DeepFaceDrawing

Mukul Kumar Vishwas

Novosibirsk State University, Russia

SHU-YU CHEN, WANCHAO SU, LIN GAO, SHIHONG XIA, HONGBO FU

24 November 2020

Problem statement

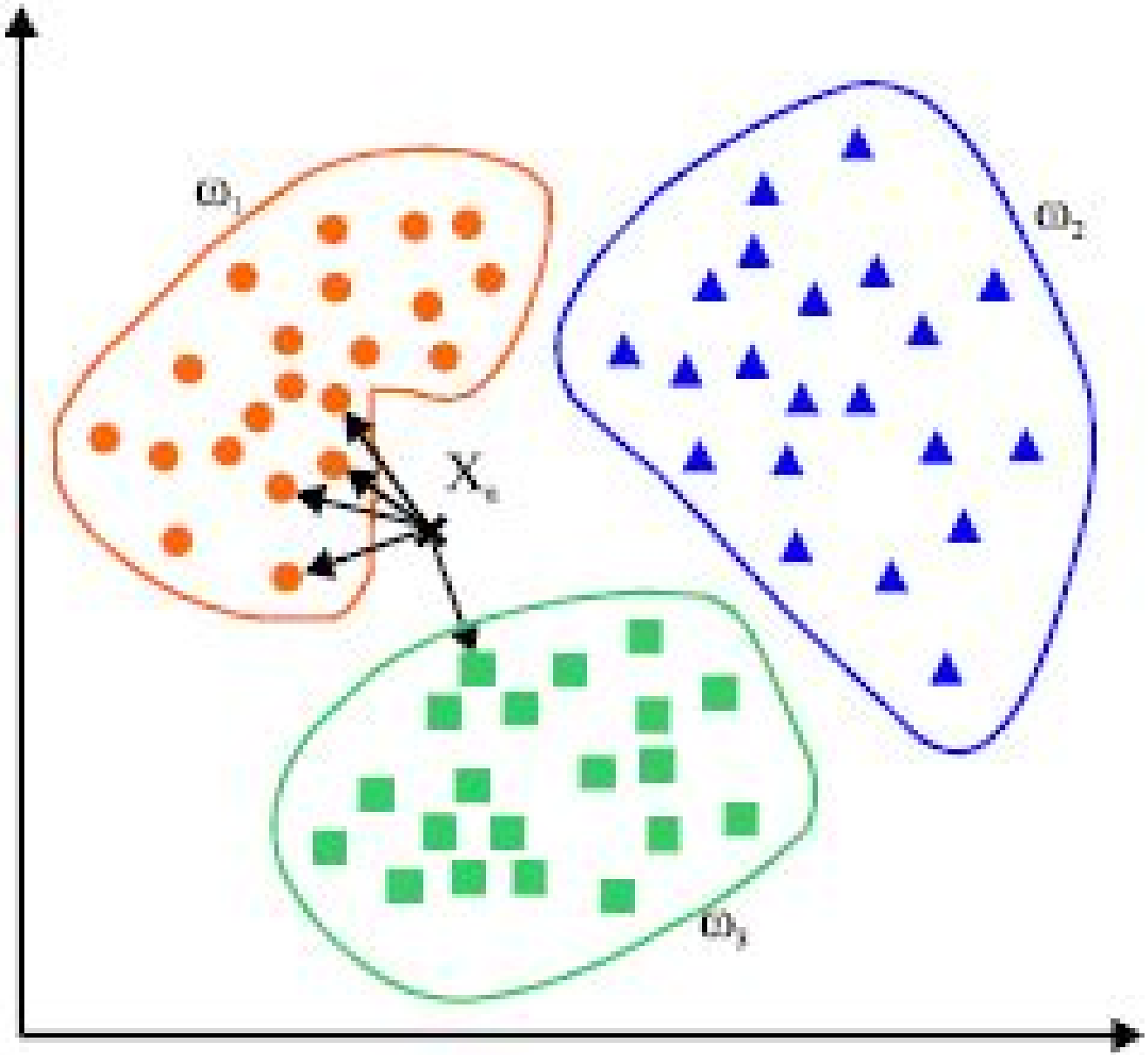
Image generating models tend to depend on sketches, thus requiring professional sketches or even edge maps as input which is hard constrains.

