Using GANs and Encoders/Decoders for Image-to-Sketch Translation

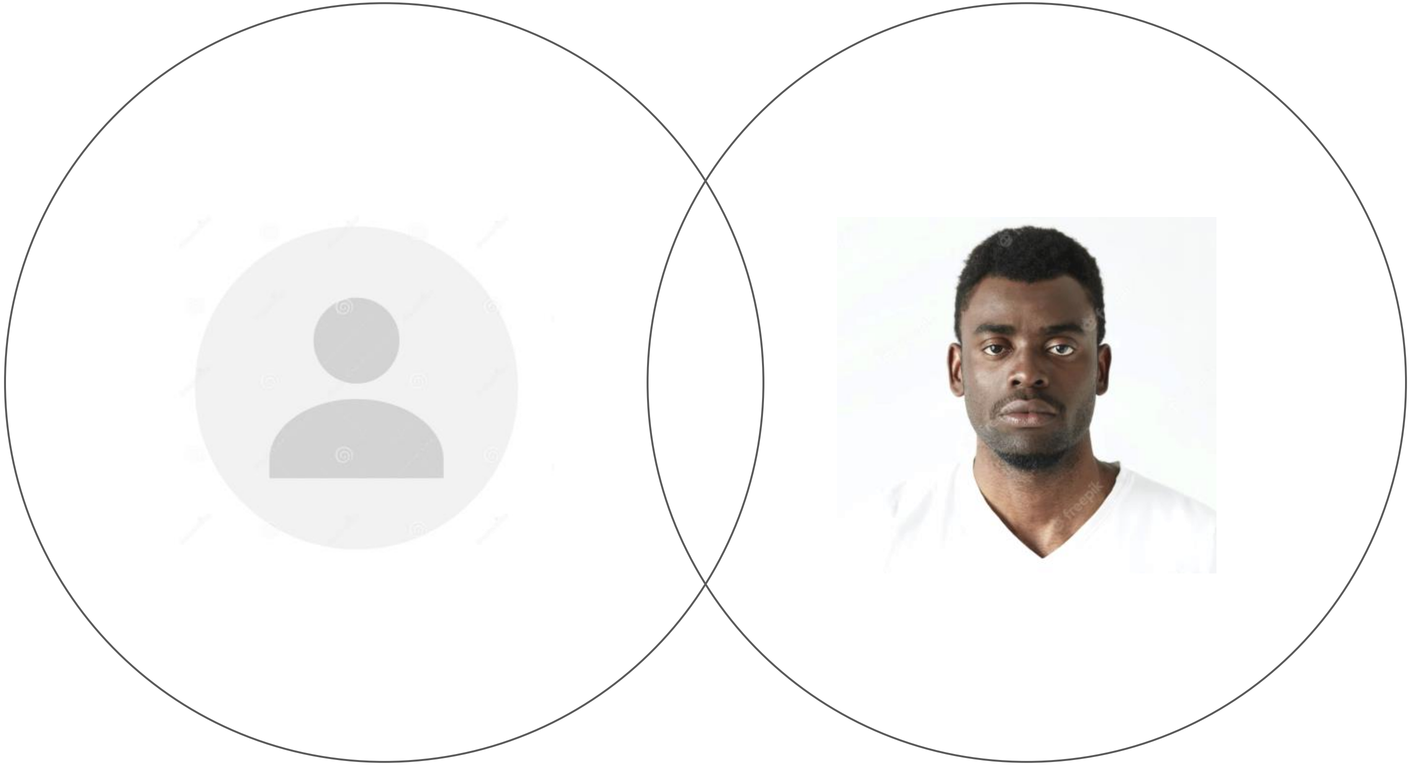
15.S04: Hands on Deep Learning

Jan Reig Torra, Shurui Cao, Xinyao Han, Daniel Chung

Introduction

Preservation of Privacy

Accurate Expression of Self



Memoji

Apple

Messaging and
communication

Digital People

Soul Machine

AI customer
service, virtual
assistant

Digital Humans

Uneeq

Frontend application
for chatbots

Objective: Convert Photographs to Sketches

Goal is to automate avatar creation within 2 dimensions, limited to grayscale

facial photograph

deep learning model

corresponding sketch

Data

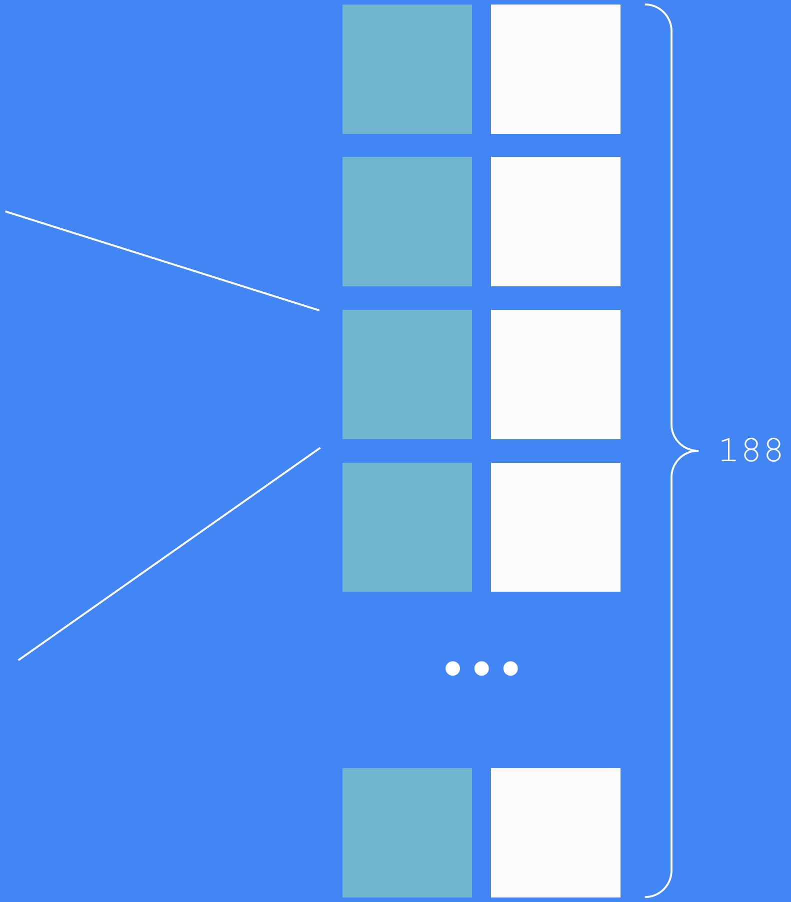
Dataset

photograph

shape: (256, 256, 3)

