ ] fid = obj1.__call__(tarimg,srcimg,15,128,None,False,None)
    print(fid)

    /usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:6: Warning: Rank deficient covariance matrix, Frechet distance will not be accurate.

    560.9619757832536
```


```
1 main()

Manhattan norm per pixel: 20.73378005324527
```

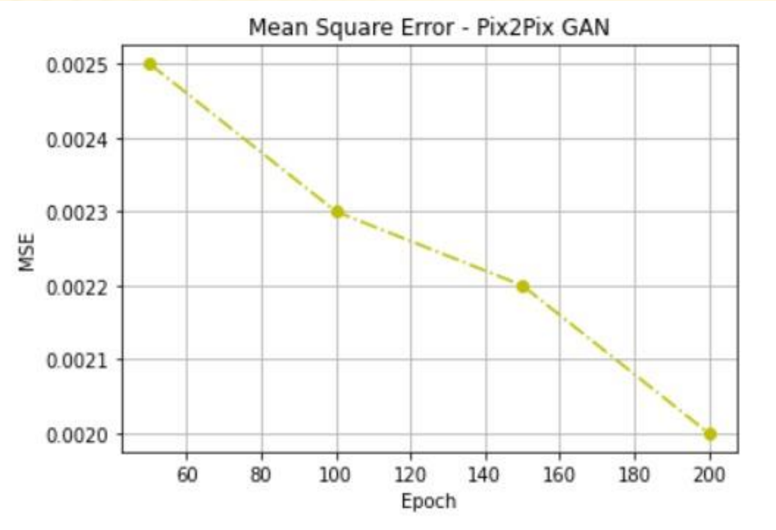
RESULT ANALYSIS

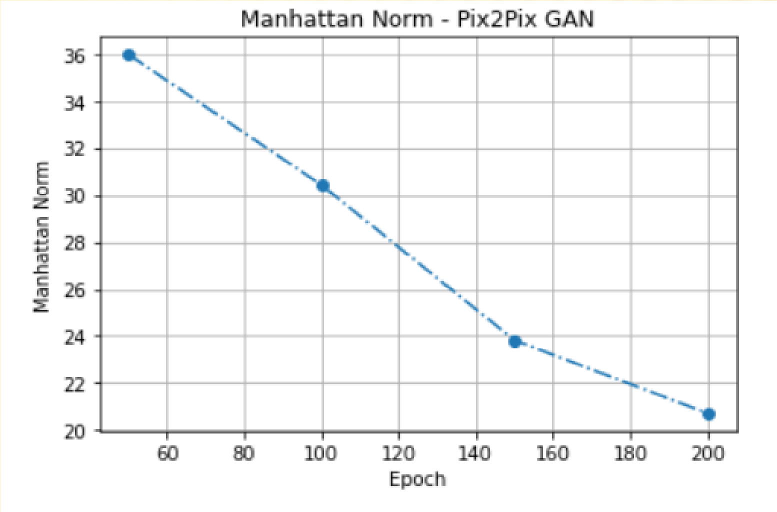
INPUT SATELLITE IMAGE	GENERATED AERIAL MAP IMAGE	OUTPUT IMAGE WITH EDIFICES DETECTED
		
		

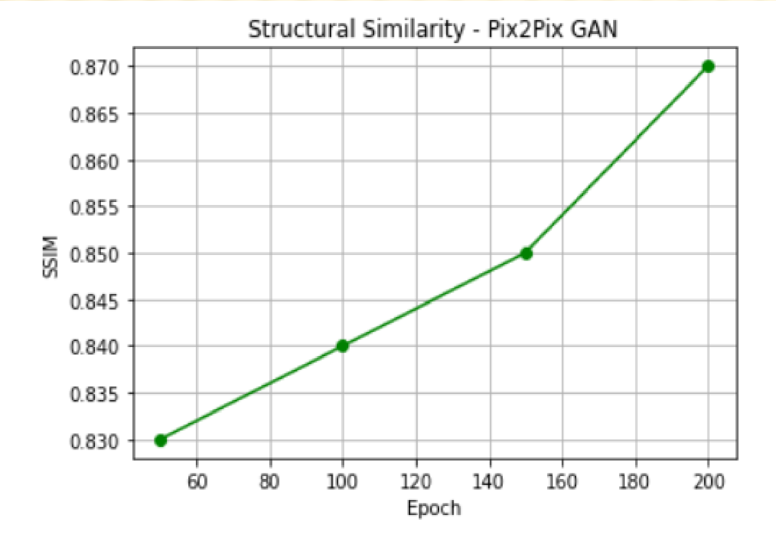
INPUT SATELLITE IMAGE	GENERATED AERIAL MAP IMAGE	OUTPUT IMAGE WITH EDIFICES DETECTED
		
		

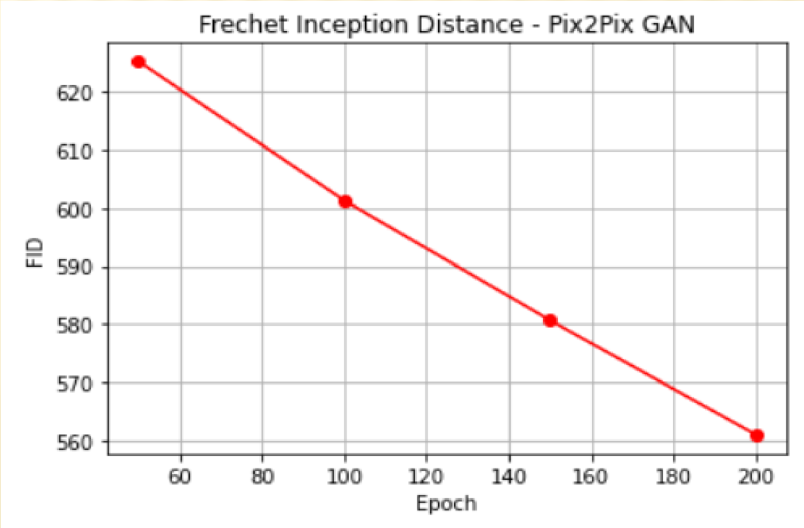
INPUT SATELLITE IMAGE	GENERATED AERIAL MAP IMAGE	OUTPUT IMAGE WITH EDIFICES DETECTED
		
		

GRAPHICAL ANALYSIS OF EVALUATION METRICS

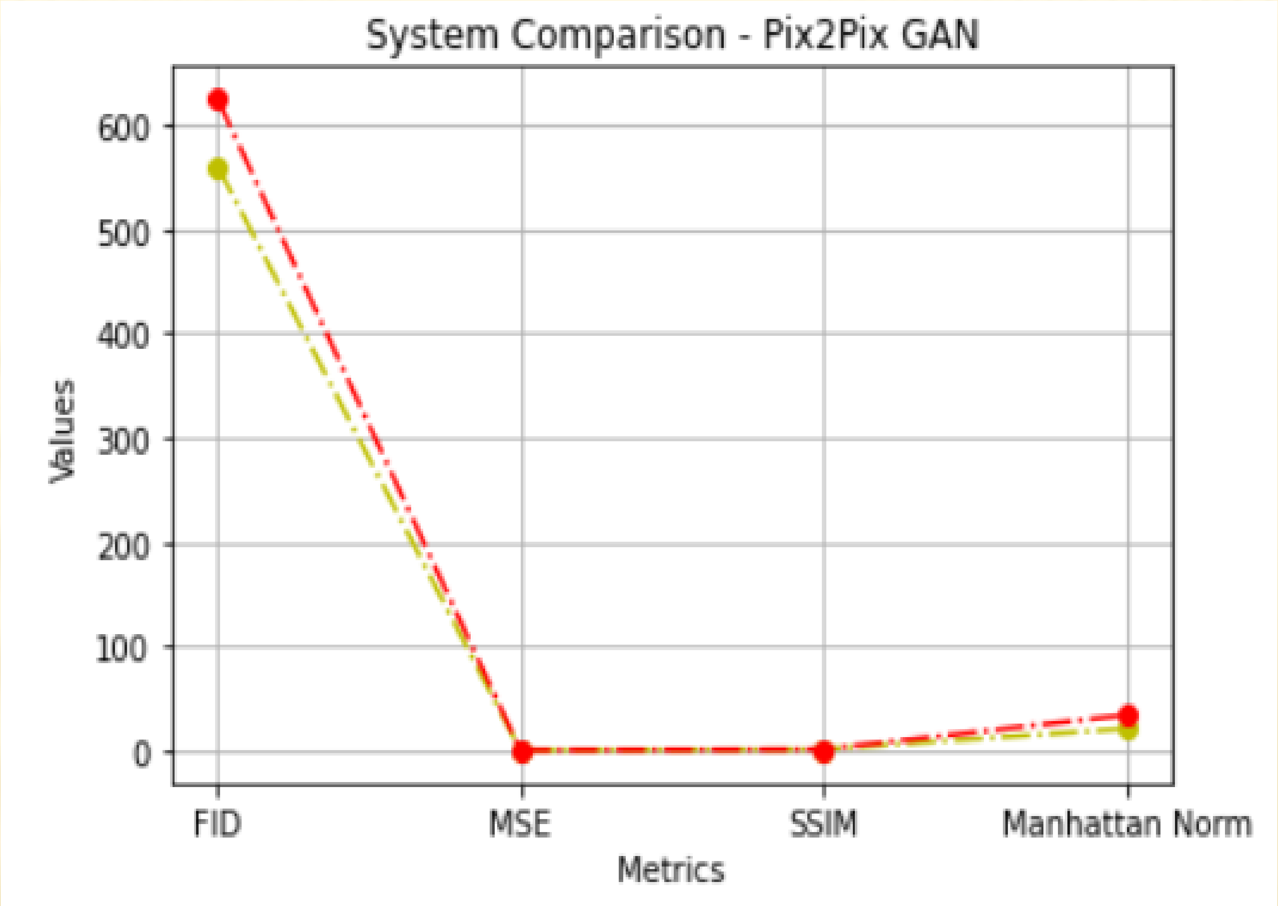








EXISTING VS PROPOSED SYSTEM



CONCLUSION AND REFERENCES

CONCLUSION

- We have developed an intelligent system that has learned to generate the aerial map version of any complicated satellite image with buildings, trees, grassland, rivers, sea, land, etc which was captured at any part of the day.
