

DEEP LEARNING FOR TEXT-DRIVEN EDITING OF IMAGES

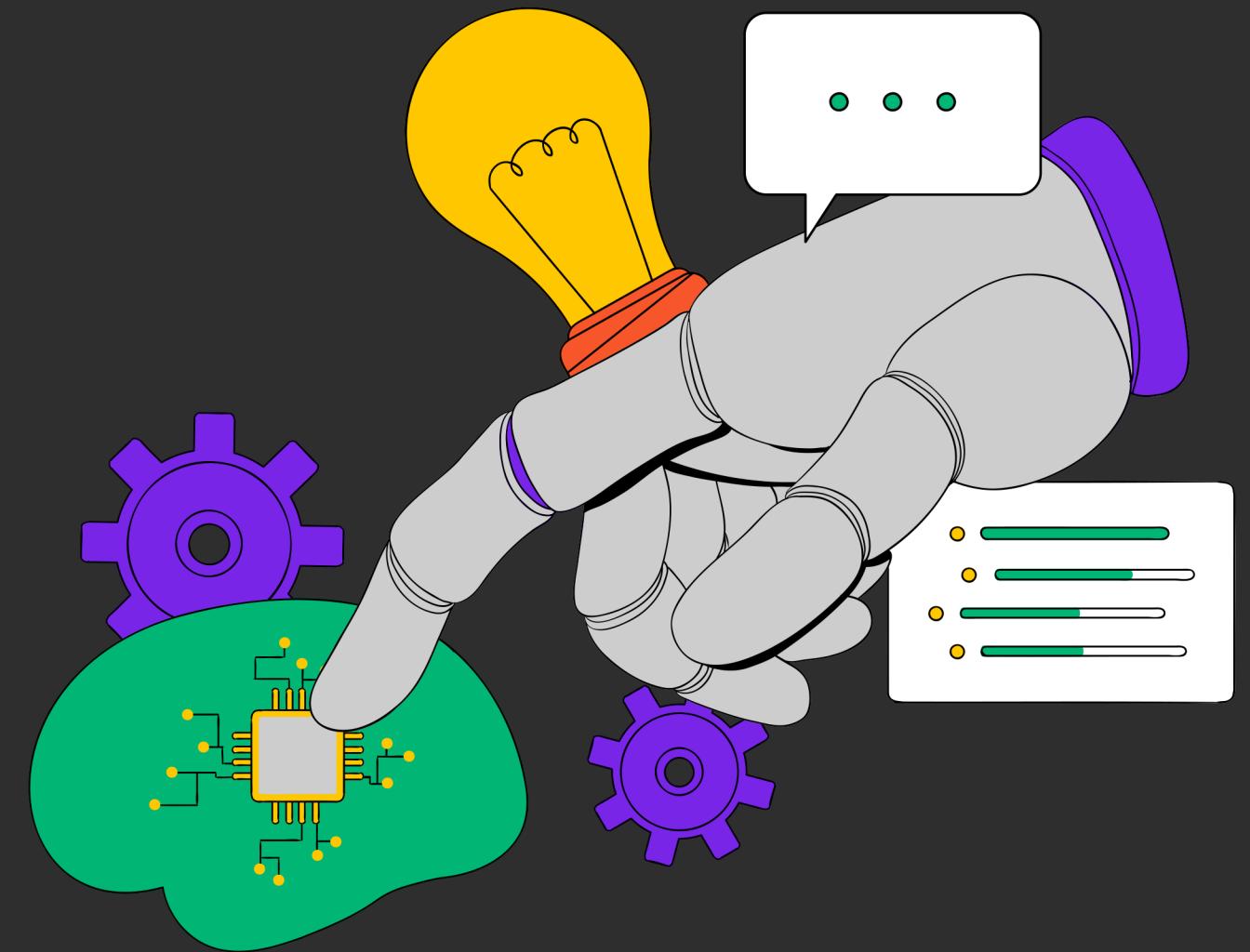
Usman Manzoor(P19-0068)
Sarim Yawar(P19-0001)
Haseeb Jan(P19-0003)

Supervised by : Mr. Shahzeb Khan

Introduction



A method for editing images from human instructions, given an input image and a written instruction that tells the model what to do, our model follows these instructions to edit the image.



System Analysis

Objective

The objective of the project is to develop a software system that enables users to edit images based on textual instructions or inputs.

Text-to-Image Translation

The system analyzes the provided text inputs and utilizes algorithms and techniques to transform and edit the corresponding images accordingly.

User Interaction

The software system provides a user-friendly interface for users to input images and make changes to the images as needed.

Problems

- Ambiguity in Textual Instructions.
- Understanding Context and Intent
- Complex Image Editing Requirements
- Algorithm Selection and Optimization





Goals Achieved



- Accomplished the development of an image-to-image editing model using CycleGANs, surpassing initial project goals.
- we achieved the transformation of color mapping, specifically converting diverse colors of shirts, pants, and shoes into a standardized palette of red, green, black, and blue.



UI/UX Design



IMAGER - Image Editing Platform

Select Model

Red

Choose an image...