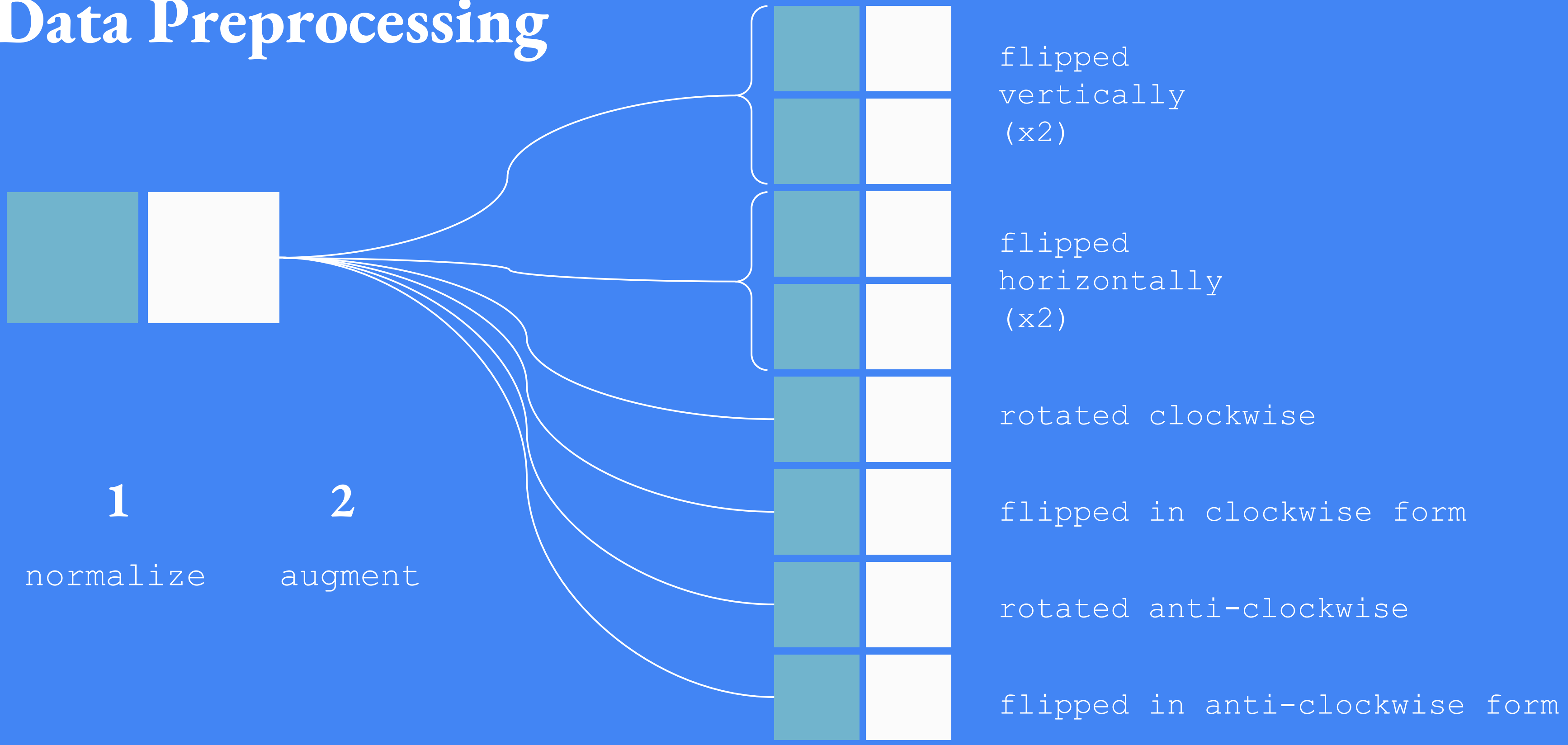
sketch

shape: (256, 256, 3)



Data collected from Chinese University of Hong Kong, publicly available on Kaggle!

Data Preprocessing



1

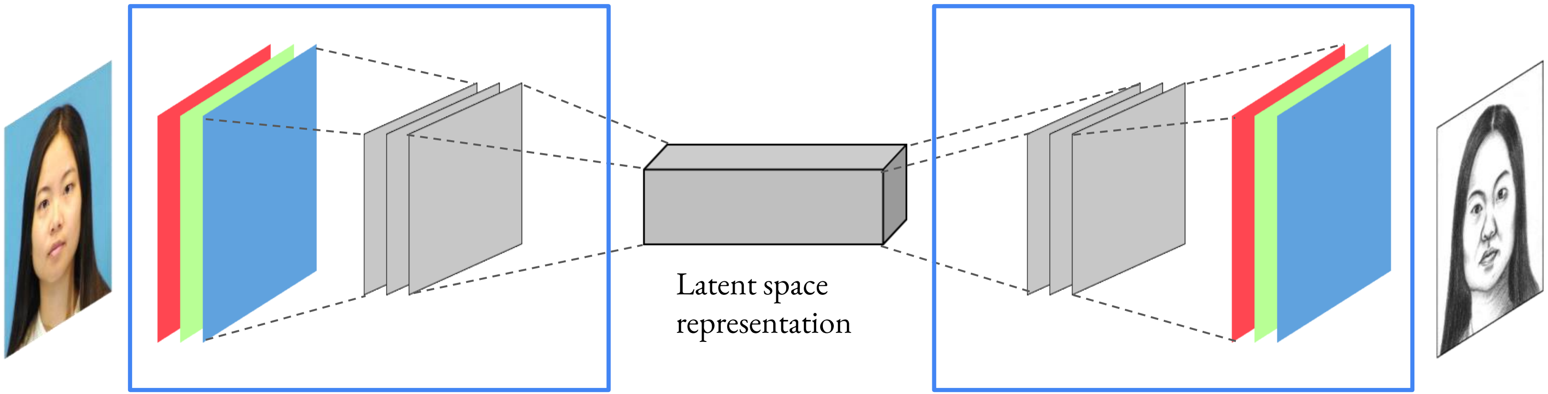
normalize

2

augment

Methodology

Method 1: Encoders / Decoders



Encoders

- Extract meaningful features & create a compressed representation
- A series of convolutional layers that capture hierarchical patterns and structures in the input image

