Image Manipulation Using Generative Machine Learning

Submitted in partial fulfillment of the requirements of the degree

BACHELOR OF ENGINEERING IN **COMPUTER ENGINEERING**

By

Hriday Keswani 2003088

Viren Keswani 2003089

Vedant Jumle 2003074

Supervisor

Prof. Manisha Dumbre

THADOMAL SHAHANI
TSEC
ENGINEERING COLLEGE

Computer Engineering Department

Thadomal Shahani Engineering College

Bandra(w), Mumbai - 400 050

University of Mumbai

(AY 2021-22)

CERTIFICATE

This is to certify that the Mini Project entitled **“Image Manipulation Using Generative Machine Learning”** is a bonafide work of **Hriday Keswani (2003088), Viren Keswani (2003089), Vedant Jumle (2003074)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **“Bachelor of Engineering”** in **“Computer Engineering”.**

(Prof. Manisha Dumbre)
Supervisor

(Prof. Tanuja Sarode)
Head of Department

(Prof. G.T. Thampi)
Principal

Mini Project Approval

This Mini Project entitled "Image Manipulation Using Generative Machine Learning**"** by **Hriday Keswani (2003088), Viren Keswani (2003089), Vedant Jumle (2003074)** is approved for the degree of **Bachelor of Engineering** in **Computer Engineering.**

Examiners

1………………………………………
(Internal Examiner Name & Sign)

2………………………………………
(External Examiner name & Sign)

Date:

Place

Content

GAN	Generative Adversarial Network
NST	Neural Style Transfer
DCGAN	Deep Convolutional Generative Adversarial Network
SRGAN	Super Resolution Generative Adversarial Network

Acknowledgement

We would like to express our gratitude and thanks to **Prof. Manisha Dumbre** for her valuable guidance and help. We are indebted for her guidance and constant supervision as well as for providing necessary information regarding the project. We would like to express our greatest appreciation to our principal **Dr. G.T. Thampi** and head of the department **Dr. Tanuja Sarode** for their encouragement and tremendous support. We take this opportunity to express our gratitude to the people who have been instrumental in the successful completion of the project.

Hriday Keswani

Viren Keswani

Vedant Jumle

Chapter 1

1.1 Introduction

Generative Adversarial Neural Network (GAN) has been part of the recent advancements in the field of Machine Learning. In the recent growth the domain of generative machine learning, our project is meant to showcase the power of the new technologies and techniques.

1.2 Motivation

Our motivation stems from our high interest in the field of Machine Learning. We were initially very fascinated when we discovered this new avenue of Machine Learning and the things that engineers and researchers were building from this new concept. For example, DLSS or Deep Learning Super Sampling which is a technology developed by Nvidia and is used in computer graphics to upscale rendered video from a lower resolution to higher resolution without loss in and apparent details. Technologies like this motivated us to study and implement the concepts and techniques that are used in them.

1.3 Problem Statement & Objectives

Image manipulation using generative machine learning.
The objective of this project is to show the power of generative machine learning. In recent years the advancements in ML have led to the birth of the Generative Adversarial Neural Network which has created a giant discourse in the Image Processing field. Here we aim to show some applications of the new technology based on the original GAN paper and other ideas in the generative machine learning space.

1.4 Organization of the Report

This report consists of three chapters. The first chapter deals with introduction of the topic, problem statement, motivation behind the topic and objectives. The second chapter is the Literature Survey. It includes all the research work done related to this topic. All information related to study of existing systems as well as learning of new tools is mentioned in this chapter. The third chapter is about the proposed system which is used in this project. The block diagram, techniques used, hardware and software used screenshots of the project are presented in this chapter. All the documents related to development of this project are mentioned in References

Chapter 2

Literature Survey

2.1 Survey of Existing System

The concepts and techniques behind generative machine learning is used in many different systems in different domains.

- Social media platforms like Instagram, and Snapchat use generative machine learning to implement certain artistic image filters.
- Deep Learning Super Sampling, as mentioned before, used to seamlessly upscale rendered frames to higher resolutions uses GANs.
- File Industry use GANs to perform all sorts of different image manipulation to make Computer Generated Imagery (CGI).
- AI image enhancement systems in mobile phones allows use to capture photos with very high details without the need of complicated camera setups.

2.2 Limitation of existing system

There are no apparent limitations in existing systems. The only thing we saw was that there was no such platform that showcases the basic form of these technologies.