Drag and drop file here
Limit 200MB per file • JPG, JPEG

Browse files

163.jpg 12.2KB

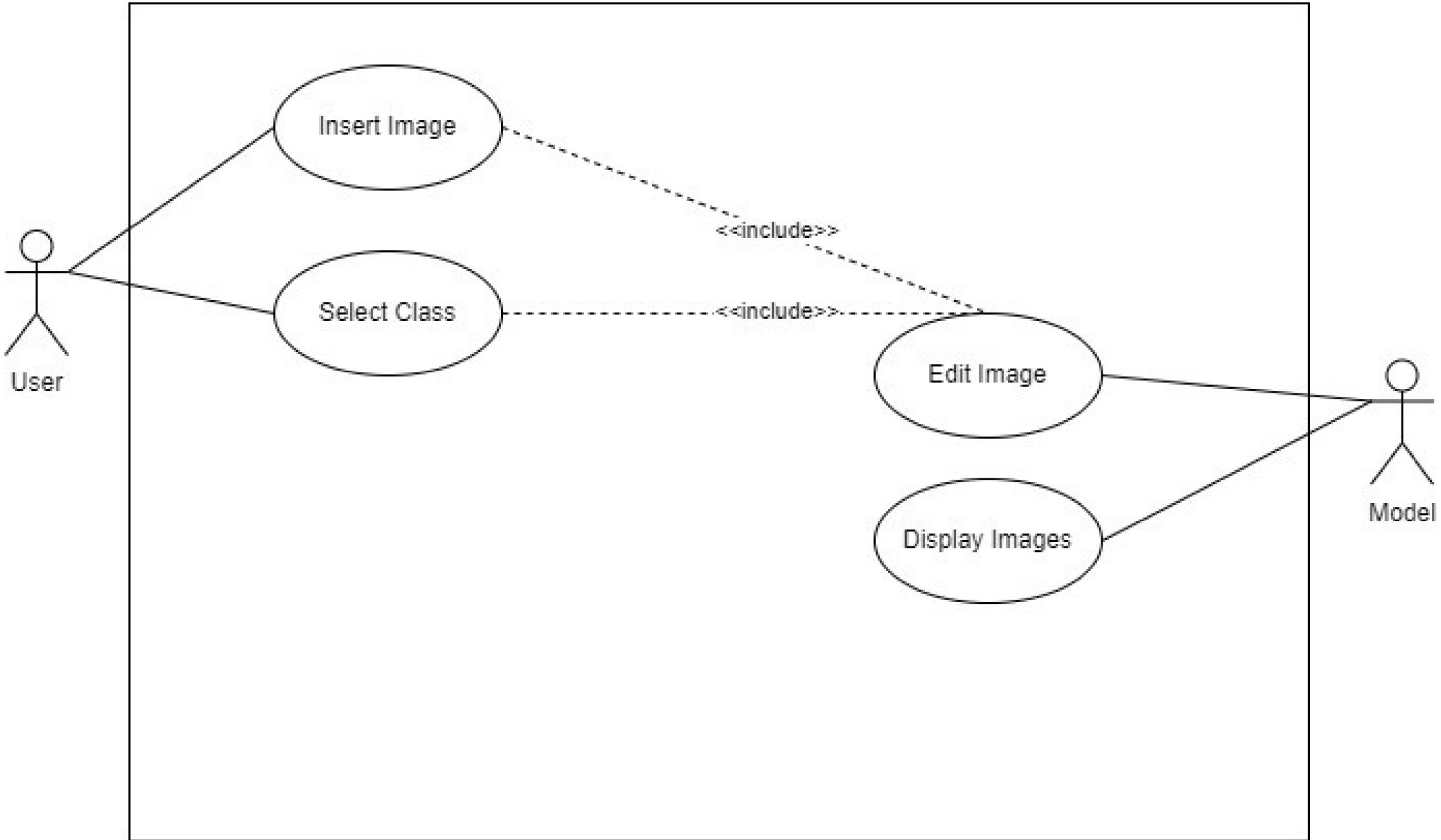
Edit Image

Uploaded Image

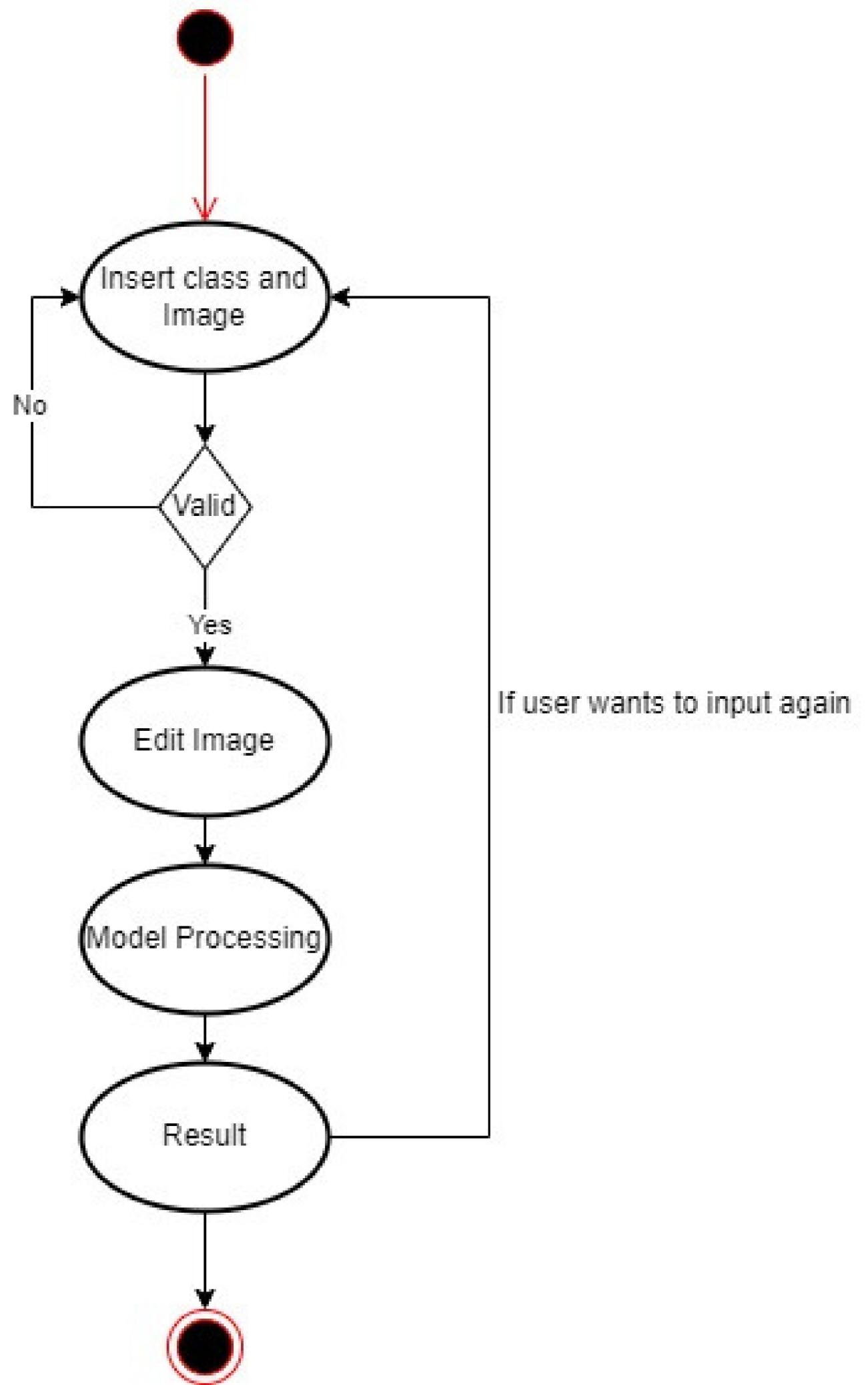
Output Image

The screenshot shows a user interface for an image editing platform. At the top, there's a dropdown menu labeled "Select Model" with "Red" selected. Below it is a file upload section with a placeholder "Choose an image..." and options to "Drag and drop file here" or "Browse files". A file named "163.jpg" (12.2KB) has been uploaded. Below the upload area are two images: the "Uploaded Image" (a blue Adidas t-shirt) and the "Output Image" (the same t-shirt colored red). There's also an "Edit Image" button.