Decoders

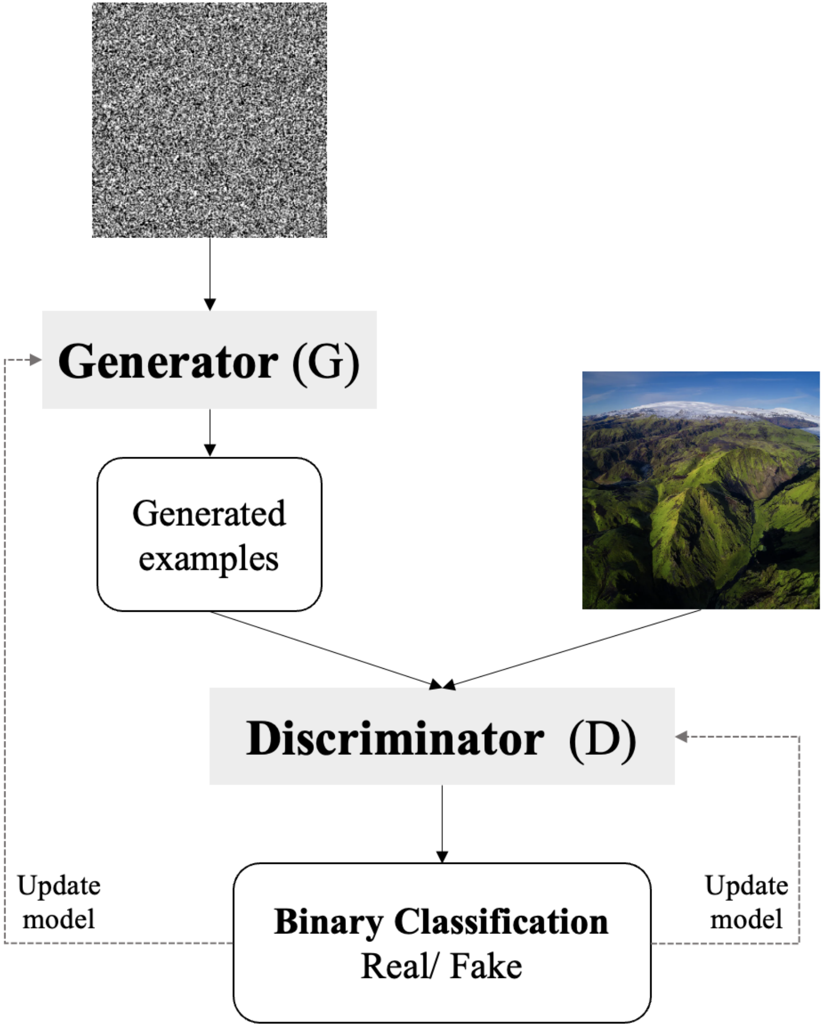
- Take the compressed representation and reconstruct a new image based on the extracted features
- A series of upsampling and convolutional layers to gradually increase the spatial resolution of the output image

Method 2: cycleGAN

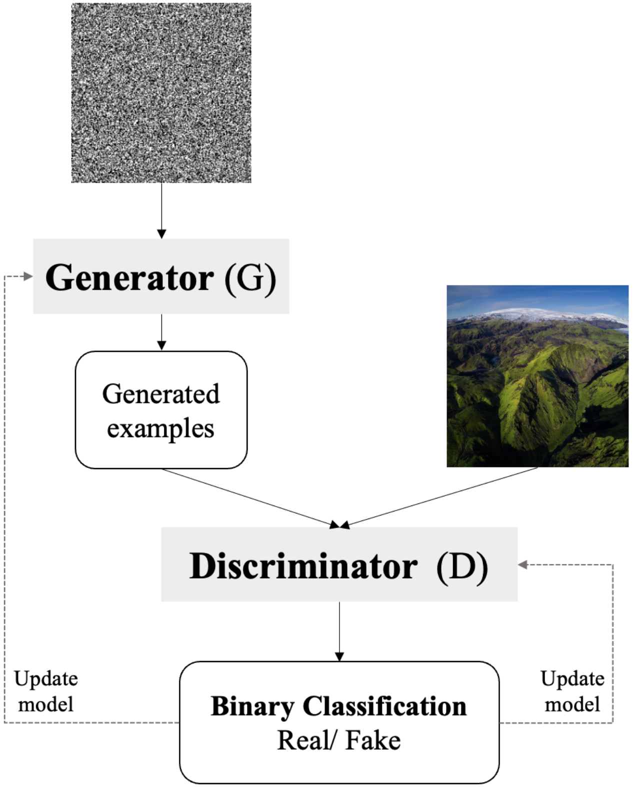
GAN (Generative Adversarial Networks)