Solution: Key idea is to implicitly learn a space of plausible face sketches (Manifold Projection) from real face sketch images and find the closest point in this space to approximate an input sketch. In this way, sketches can be used more like soft constraints to guide image synthesis.

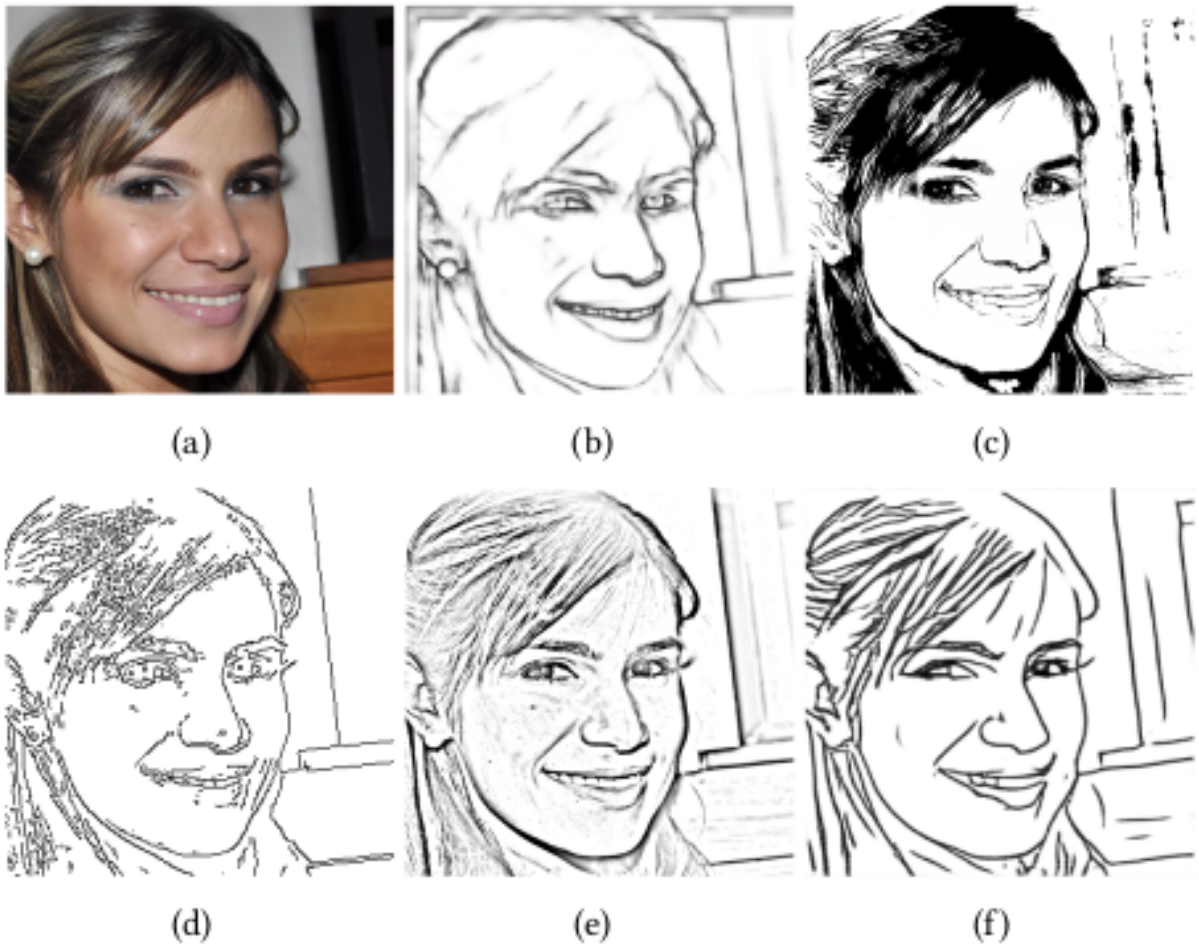
Introduction

- Allow user to draw high-quality face image with little training in drawing.
- From rough or even incomplete freehand sketches.
- Method faithfully respects user intentions in input strokes.
- It serve more like soft constraints to guide image synthesis.