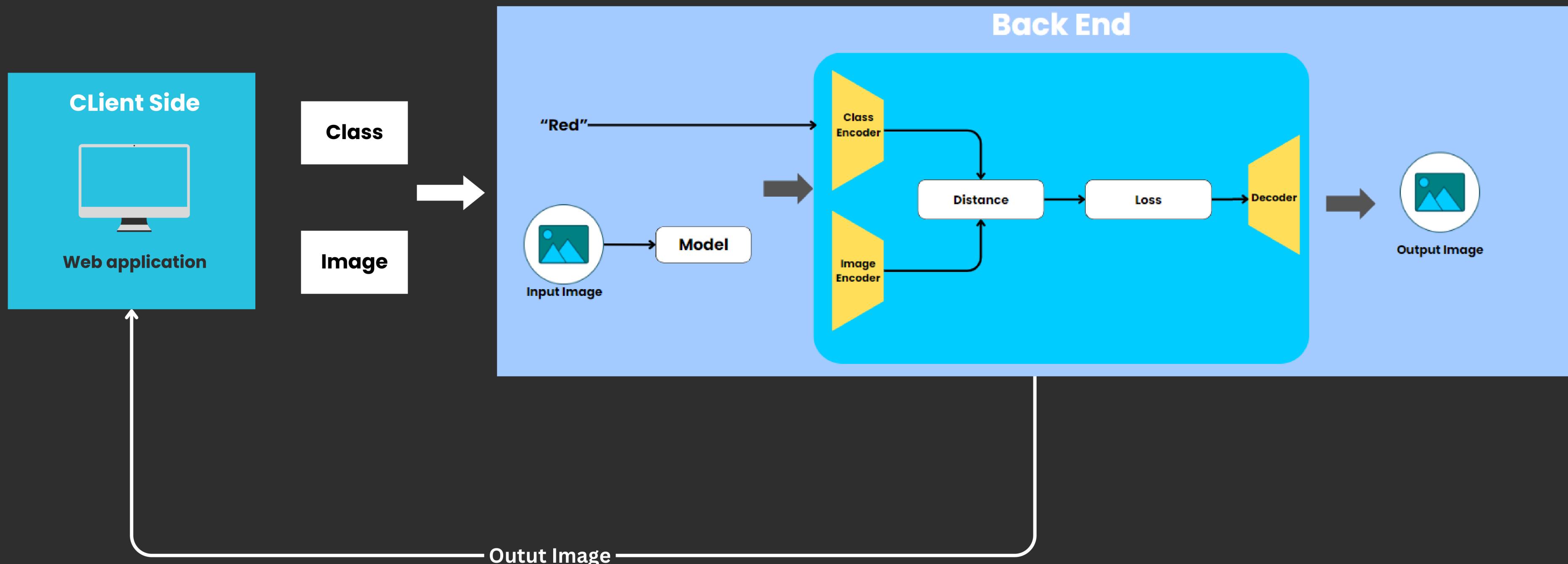
UseCase



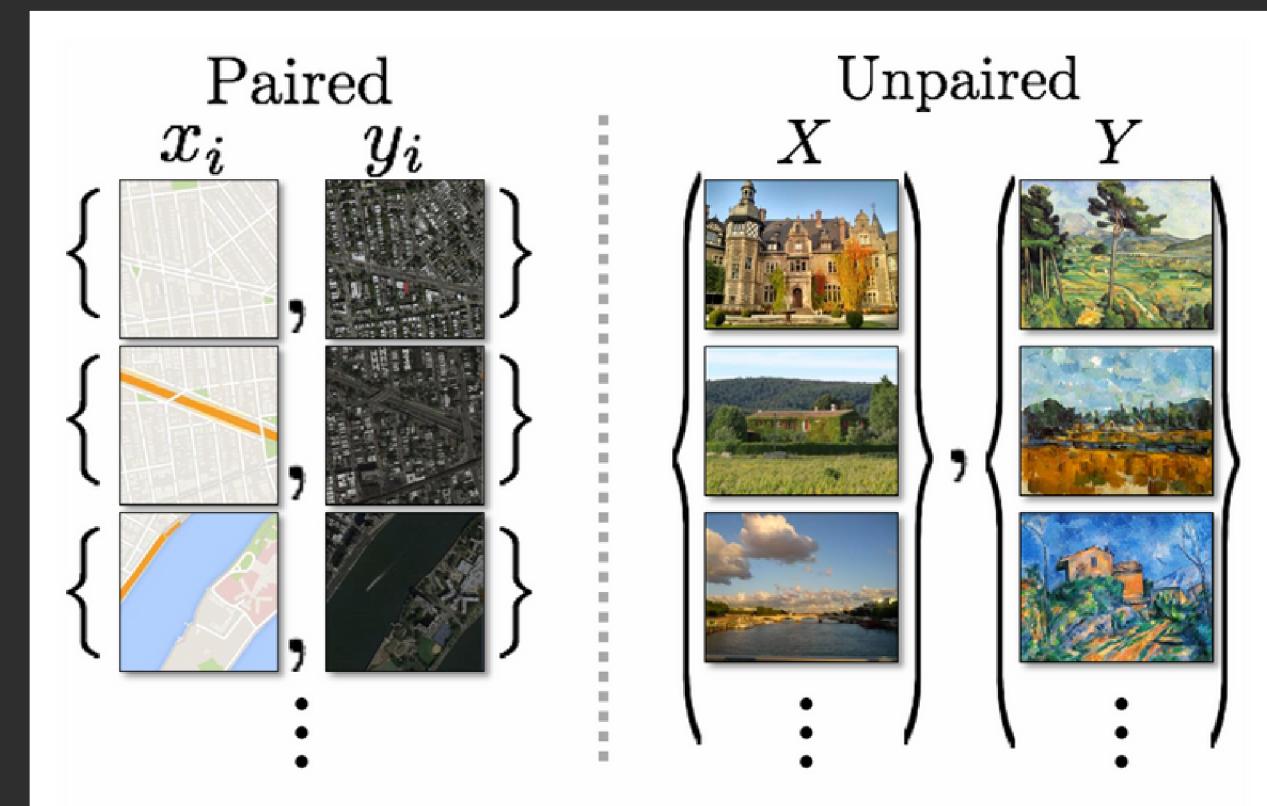
Activity Diagram



System Design



Dataset



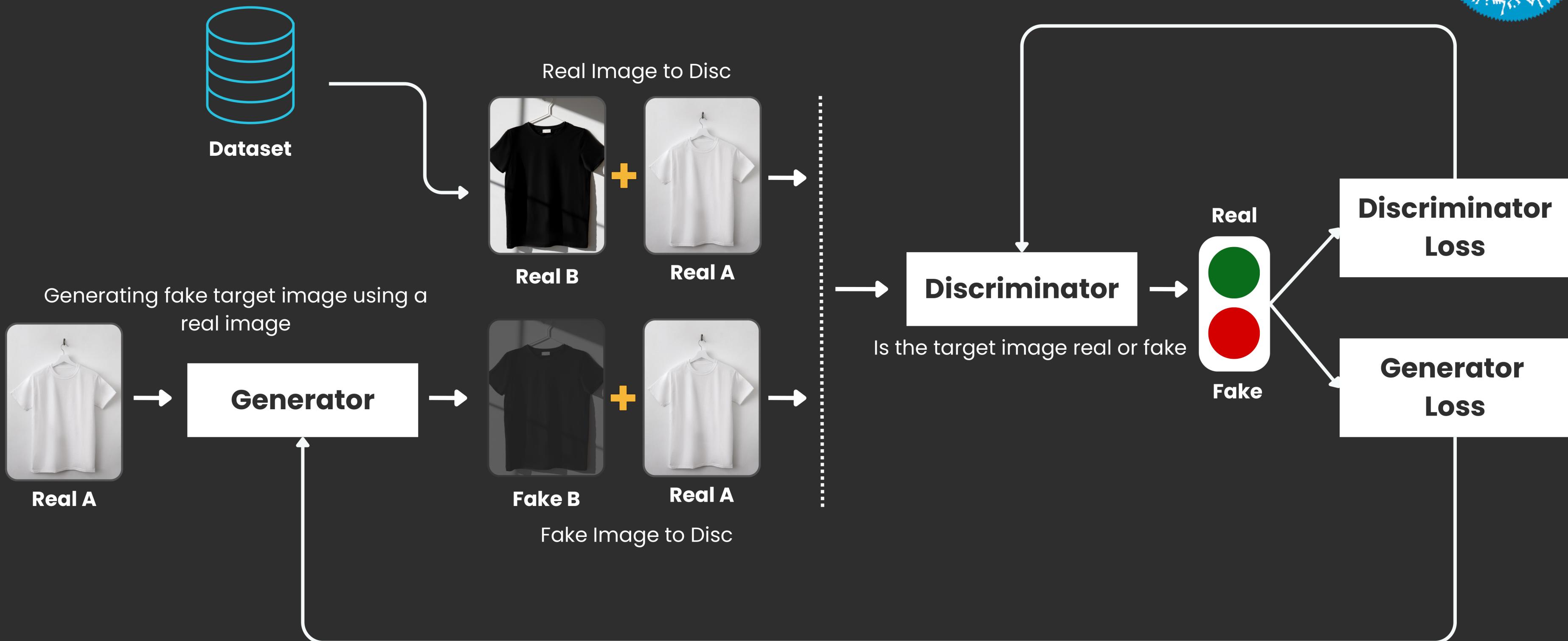
Apparel Dataset

- 16,170 images
- 8 different clothing categories in 9 different colors
- Multi-labeled