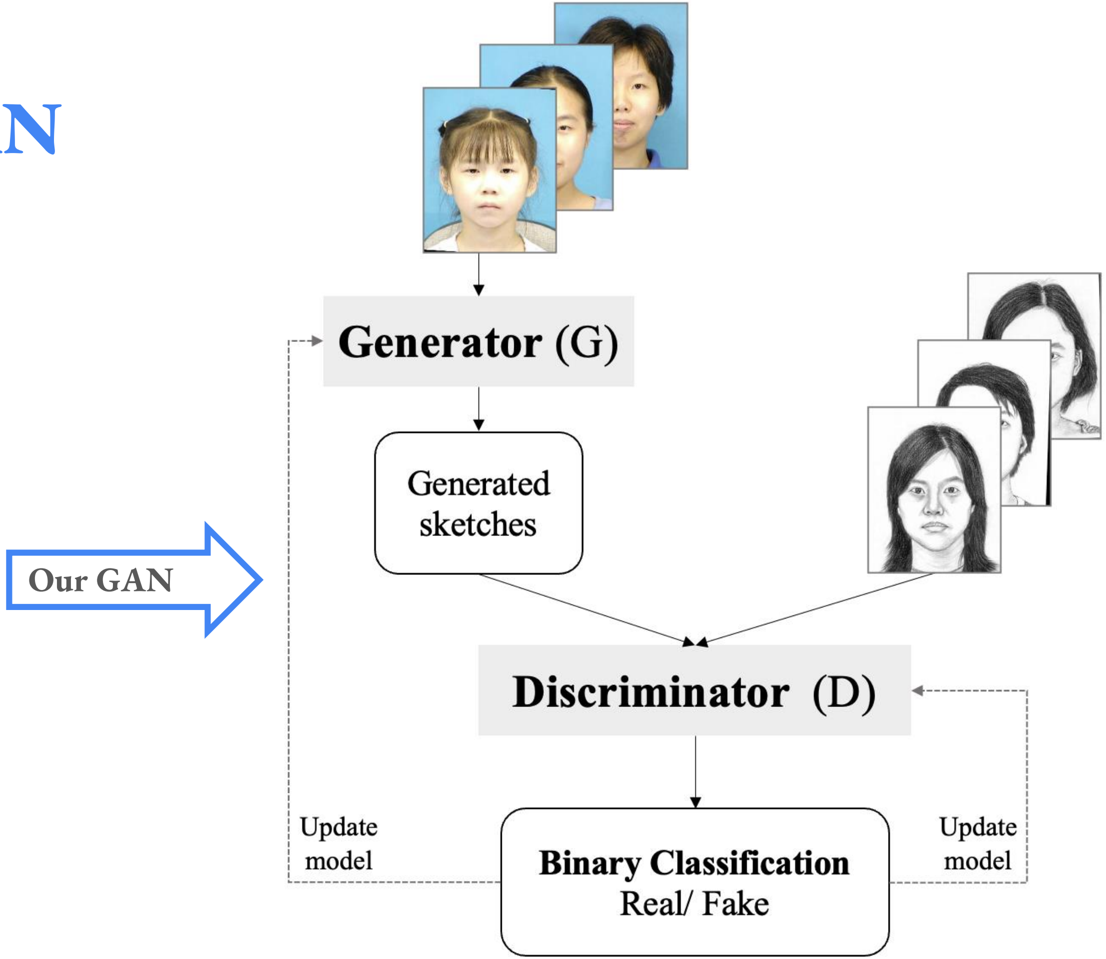
Generate images

- **GENERATOR** → Generates synthetic images
- **DISCRIMINATOR**
  - **Input** → Real image vs Synthetic images
  - Evaluates the authenticity of the generated data
- Trained simultaneously through **adversarial training**
- **Backpropagation** → update weights