2.3 Mini Project Contribution

Our project is an implementation of the original research papers for NST, DCGAN, and SRGAN. NST can be used to style different images with a particular styling image, DCGAN can generate random realistic faces that can unique and never seen before, and SRGAN can be used to upscale images without loss in details. Essentially, this project showcases the new advancements in AI and ML.

Chapter 3

Proposed System

3.1 Introduction

- [1]Neural Style Transfer (NST):

  Neural Style Transfer is a model which takes in to input: content image and style image, and outputs an image which is 'stylized' formed of the content image. Eg:

Fig 3.1.1 Content Image

Fig 3.1.2 Style Image

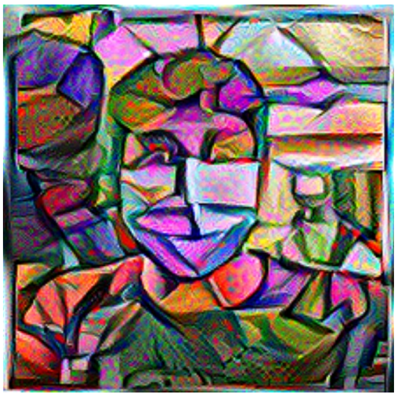


Fig 3.1.3 Stylized image

- [2] Deep Convolutional GAN (DCGAN):

  DCGAN is a model that generates random and novel images from a given set of noise. The property of this model is that it can generate images without any simulations.

Fig 3.1.4 DCGAN example

- [3] Super Resolution GAN (SRGAN):
  SRGAN is a model that takes a low-resolution image and upscales it to a higher resolution, with minimal loss in details of the image unlike traditional upscaling methods like Cubic Interpolation or Nearest Neighbor Interpolation.

Fig 3.1.5 Original Low resolution image

Fig 3.1.6 SRGAN output

Fig 3.1.7 Cubic Interpolation

Fig 3.1.8 Nearest Neighbor Interpolation

3.2 Architecture

In Neural Style Transfer, we use a pretrained model (in this case VGG-19), and pass both our content and style image through it. We select one of the middle layers of the model as the output for our content image representation, and 5 layers from the start to the middle of the model as the style representation. We the collected images for style representation, we create a 'Gram Matrix' and it is then combined with content representation to create our stylized image, we then calculate the [1] vgg-loss value as described in the paper. This process is repeated over a large number of time to generate an output.