Deep Fashion

- 80,000 images
- Clothes(dresses, shirts, pants), bags, shoes, jewellery
- Rich Annotations

Cycle GANs Architecture



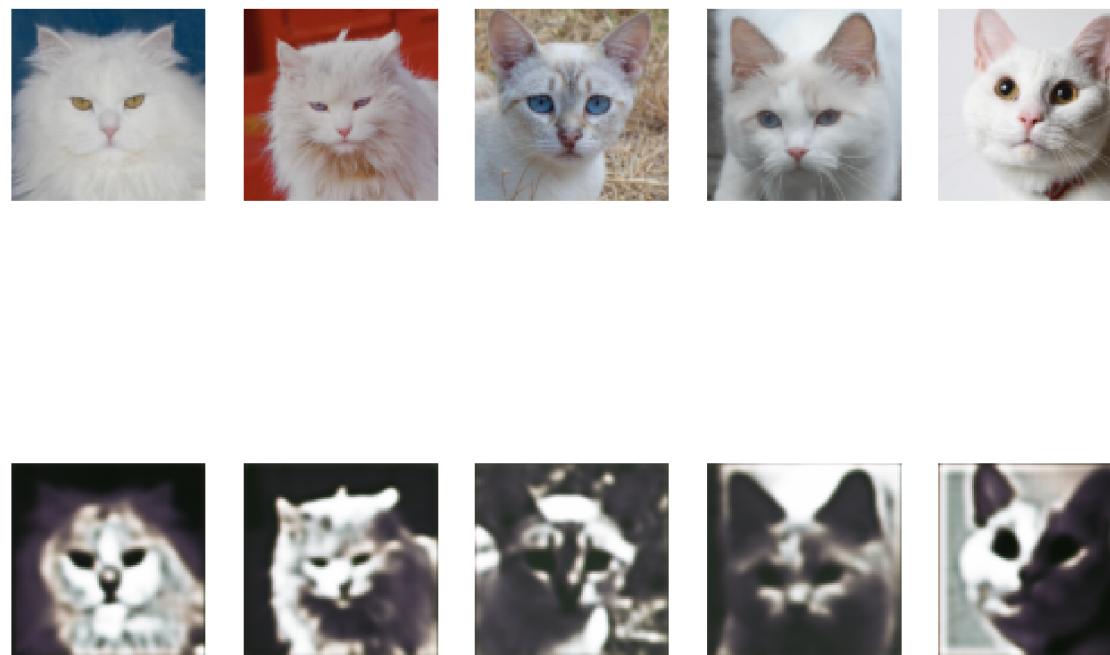


Model Training Progression

Class A → Class B



Issues Faced



```

from datetime import datetime
start1 = datetime.now()
# train models
train(d_model_A, d_model_B, g_model_AtoB, g_model_BtoA, c_model_AtoB, c_model_BtoA, dataset, epochs=3)

stop1 = datetime.now()
#Execution time of the model
execution_time = stop1-start1
print("Execution time is: ", execution_time)
1/1 [-----] - 4s 4s/step
Iteration>297, dA[0.042,0.101] dB[0.075,0.034] g[5.122,5.431]
1/1 [-----] - 4s 4s/step

1/1 [-----] - 4s 4s/step
Iteration>298, dA[0.184,0.238] dB[0.016,0.488] g[5.488,6.797]
1/1 [-----] - 4s 4s/step
1/1 [-----] - 4s 4s/step
Iteration>299, dA[0.226,0.030] dB[0.018,0.010] g[7.928,8.255]
1/1 [-----] - 4s 4s/step
1/1 [-----] - 4s 4s/step
Iteration>300, dA[0.187,0.072] dB[0.113,0.015] g[5.789,5.792]
1/1 [-----] - 17s 17s/step
1/1 [-----] - 17s 17s/step
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. 'model.compile_metrics' will be
empty until you train or evaluate the model.
WARNING:tensorflow:Compiled the loaded model, but the compiled metrics have yet to be built. 'model.compile_metrics' will be
empty until you train or evaluate the model.
>Saved: g_model_AtoB_000300.h5 and g_model_BtoA_000300.h5
Execution time is: 9:52:21.984208

```

Our model was taking up to 10 hours to complete a training session. Even more, despite this extensive training duration, the results were far from satisfactory. Our objective was to achieve the transformation of a white cat image into a black cat image.