Example of shape space

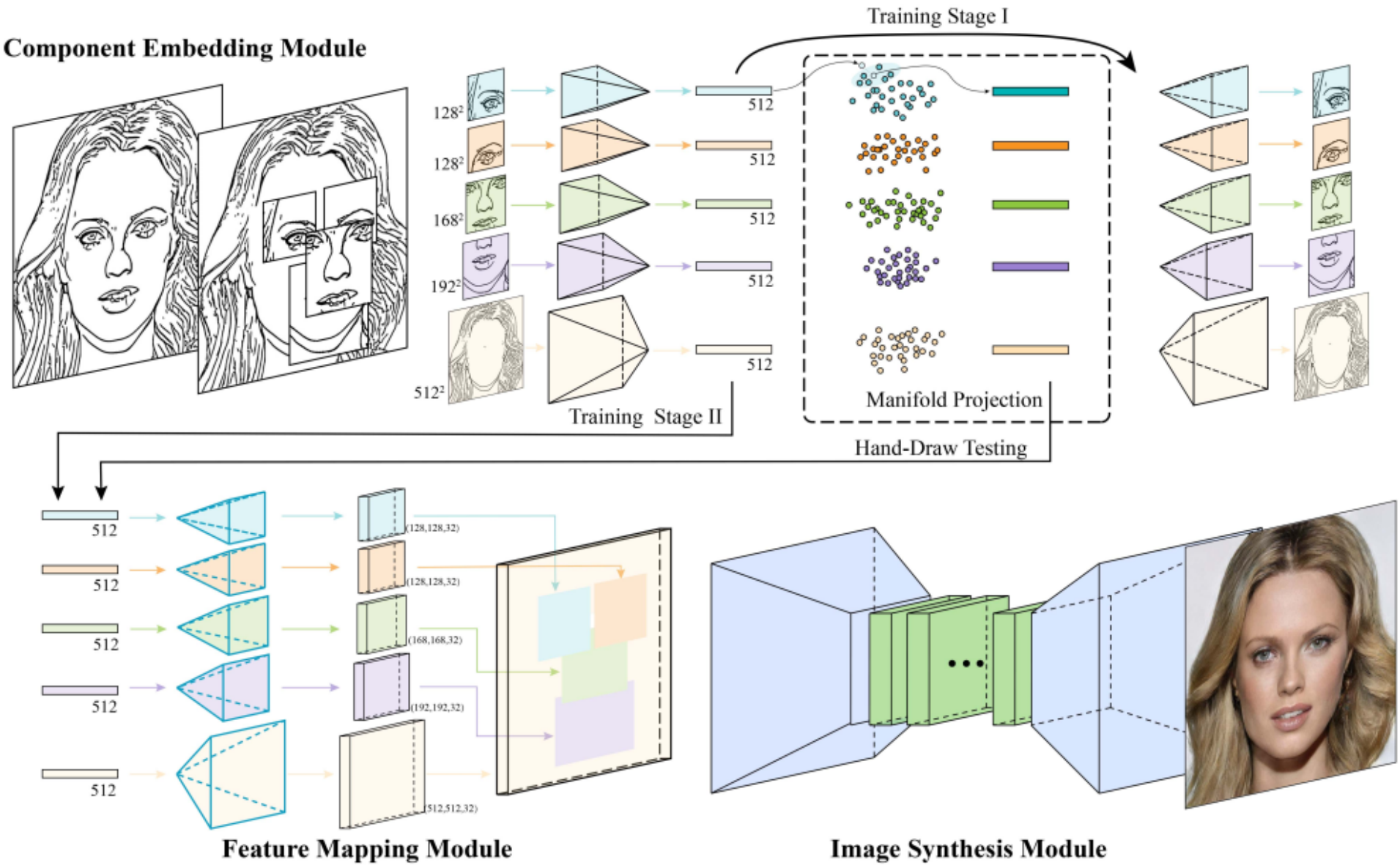


Training Data



a) original, b) Holistically-nested Edge detection c) APDrawingGAN d) Canny edges e) Photocopy filter in Photoshop f) Photocopy + sketch simplification)