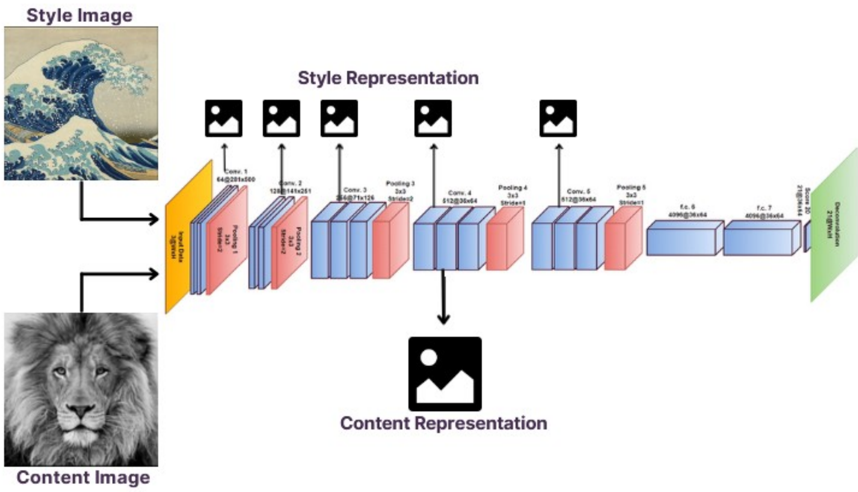


Fig 3.2.1 Neural Style Transfer Layer selection

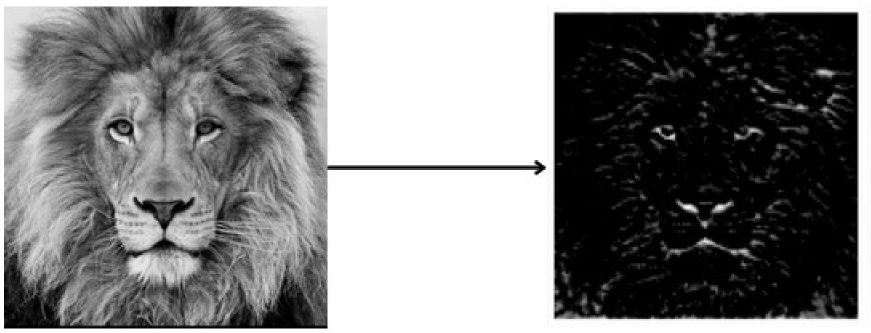


Fig 3.2.2 Content Representation

Fig 3.2.3 Style representation

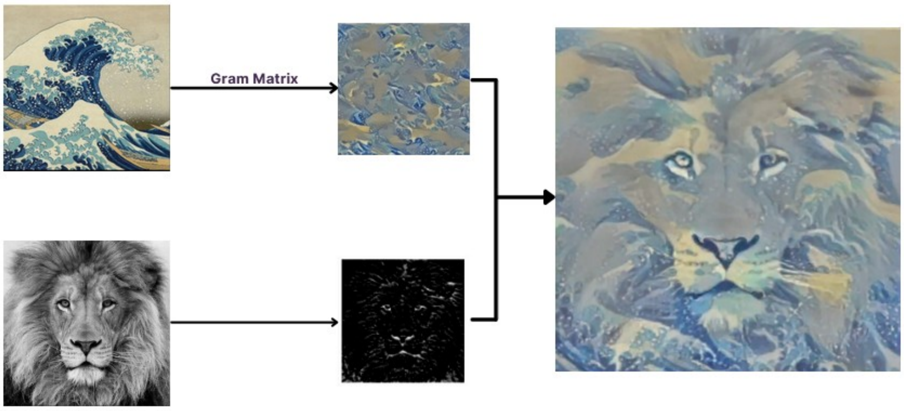


Fig 3.2.4 Final Stylized image

Generative Adversarial Neural Networks or GAN networks consists of two distinct models: A generator and a discriminator. The generator uses Convolutional Transpose layers to generate images which are then fed into the discriminator which then determines if the image is fake or not. This result is then used to calculate the [1] GAN loss value which is used to adjust the weights in the generator and the discriminator.

Fig 3.2.5 Concept of GAN

Deep Convolutional GANs uses the same method as GANs to generate random images. For this we use the [4] Google Cartoon faces dataset to train the generator and discriminator to generate random images of cartoon faces.

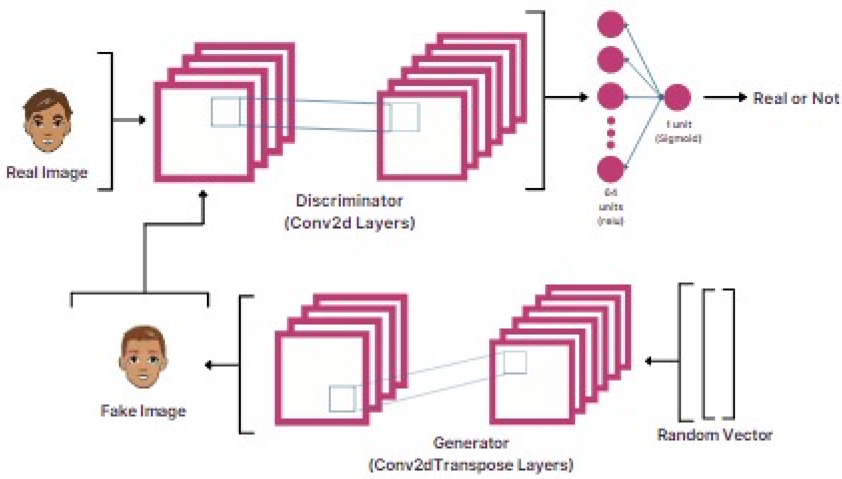


Fig 3.2.6 DCGAN model architecture

Super Resolution GAN has an architecture similar to DCGAN but it takes in low resolution image input for the generator which then creates a higher resolution version of the input image. Here we use a pretrained model (in this case VGG-19) to calculate the [3] perceptual loss value, which is a measure of how real the generated images look.