Error and Issues



Error and Issues



Error and Issues



```
20/20 [=====] - 28s 1s/step - gen_w2b_loss: 10.9186 - gen_b2w_loss: 9.9632 - disc_white_loss: 0.2854 - disc_black_loss: 0.0853 - lr: 0.6327
Epoch 90/100
20/20 [=====] - ETA: 0s - gen_w2b_loss: 10.1501 - gen_b2w_loss: 9.5885 - disc_white_loss: 0.1702 - disc_black_loss: 0.2131-->Saved: g_model_AtoB.h5, g_model_BtoA.h5
20/20 [=====] - 28s 1s/step - gen_w2b_loss: 10.3676 - gen_b2w_loss: 9.7163 - disc_white_loss: 0.1675 - disc_black_loss: 0.2077 - lr: 0.6992
Epoch 91/100
20/20 [=====] - ETA: 0s - gen_w2b_loss: 10.1326 - gen_b2w_loss: 9.7870 - disc_white_loss: 0.2891 - disc_black_loss: 0.2006-->Saved: g_model_AtoB.h5, g_model_BtoA.h5
20/20 [=====] - 28s 1s/step - gen_w2b_loss: 10.1191 - gen_b2w_loss: 9.8067 - disc_white_loss: 0.2772 - disc_black_loss: 0.1962 - lr: 0.7728
Epoch 92/100
20/20 [=====] - ETA: 0s - gen_w2b_loss: 11.6338 - gen_b2w_loss: 9.6229 - disc_white_loss: 0.2205 - disc_black_loss: 0.0829-->Saved: g_model_AtoB.h5, g_model_BtoA.h5
20/20 [=====] - 28s 1s/step - gen_w2b_loss: 11.6010 - gen_b2w_loss: 9.6191 - disc_white_loss: 0.2130 - disc_black_loss: 0.0853 - lr: 0.8541
Epoch 93/100
20/20 [=====] - ETA: 0s - gen_w2b_loss: 11.8248 - gen_b2w_loss: 10.1358 - disc_white_loss: 0.2708 - disc_black_loss: 0.0958-->Saved: g_model_AtoB.h5, g_model_BtoA.h5
20/20 [=====] - 29s 1s/step - gen_w2b_loss: 11.9397 - gen_b2w_loss: 10.2327 - disc_white_loss: 0.2634 - disc_black_loss: 0.0931 - lr: 0.9439
Epoch 94/100
20/20 [=====] - ETA: 0s - gen_w2b_loss: 10.5100 - gen_b2w_loss: 9.6514 - disc_white_loss: 0.2411 - disc_black_loss: 0.2052-->Saved: g_model_AtoB.h5, g_model_BtoA.h5
20/20 [=====] - 28s 1s/step - gen_w2b_loss: 10.4553 - gen_b2w_loss: 9.6945 - disc_white_loss: 0.2403 - disc_black_loss: 0.2083 - lr: 1.0432
Epoch 95/100
20/20 [=====] - ETA: 0s - gen_w2b_loss: 10.4246 - gen_b2w_loss: 10.3211 - disc_white_loss: 0.1959 - disc_black_loss: 0.2401-->Saved: g_model_AtoB.h5, g_model_BtoA.h5
20/20 [=====] - 28s 1s/step - gen_w2b_loss: 10.3702 - gen_b2w_loss: 10.3577 - disc_white_loss: 0.2035 - disc_black_loss: 0.2435 - lr: 1.1529
Epoch 96/100
20/20 [=====] - ETA: 0s - gen_w2b_loss: 10.7182 - gen_b2w_loss: 10.2660 - disc_white_loss: 0.2979 - disc_black_loss: 0.2533-->Saved: g_model_AtoB.h5, g_model_BtoA.h5
20/20 [=====] - 29s 1s/step - gen_w2b_loss: 10.6960 - gen_b2w_loss: 10.3086 - disc_white_loss: 0.2864 - disc_black_loss: 0.2537 - lr: 1.2741
Epoch 97/100
20/20 [=====] - ETA: 0s - gen_w2b_loss: 11.3962 - gen_b2w_loss: 10.7600 - disc_white_loss: 0.6152 - disc_black_loss: 0.1864-->Saved: g_model_AtoB.h5, g_model_BtoA.h5
20/20 [=====] - 28s 1s/step - gen_w2b_loss: 11.4400 - gen_b2w_loss: 10.7669 - disc_white_loss: 0.5916 - disc_black_loss: 0.1810 - lr: 1.4081
Epoch 98/100
20/20 [=====] - ETA: 0s - gen_w2b_loss: 10.7896 - gen_b2w_loss: 9.3380 - disc_white_loss: 0.2512 - disc_black_loss: 0.1527-->Saved: g_model_AtoB.h5, g_model_BtoA.h5
20/20 [=====] - 28s 1s/step - gen_w2b_loss: 10.6777 - gen_b2w_loss: 9.3710 - disc_white_loss: 0.2422 - disc_black_loss: 0.1561 - lr: 1.5562
Epoch 99/100
20/20 [=====] - ETA: 0s - gen_w2b_loss: 10.5622 - gen_b2w_loss: 9.3073 - disc_white_loss: 0.3624 - disc_black_loss: 0.2403-->Saved: g_model_AtoB.h5, g_model_BtoA.h5
20/20 [=====] - 28s 1s/step - gen_w2b_loss: 10.6507 - gen_b2w_loss: 9.2956 - disc_white_loss: 0.3573 - disc_black_loss: 0.2462 - lr: 1.7199
Epoch 100/100
20/20 [=====] - ETA: 0s - gen_w2b_loss: 10.2107 - gen_b2w_loss: 8.8363 - disc_white_loss: 0.2373 - disc_black_loss: 0.1875-->Saved: g_model_AtoB.h5, g_model_BtoA.h5
20/20 [=====] - 28s 1s/step - gen_w2b_loss: 10.2860 - gen_b2w_loss: 8.8021 - disc_white_loss: 0.2430 - disc_black_loss: 0.1875 - lr: 1.9007
```

Error and Issues



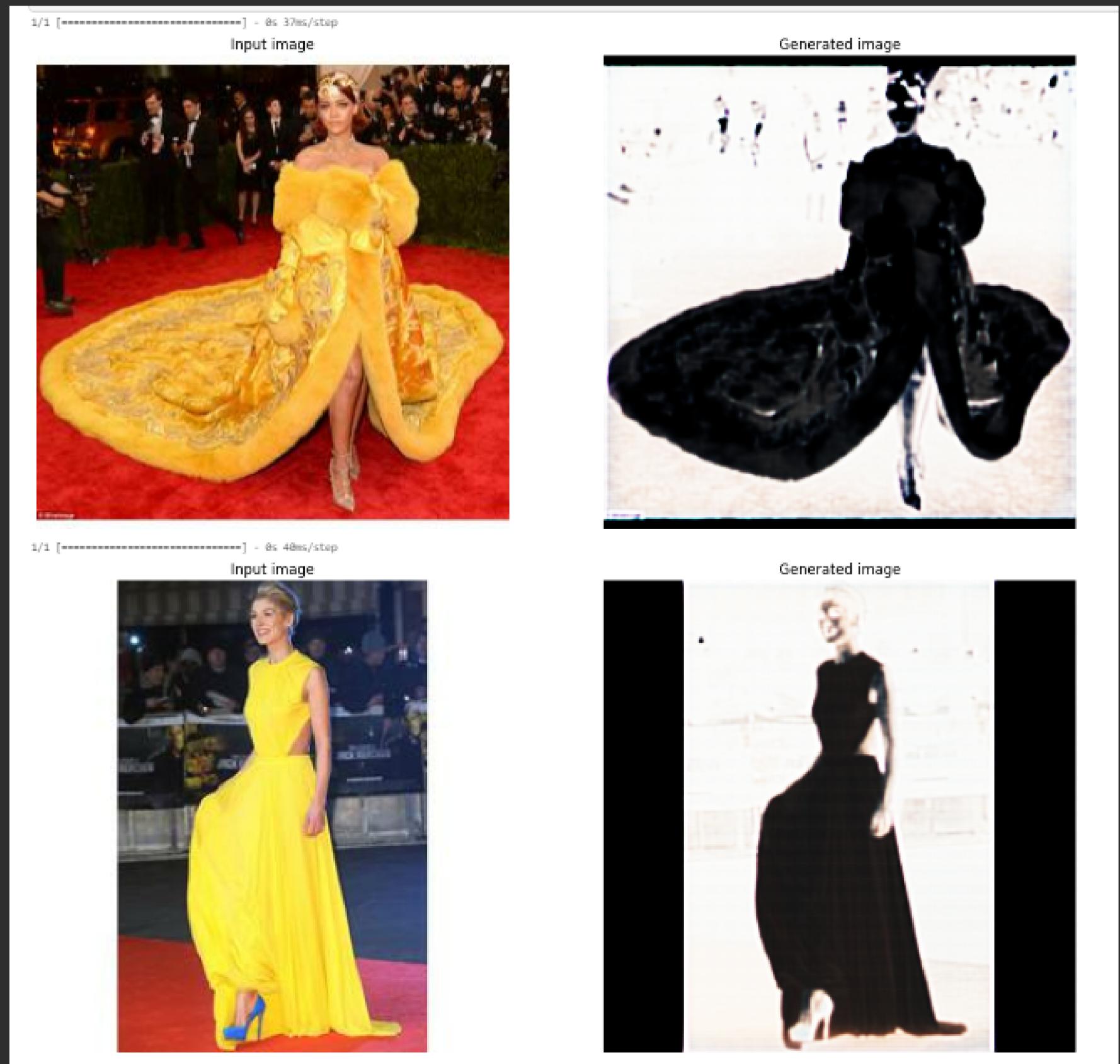
Failed to save draft.