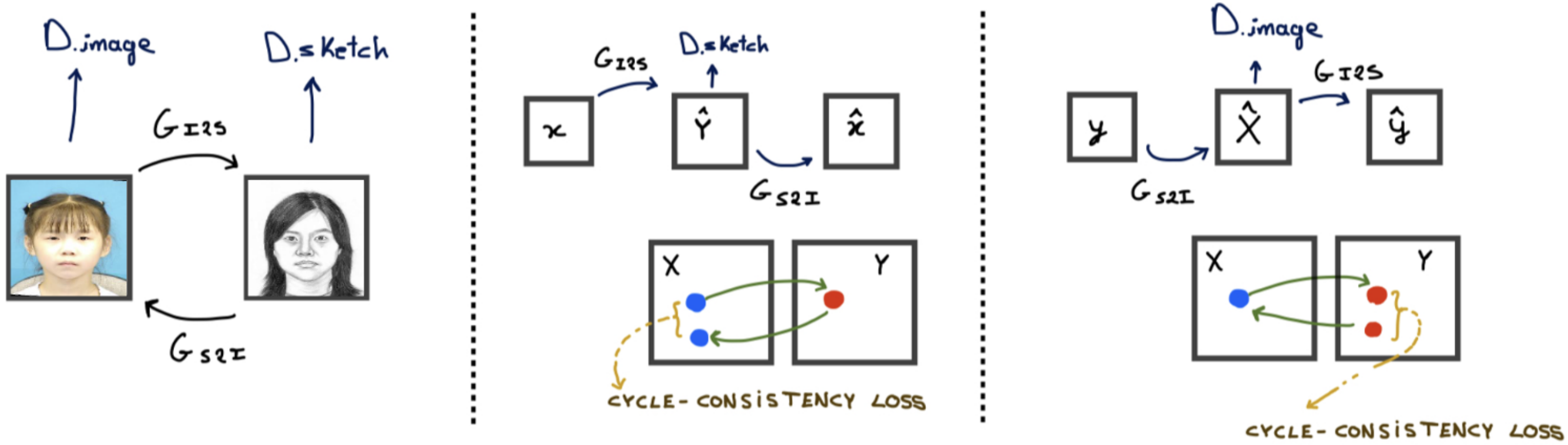
Method 2: cycleGAN





Method 2: cycleGAN

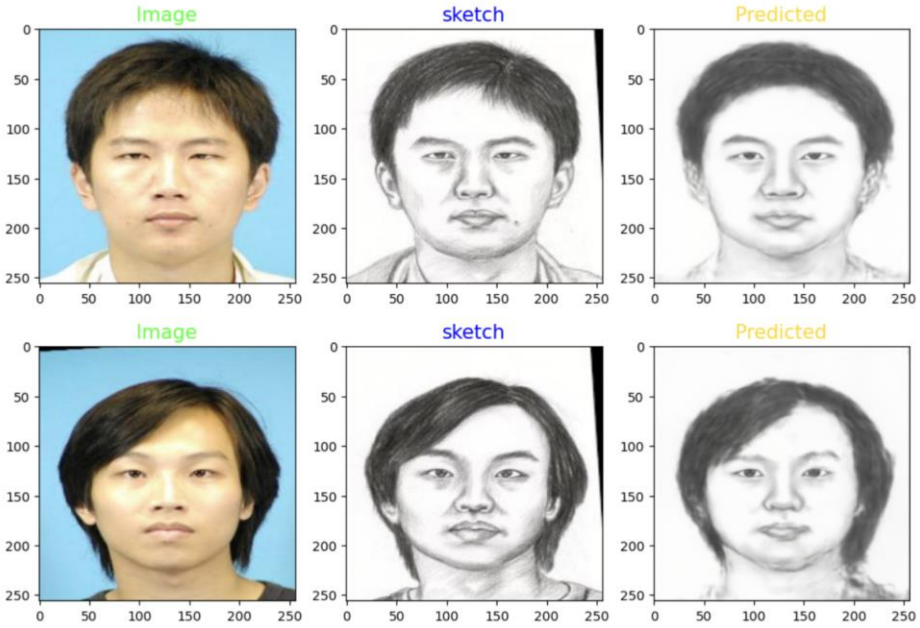
- Unpaired image-to-image translation
- **2 GANs** → (**I2S**) GAN is responsible for generating sketches from images *and* (**S2I**) the other GAN generates images from sketches



- **Consistency loss** → Enforces the constraint that **if an image is translated from one domain to another and then back again**, it should closely **resemble the original image**