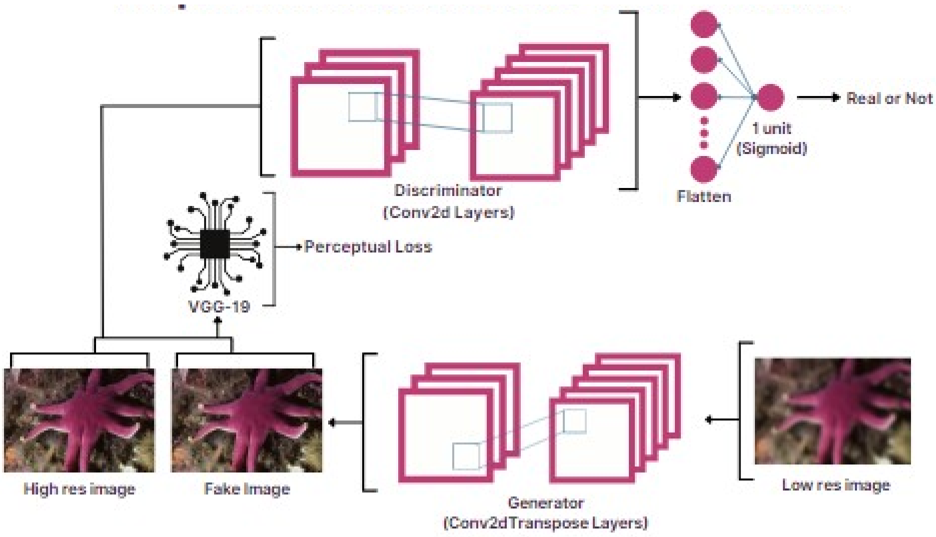


Fig 3.2.7 SRGAN architecture

3.3 Algorithm and Process Design

1. Formulating the Problem statement:

   After much thought and research, we found that GANs are a very interesting topic in ML. These being the new form of image generation, we got curious and finalized to make SRGAN as the most complex sub-topic of our project

2. Understanding the framework and requirements:

   We had a relatively good understanding of the concepts that were discussed in the original papers. We only had to figure out how to make custom models in TensorFlow.

3. Identifying tools/technology to be used:

   - Python
   - TensorFlow
   - Flask
   - JavaScript
   - HTML/CSS

4. Finalizing the features to be included:

We finitized to the project to have three levels of increasingly complex ML models. First being Neural Style Transfer, second being DCGAN, third being SRGAN.

5. Development:

A lot of the time of development was spent in optimizing the machine learning model so that it does not bottle neck the system. We used two datasets to train our models:

1. Google's Cartoon Faces dataset: This dataset consists of 10,000 unique cartoon faces which was used to train the DCGAN model.

Fig 3.3.1 Google cartoon faces dataset example

2. Div2K dataset: This dataset consists 2,000 images in 4k resolution which was use to train the SRGAN model.

Fig 3.3.2 Div2k dataset example

Once the model was ready, we worked on integrating it into an API which then can be used in the web server.

6. Testing:

7. Testing of ML model was fairly easy. Since it generated images, we can visually check the results and validate if the models were performing as expected, which they did.

3.4 Details of Hardware & Software

The project was developed in the High-end configuration of:

8X2 GB DDR4 3200MHz RAM

AMD RYZEN 5 5600X

RTX 3080 OC 10GB VRAM

Windows 10/11

VS Code text editor

and it can run on any device capable of rendering a website

3.5 Results

Results for Neural Style Transfer:

Fig 3.5.1 NST results

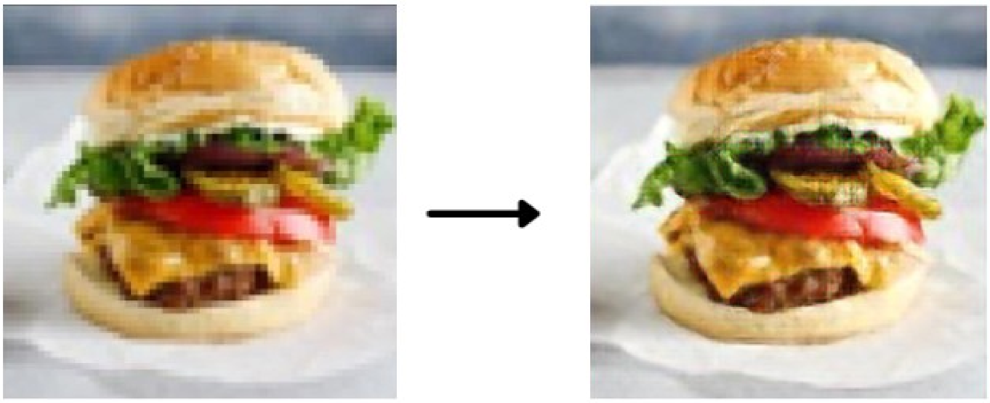
Here we see two examples from the results of Neural Style Transfer implementation, first being the combination of a scenery picture from Holland and 'The Starry Night' a painting by Vincent Van Gogh. The second being a combination of the 'Mona Lisa' painting and some abstract art.