Run Add-ons Help

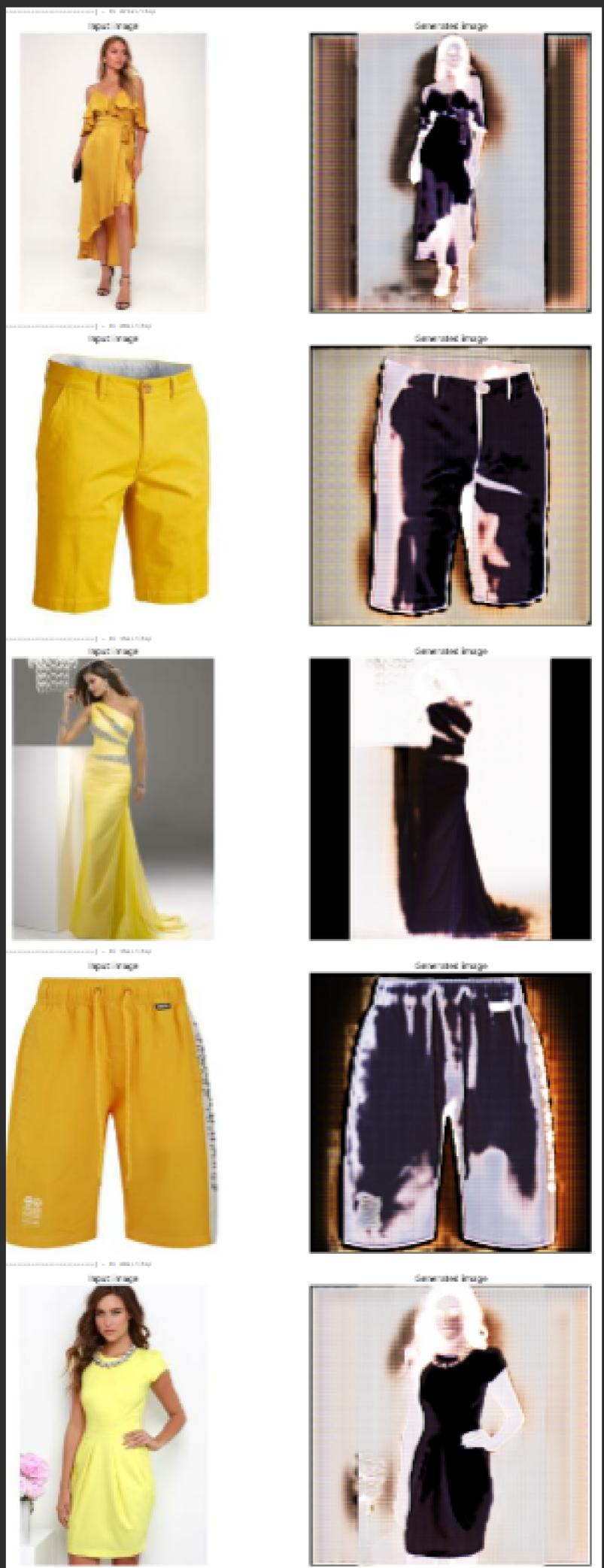
Cancel Run | Code | Draft Session (1h:0m) | H D S C P R A G O P U Q :

```
10/30
[=====] - ETA: 0s - gen_a2b_loss: 10.4466 - gen_b2a_loss: 10.2018 - disc_A_loss: 0.4217 - disc_B_loss: 0.4745-->Saved: g_model_AtoB_000010.h5, g_model_BtoA_000010.h5
[=====] - 48s 3s/step - gen_a2b_loss: 10.5606 - gen_b2a_loss: 10.2183 - disc_A_loss: 0.4368 - disc_B_loss: 0.4766 - lr: 0.0010
11/30
[=====] - ETA: 0s - gen_a2b_loss: 10.3311 - gen_b2a_loss: 10.2181 - disc_A_loss: 0.3855 - disc_B_loss: 0.4339-->Saved: g_model_AtoB_000011.h5, g_model_BtoA_000011.h5
[=====] - 48s 3s/step - gen_a2b_loss: 10.5267 - gen_b2a_loss: 10.2889 - disc_A_loss: 0.4114 - disc_B_loss: 0.4643 - lr: 0.0011
12/30
[=====] - ETA: 0s - gen_a2b_loss: 9.7215 - gen_b2a_loss: 9.8180 - disc_A_loss: 0.4350 - disc_B_loss: 0.3801-->Saved: g_model_AtoB_000012.h5, g_model_BtoA_000012.h5
[=====] - 48s 3s/step - gen_a2b_loss: 9.6317 - gen_b2a_loss: 9.7017 - disc_A_loss: 0.4500 - disc_B_loss: 0.3804 - lr: 0.0012
13/30
[=====] - ETA: 0s - gen_a2b_loss: 9.9963 - gen_b2a_loss: 9.9489 - disc_A_loss: 0.3987 - disc_B_loss: 0.3989-->Saved: g_model_AtoB_000013.h5, g_model_BtoA_000013.h5
[=====] - 48s 3s/step - gen_a2b_loss: 9.9514 - gen_b2a_loss: 10.0063 - disc_A_loss: 0.4025 - disc_B_loss: 0.4018 - lr: 0.0012
14/30
[=====] - ETA: 0s - gen_a2b_loss: 10.4697 - gen_b2a_loss: 10.0682 - disc_A_loss: 0.3553 - disc_B_loss: 0.2354-->Saved: g_model_AtoB_000014.h5, g_model_BtoA_000014.h5
[=====] - 48s 3s/step - gen_a2b_loss: 10.4400 - gen_b2a_loss: 10.0094 - disc_A_loss: 0.3477 - disc_B_loss: 0.2348 - lr: 0.0013
15/30
[=====] - ETA: 0s - gen_a2b_loss: 10.5847 - gen_b2a_loss: 9.9895 - disc_A_loss: 0.4749 - disc_B_loss: 0.3065-->Saved: g_model_AtoB_000015.h5, g_model_BtoA_000015.h5
[=====] - 48s 3s/step - gen_a2b_loss: 10.4475 - gen_b2a_loss: 9.8836 - disc_A_loss: 0.4894 - disc_B_loss: 0.3080 - lr: 0.0013
16/30
[=====] - ETA: 0s - gen_a2b_loss: 9.8073 - gen_b2a_loss: 9.6928 - disc_A_loss: 0.3592 - disc_B_loss: 0.3556-->Saved: g_model_AtoB_000016.h5, g_model_BtoA_000016.h5
[=====] - 48s 3s/step - gen_a2b_loss: 9.7671 - gen_b2a_loss: 9.6159 - disc_A_loss: 0.3697 - disc_B_loss: 0.3670 - lr: 0.0014
17/30
[=====] - ETA: 0s - gen_a2b_loss: 10.8036 - gen_b2a_loss: 10.4806 - disc_A_loss: 0.4041 - disc_B_loss: 0.4226-->Saved: g_model_AtoB_000017.h5, g_model_BtoA_000017.h5
[=====] - 48s 3s/step - gen_a2b_loss: 10.7441 - gen_b2a_loss: 10.4317 - disc_A_loss: 0.3941 - disc_B_loss: 0.4033 - lr: 0.0015
18/30
[=====] - ETA: 0s - gen_a2b_loss: 10.2075 - gen_b2a_loss: 10.1886 - disc_A_loss: 0.3837 - disc_B_loss: 0.3277-->Saved: g_model_AtoB_000018.h5, g_model_BtoA_000018.h5
[=====] - 48s 3s/step - gen_a2b_loss: 10.2271 - gen_b2a_loss: 10.2000 - disc_A_loss: 0.3882 - disc_B_loss: 0.3166 - lr: 0.0016
19/30
[=====] - ETA: 0s - gen_a2b_loss: 10.1301 - gen_b2a_loss: 10.1923 - disc_A_loss: 0.3873 - disc_B_loss: 0.3410-->Saved: g_model_AtoB_000019.h5, g_model_BtoA_000019.h5
[=====] - 48s 3s/step - gen_a2b_loss: 10.1722 - gen_b2a_loss: 10.2488 - disc_A_loss: 0.3991 - disc_B_loss: 0.3444 - lr: 0.0016