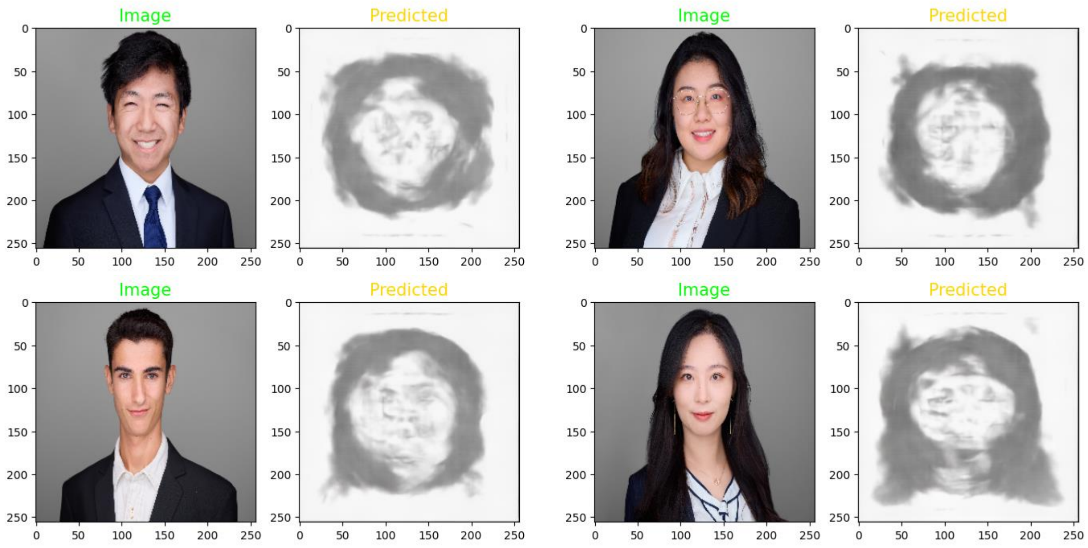
Results

Result I - Encoder & Decoder



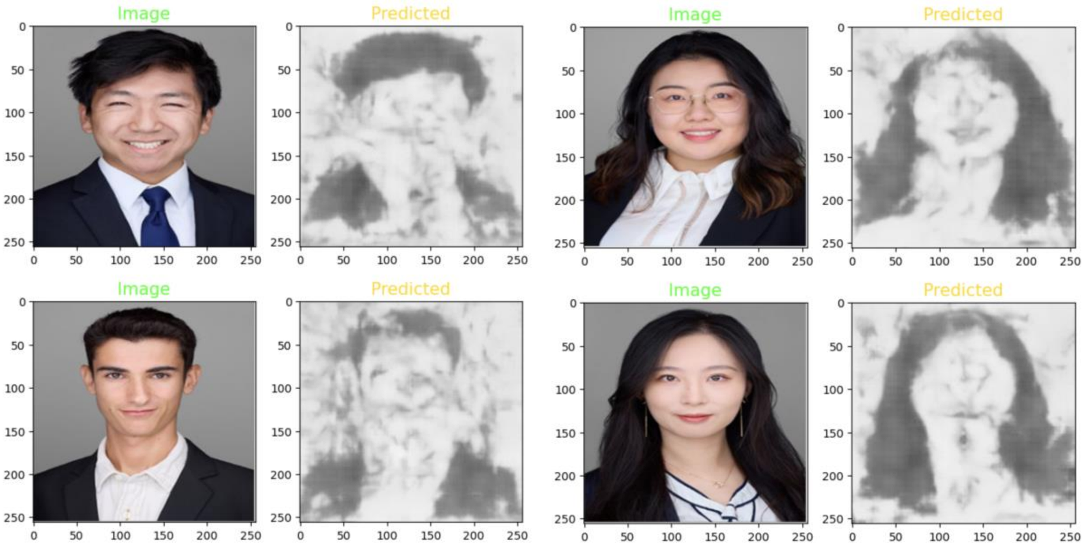
- Predictions have high resemblance to the original sketches.
- While some details of brushstrokes are lost, the model is able to reconstruct the original human faces clearly.