Architecture

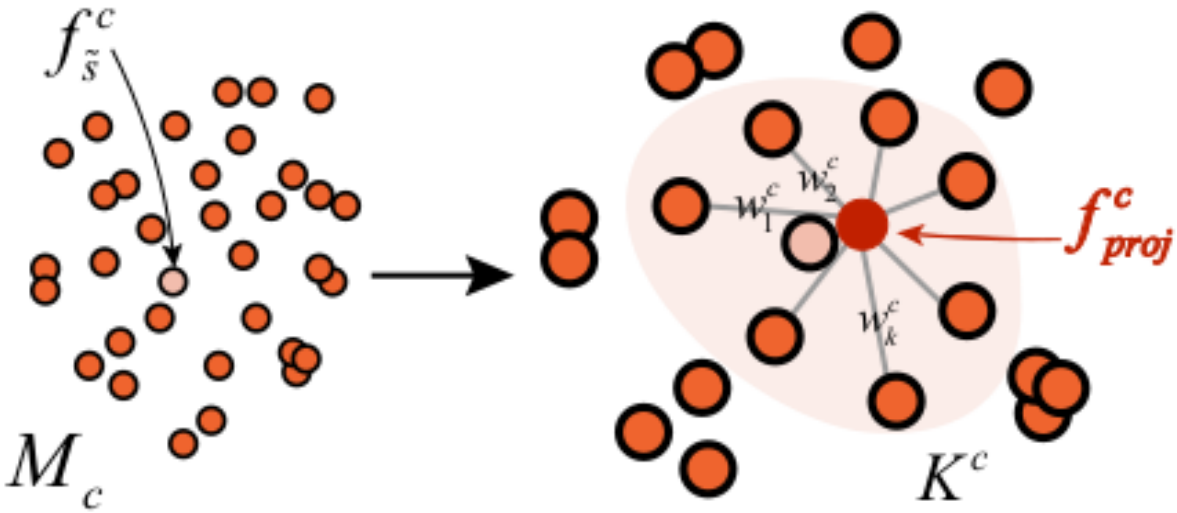


Architecture

- hand draw sketch as input and generate a 512*512 high resolution facial image.
- divided into 3 module a) Component Embedding b) Feature Mapping c) Image Synthesis.
- CE is responsible for learning feature embedding of individual face component.
- All human face share common feature hence the encoder divided into 5 sub-parts.

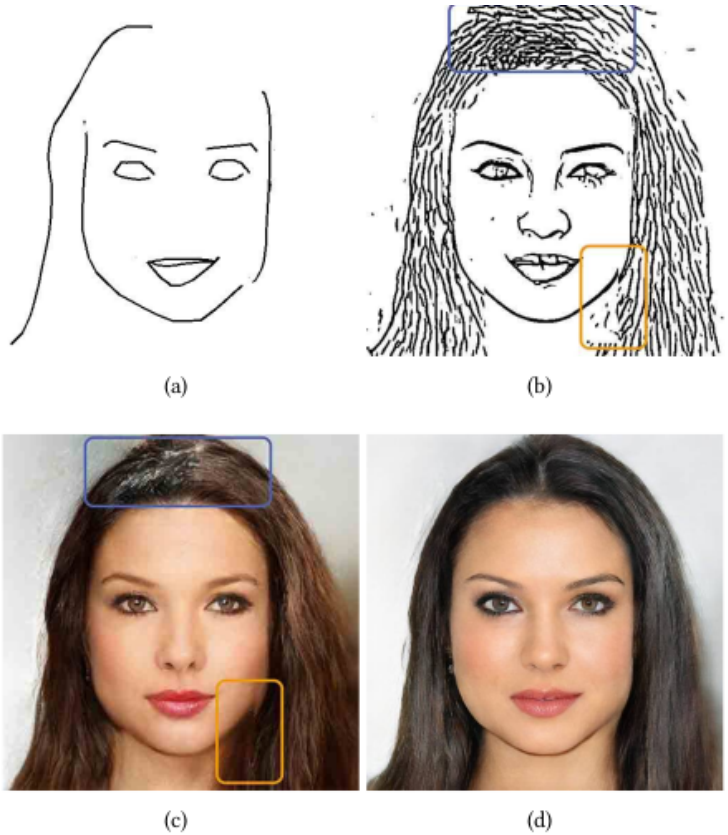
  a) Left Eye b) Right Eye c) Nose d) Mouth e) Remimder

Architecture (CE)