20/30
[=====] - ETA: 0s - gen_a2b_loss: 10.6383 - gen_b2a_loss: 10.0582 - disc_A_loss: 0.5598 - disc_B_loss: 0.3914-->Saved: g_model_AtoB_000020.h5, g_model_BtoA_000020.h5
[=====] - 48s 3s/step - gen_a2b_loss: 10.5799 - gen_b2a_loss: 10.0122 - disc_A_loss: 0.5463 - disc_B_loss: 0.3966 - lr: 0.0017
21/30
[=====] - ETA: 0s - gen_a2b_loss: 9.3640 - gen_b2a_loss: 9.5951 - disc_A_loss: 0.4591 - disc_B_loss: 0.4731-->Saved: g_model_AtoB_000021.h5, g_model_BtoA_000021.h5
[=====] - 48s 3s/step - gen_a2b_loss: 9.3516 - gen_b2a_loss: 9.5412 - disc_A_loss: 0.4616 - disc_B_loss: 0.4614 - lr: 0.0018
22/30
[=====] - ETA: 0s - gen_a2b_loss: 9.5810 - gen_b2a_loss: 9.9168 - disc_A_loss: 0.4330 - disc_B_loss: 0.3410-->Saved: g_model_AtoB_000022.h5, g_model_BtoA_000022.h5
[=====] - 48s 3s/step - gen_a2b_loss: 9.5008 - gen_b2a_loss: 9.9213 - disc_A_loss: 0.4231 - disc_B_loss: 0.3292 - lr: 0.0019
23/30
[=====] - ETA: 0s - gen_a2b_loss: 10.3370 - gen_b2a_loss: 10.5314 - disc_A_loss: 0.4175 - disc_B_loss: 0.5236-->Saved: g_model_AtoB_000023.h5, g_model_BtoA_000023.h5
[=====] - 48s 3s/step - gen_a2b_loss: 10.1831 - gen_b2a_loss: 10.5555 - disc_A_loss: 0.4141 - disc_B_loss: 0.5234 - lr: 0.0020
24/30
[=====] - ETA: 0s - gen_a2b_loss: 10.2257 - gen_b2a_loss: 10.0380 - disc_A_loss: 0.5322 - disc_B_loss: 0.4456-->Saved: g_model_AtoB_000024.h5, g_model_BtoA_000024.h5
[=====] - 48s 3s/step - gen_a2b_loss: 10.1294 - gen_b2a_loss: 10.0646 - disc_A_loss: 0.5247 - disc_B_loss: 0.4533 - lr: 0.0021
25/30
[=====] - ETA: 0s - gen_a2b_loss: 9.3743 - gen_b2a_loss: 9.2237 - disc_A_loss: 0.4885 - disc_B_loss: 0.3928-->Saved: g_model_AtoB_000025.h5, g_model_BtoA_000025.h5
[=====] - 48s 3s/step - gen_a2b_loss: 9.2945 - gen_b2a_loss: 9.1450 - disc_A_loss: 0.5000 - disc_B_loss: 0.3872 - lr: 0.0022
26/30
[=====] - ETA: 0s - gen_a2b_loss: 10.7055 - gen_b2a_loss: 10.3819 - disc_A_loss: 0.4238 - disc_B_loss: 0.3554-->Saved: g_model_AtoB_000026.h5, g_model_BtoA_000026.h5
[=====] - 48s 3s/step - gen_a2b_loss: 10.7151 - gen_b2a_loss: 10.4470 - disc_A_loss: 0.4150 - disc_B_loss: 0.3827 - lr: 0.0023
27/30
[=====] - ETA: 0s - gen_a2b_loss: 10.0911 - gen_b2a_loss: 10.0855 - disc_A_loss: 0.6436 - disc_B_loss: 0.4494-->Saved: g_model_AtoB_000027.h5, g_model_BtoA_000027.h5
[=====] - 48s 3s/step - gen_a2b_loss: 10.0237 - gen_b2a_loss: 9.9865 - disc_A_loss: 0.6337 - disc_B_loss: 0.4449 - lr: 0.0025
28/30
[=====] - ETA: 0s - gen_a2b_loss: 9.1725 - gen_b2a_loss: 9.3657 - disc_A_loss: 0.4848 - disc_B_loss: 0.3880-->Saved: g_model_AtoB_000028.h5, g_model_BtoA_000028.h5
[=====] - 48s 3s/step - gen_a2b_loss: 9.2056 - gen_b2a_loss: 9.4124 - disc_A_loss: 0.4817 - disc_B_loss: 0.3892 - lr: 0.0026
29/30
[=====] - ETA: 0s - gen_a2b_loss: 9.8966 - gen_b2a_loss: 9.6365 - disc_A_loss: 0.4248 - disc_B_loss: 0.3351-->Saved: g_model_AtoB_000029.h5, g_model_BtoA_000029.h5
[=====] - 48s 3s/step - gen_a2b_loss: 9.8154 - gen_b2a_loss: 9.5909 - disc_A_loss: 0.4092 - disc_B_loss: 0.3239 - lr: 0.0027
30/30
[=====] - ETA: 30s - gen_a2b_loss: 9.3465 - gen_b2a_loss: 9.5320 - disc_A_loss: 0.2915 - disc_B_loss: 0.4205
```