- We successfully and accurately detect all the visible as well as hidden edifices that are part of the image which helps the secret services, intelligence bureaus and other related international agencies to figure out the locations of anti - nationalists and terrorists and bring down the crime rate by a significant amount.
- The system also plays an extremely important role when it comes go intra - planetary transport and space exploration.

REFERENCES

[1] P. Isola, J. Zhu, T. Zhou and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 5967-5976, doi: 10.1109/CVPR.2017.632.

[2] H. Tang, H. Liu and N. Sebe, "Unified Generative Adversarial Networks for Controllable Image-to-Image Translation," in IEEE Transactions on Image Processing, vol. 29, pp. 8916-8929, 2020, doi: 10.1109/TIP.2020.3021789.

[3] N. -T. Tran, V. -H. Tran, N. -B. Nguyen, T. -K. Nguyen and N. -M. Cheung, "On Data Augmentation for GAN Training," in IEEE Transactions on Image Processing, vol. 30, pp. 1882-1897, 2021, doi: 10.1109/TIP.2021.3049346.

[4] C. Min, Y. Li, L. Fang and P. Chen, "Conditional Generative Adversarial Network on Semi-supervised Learning Task," 2019 IEEE 5th International Conference on Computer and Communications (ICCC), Chengdu, China, 2019, pp. 1448-1452, doi: 10.1109/ICCC47050.2019.9064268.

[5] M. Oza, H. Vaghela and S. Bagul, "Semi-Supervised Image-to-Image Translation," 2019 International Conference of Artificial Intelligence and Information Technology (ICAIIT), Yogyakarta, Indonesia, 2019, pp. 16-20, doi: 10.1109/ICAIIT.2019.8834613.

[6] I. B. Akkaya and U. Halici, "GAIT: Gradient Adjusted Unsupervised Image-To-Image Translation," 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 2020, pp. 1586-1590, doi: 10.1109/ICIP40778.2020.9191271.

[7] R. Yin, "Multi-Resolution Generative Adversarial Networks for Tiny-Scale Pedestrian Detection," 2019 IEEE International Conference on Image Processing (ICIP), 2019, pp. 1665-1669, doi: 10.1109/ICIP.2019.8803030.