Result II - Encoder & Decoder (Incomplete Training)



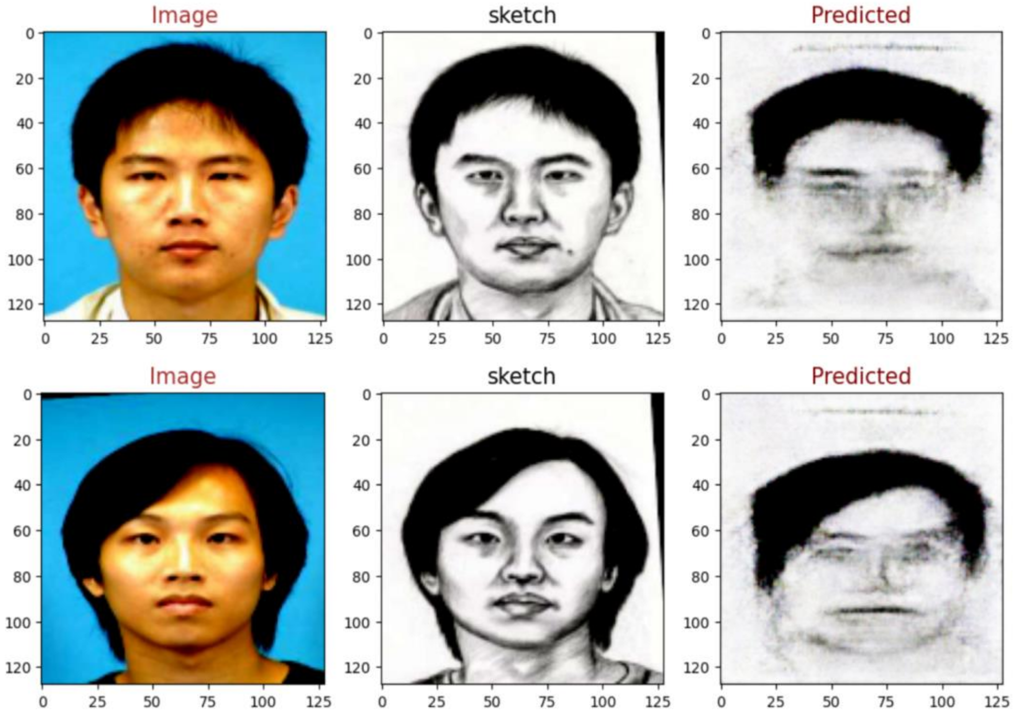
- Model struggles to capture specificities for each face, seems to output an averaged representation of all faces in its training data