Error and Issues



Error and Issues



Error and Issues



Results



Results



Results



IMAGER - Image Editing Platform

Select Model

Red

Choose an image...



Drag and drop file here

Limit 200MB per file • JPG, JPEG

Browse files



11.jpg 16.2KB



Edit Image



Uploaded Image

Output Image

Code

Preprocessing --Training

```
import os
import pathlib
import time
import datetime
import imageio
from glob import glob

import tensorflow as tf
import numpy as np
import tensorflow.keras.backend as K
import tensorflow_addons as tfa

from matplotlib import pyplot as plt
from IPython import display
from termcolor import colored
from tqdm import tqdm
from IPython.display import Image
import PIL
from PIL import ImageDraw
from IPython import display
```

+ Code + Markdown

```
def color_print(print_str, print_color='green'):

    '''print in given color (default green)'''
    print(colored(print_str, print_color))

def set_seed(seed):
    np.random.seed(seed)
    tf.random.set_seed(seed)
    os.environ['PYTHONHASHSEED'] = str(seed)
    os.environ['TF_DETERMINISTIC_OPS'] = '1'
    print(f'setting seed to {seed}')

class CFG:
```

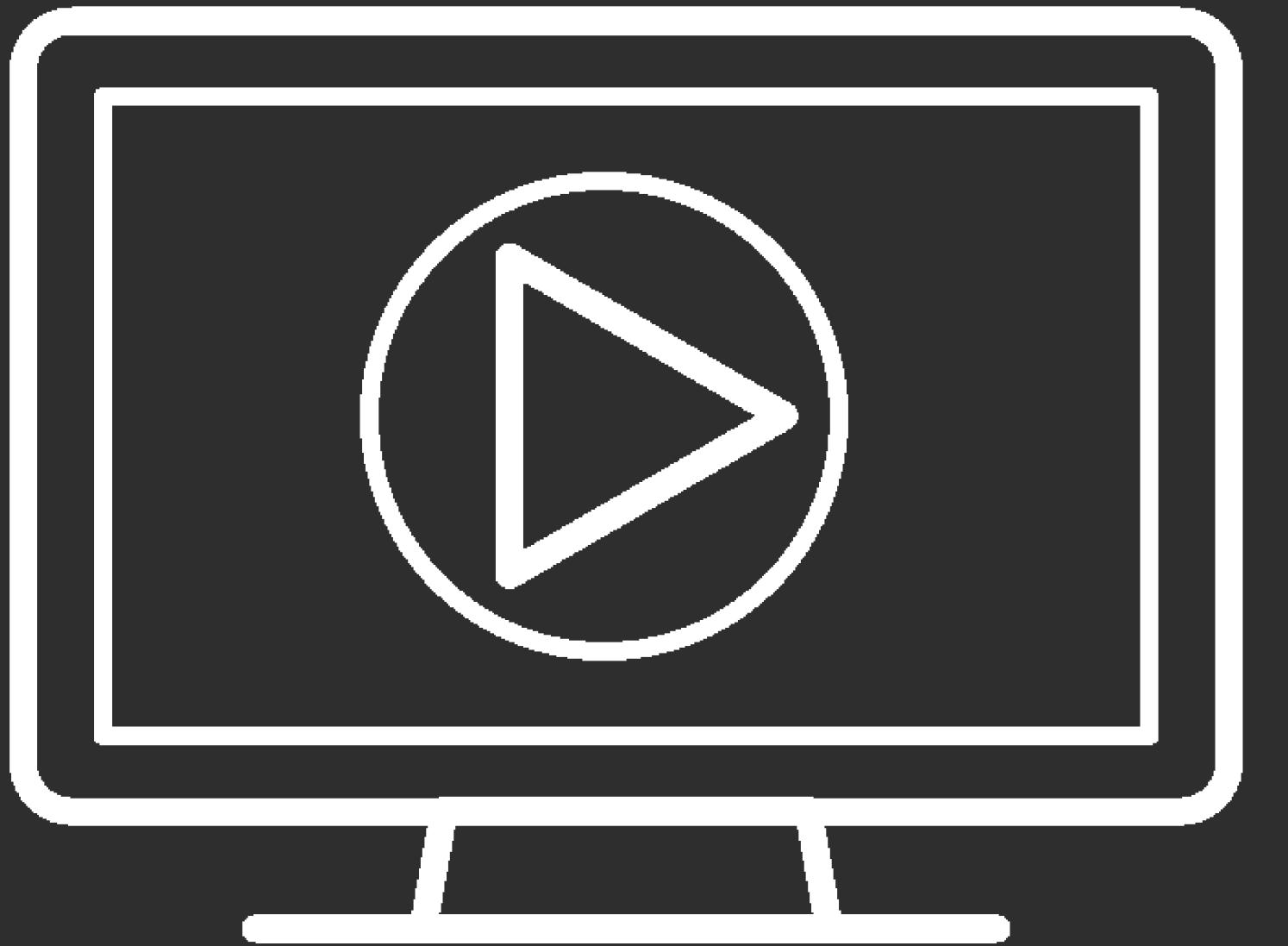
FrontEnd--Model Integration

```
20 st.title("IMAGER - Image Editing Platform")
21 # Initial model selection
22 selected_model_name = st.selectbox("Select Model", list(model_paths.keys()), index=1)
23 selected_model_path = model_paths[selected_model_name]
24 model_AtoB = load_model(selected_model_path, custom_objects=custom_objects)
25

26 def main():
27     # Upload image through Streamlit
28     uploaded_file = st.file_uploader("Choose an image...", type="jpg")
29
30     edit_btn = st.button("Edit Image")
31     col1, col2 = st.columns(2)
32
33     if uploaded_file is not None:
34         # Display the uploaded image
35         with col1:
36             st.image(uploaded_file, caption="Uploaded Image", use_column_width=True)
37
38         if edit_btn:
39             # Send image to the deep learning model's API endpoint
40             response = handle_file_upload(uploaded_file)
41
42             # Display the output image
43             with col2:
44                 st.image(response, caption="Output Image", channels='RGB', clamp=True, use_column_width=True)
45
46     def handle_file_upload(uploaded_file):
47         # Prepare the image data
```



Demo





Team Work



USMAN

HASEEB

SARIM

TIMELINE

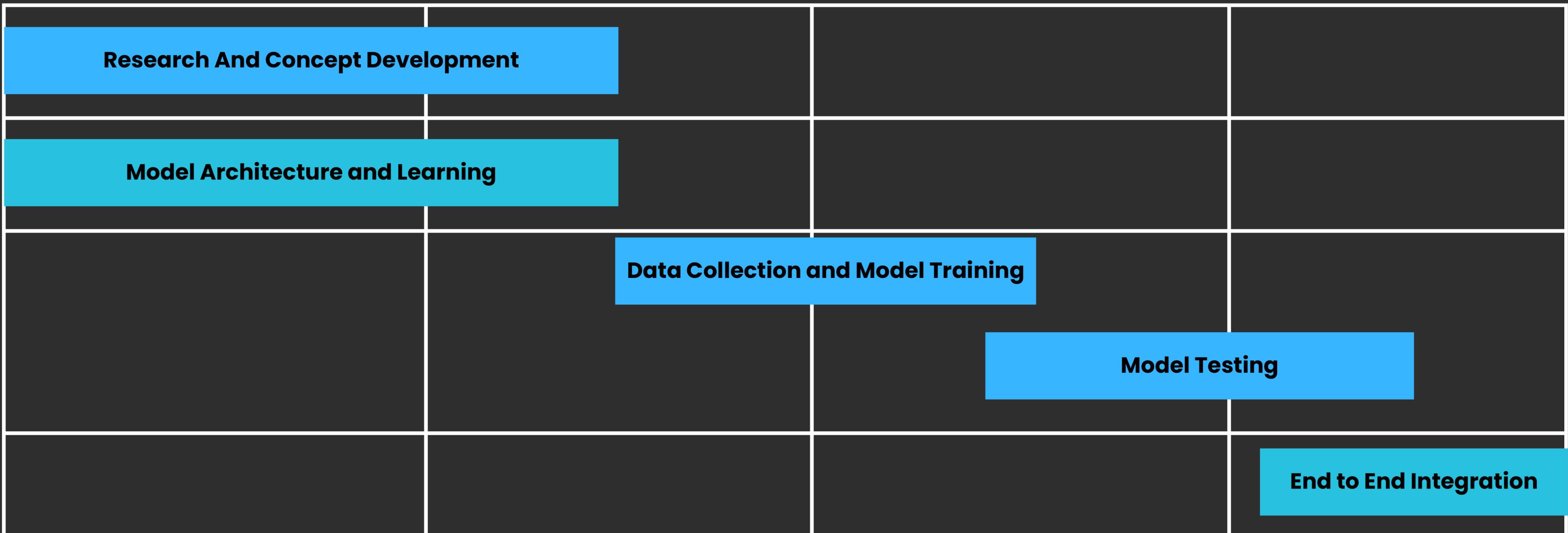


1st Quarter

2nd Quarter

3rd Quarter

4th Quarter





References

- [1] Zhu, J.Y., Park, T., Isola, P. and Efros, A.A., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).
- [2] Ravuri, Suman, and Oriol Vinyals. "Classification accuracy score for conditional generative models." Advances in neural information processing systems 32 (2019).
- [3] Pernuš, Martin, et al. "FICE: Text-Conditioned Fashion Image Editing With Guided GAN Inversion." arXiv preprint arXiv:2301.02110 (2023).
- [4] Kawar, Bahjat, et al. "Imagic: Text-based real image editing with diffusion models." arXiv preprint arXiv:2210.09276 (2022).
- [5] Kwon, Gihyun, and Jong Chul Ye. "Clipstyler: Image style transfer with a single text condition." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022.
- <https://www.kaggle.com/datasets/sunnykusawa/tshirts>
- <https://www.kaggle.com/datasets/kaiska/apparel-dataset>