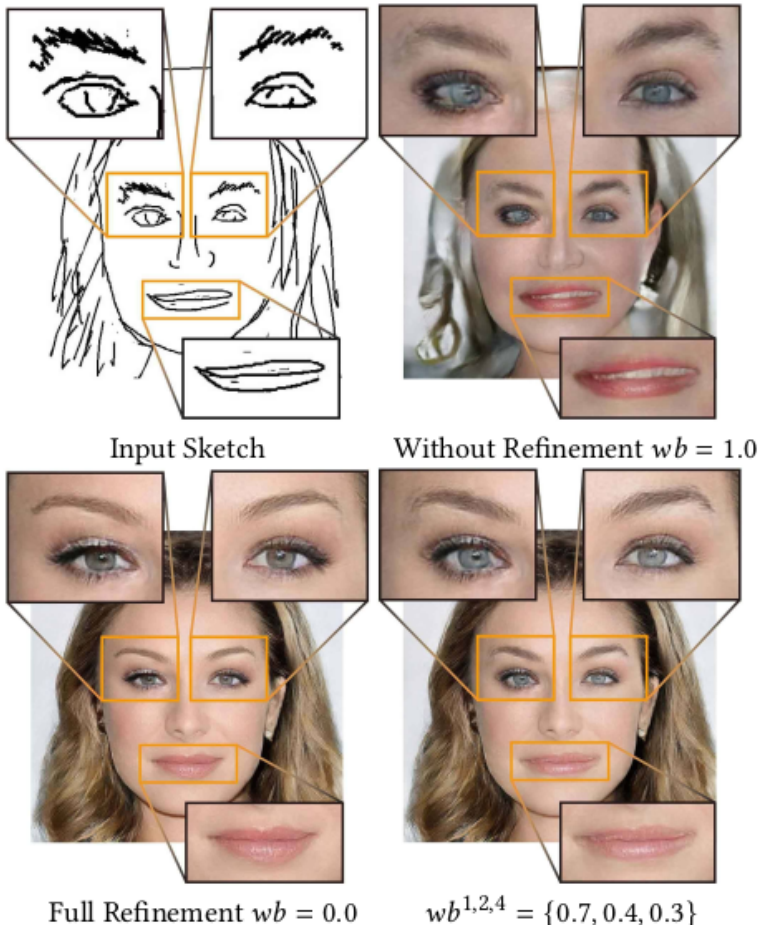
- Each auto-encoder consist of 5 encoder and decoder.
- features are converted into 512 dimensions for each.
- Loss: Mean squared error (MSE)
- In manifold projection KNN (K=10) used.

Architecture(FM)



- Individual encoders project the vectors to manifold.
- the intersection mapping are the most error prone.
- The FM can be done in 3 way the most success-full way is converting dimensions to 3D tensor.

Architecture(IS)



- Conditional GAN archetecture.
- generator-discriminator works in multi scale manner to learn high level co-relation among different face parts.

Training

- Model train in 2 stages.

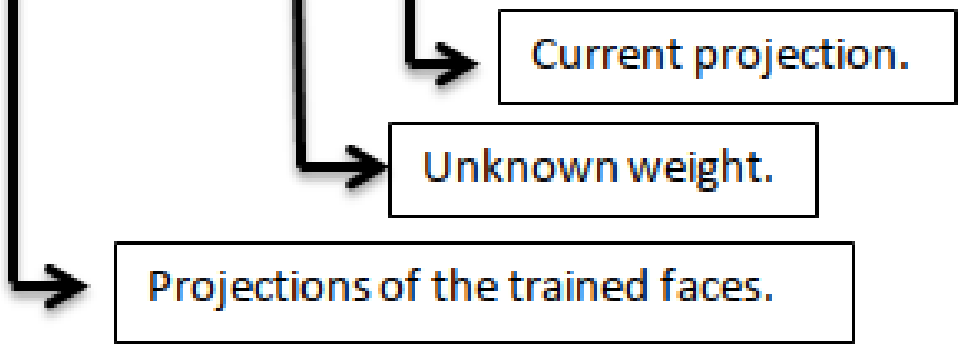
  a) Stage-1 only CE is trained.

  b) Stage -2 full model trained.

- Trained model with all training sample but generation of male and female face is separated.