Result II - Encoder & Decoder (Complete Training)



- Model captures the most obvious features of the photos, for example, the area with the darkest colors like hair and clothes.
- It fails to capture the details in the face area.

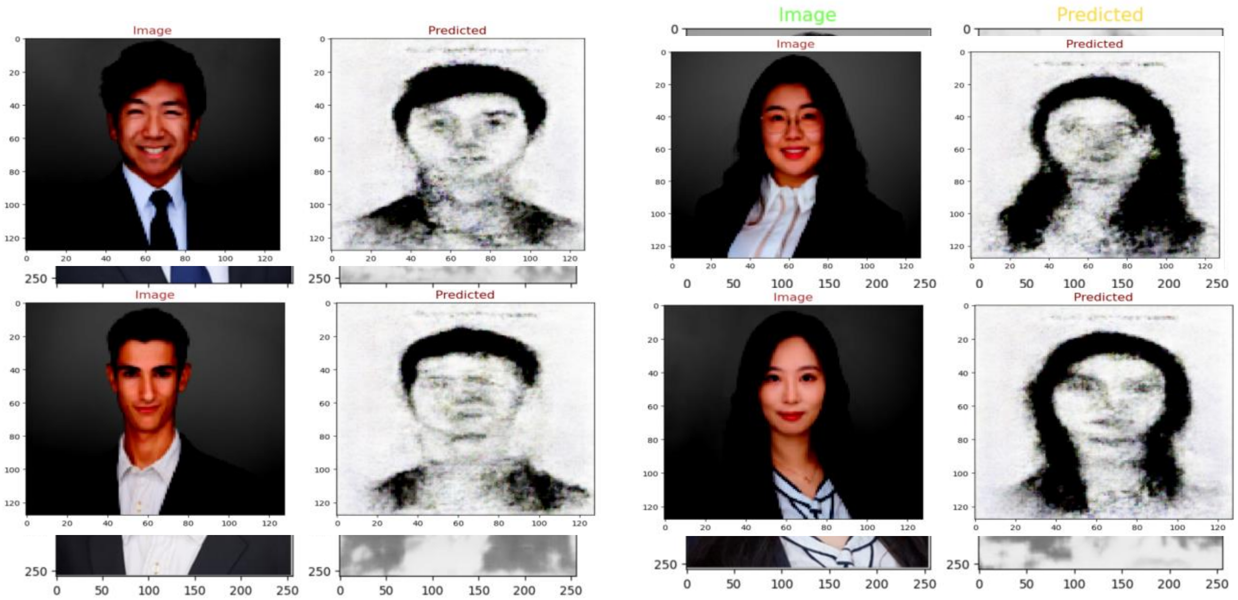
Result III - cycleGAN



- Sketches predicted from random samples with the generator G_{I2S}
- Blurred
- Worse off compared with Encoder/Decoder methods - abstract facial features & hair shape

Result IV - Best Model for new images (cycleGAN)

photograph sketch photograph sketch



- New image with different styles and lighting conditions
- Better off compared to Encoder/Decoder method
- Slightly worse off compared to same testing dataset for cycleGAN

Result V - Sketches to Images Generation

- Images predicted from random sketches with a different generator G_{S2I}
- blurred

Lessons Learned

Overfitting

Lack of diversity in the training dataset (style, race, lightening...)

Insufficient training data and training time