Results for Deep Convolutional GAN:

Fig 3.5.2 Result for DCGAN

Here we see a Result from the DCGAN's implementation where both the images were generated from random noise and hence are completely unique.

Results of Super Resolution GAN:



(a) low resolution image (b) super-resolution image

Fig 3.5.3 Result of SRGAN

Here we see a result from the SRGAN's implementation, where we upscale a 64x64 **(3.5.3 fig(a))** image of a burger to 256x256 image **(3.5.3 fig(b))** without loss in detail.

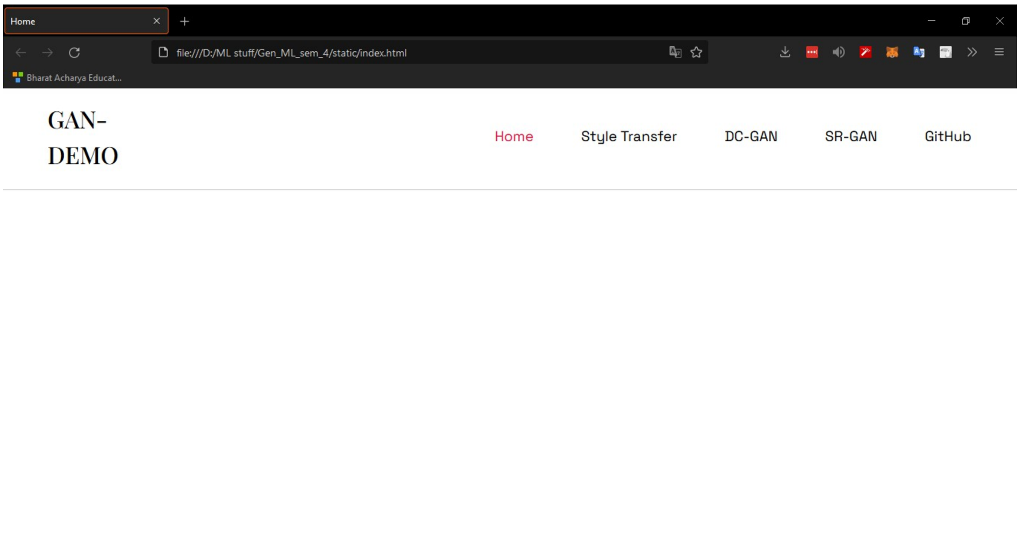


Fig 3.5.4 Website

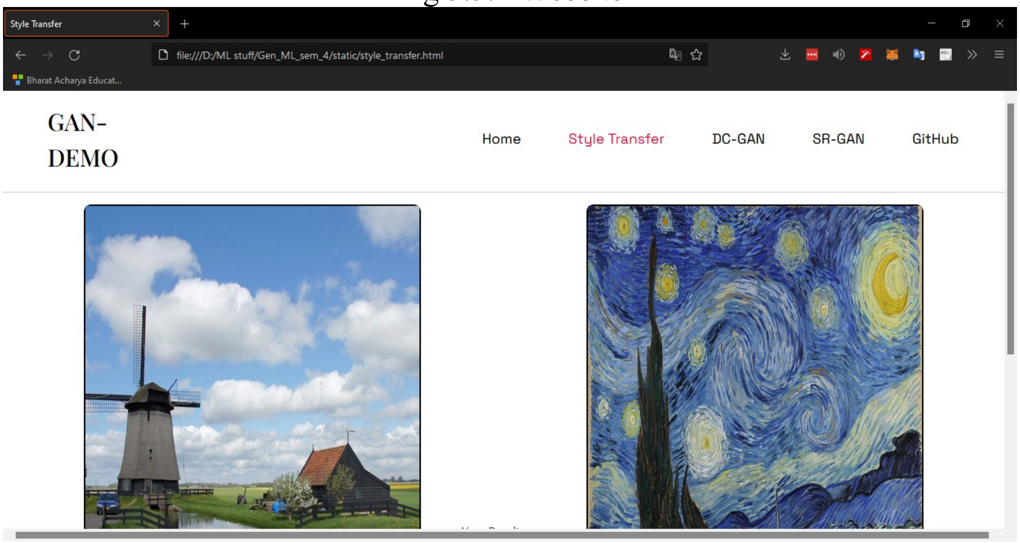


Fig 3.5.5 NST website

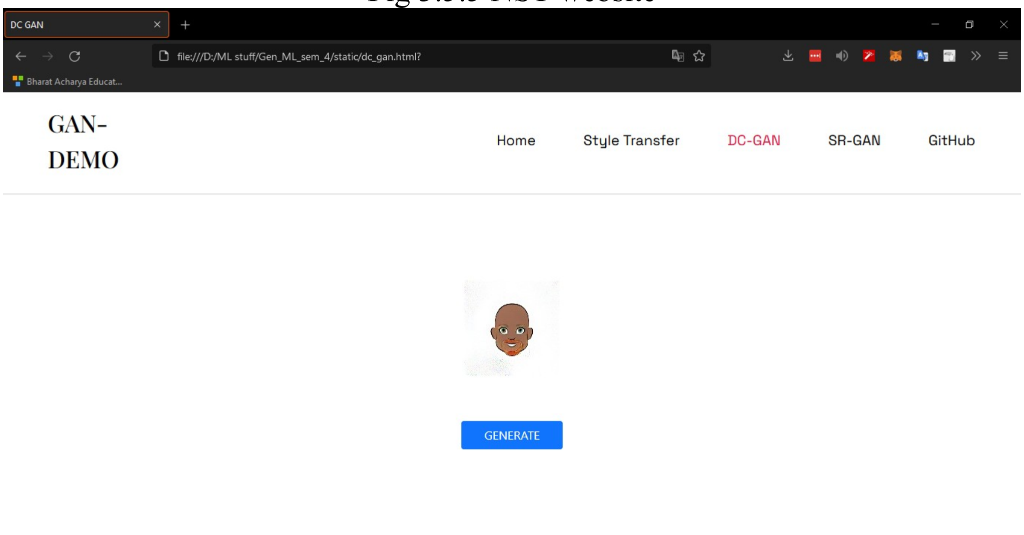


Fig 3.5.6 DCGAN website

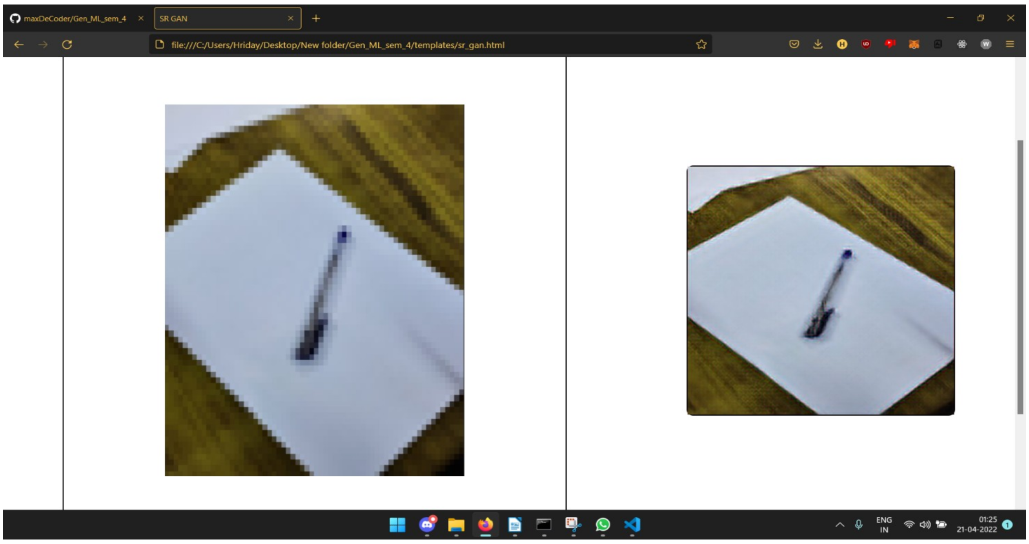


Fig 3.5.7 SRGAN website

In the above figures from 3.5.4 to 3.5.7, we see how the website looks likes.

3.6 Conclusion and Future Work

In future we can implement multi-topical generator for DCGAN, which can generate random images of different topic. We can also train the DCGAN model to take in required features for the image and make it generate images that align with the given set of features of the image. For SRGAN, we can improve upon it by implementing the Extended SRGAN model (ESRGAN).

References

[1] L. Gatys, A. Ecker, and M. Bethge, “A neural algorithm of artistic style,” Journal of Vision, vol. 16, no. 12, p. 326, 2016.

[2] A. Radford, L. Metz, and S. Chintala, Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, Aug. 2015.

[3] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-realistic single image super-resolution using a generative adversarial network,” 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.