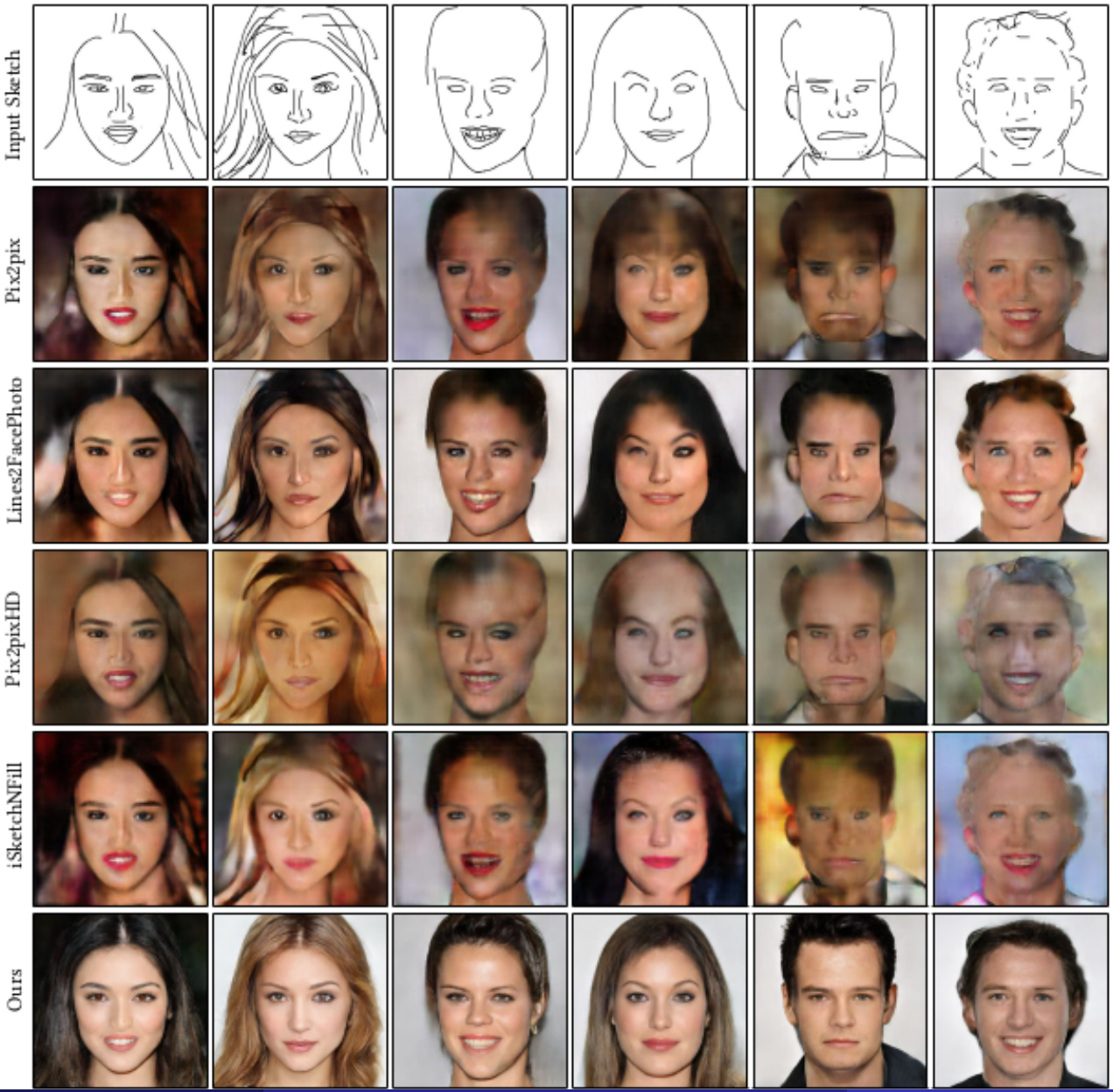
Reconstruction Error/Loss

$$\min ||f_{\tilde{s}}^{c} - \sum_{k \in \mathcal{K}^c} w_k^c \cdot f_k^c||_2^2, \quad s.t. \sum_{k \in \mathcal{K}} w_k^c = 1,$$



- Minimize the distance between model projections and current projection.
- CE use MSE loss.
- IS use GAN loss.

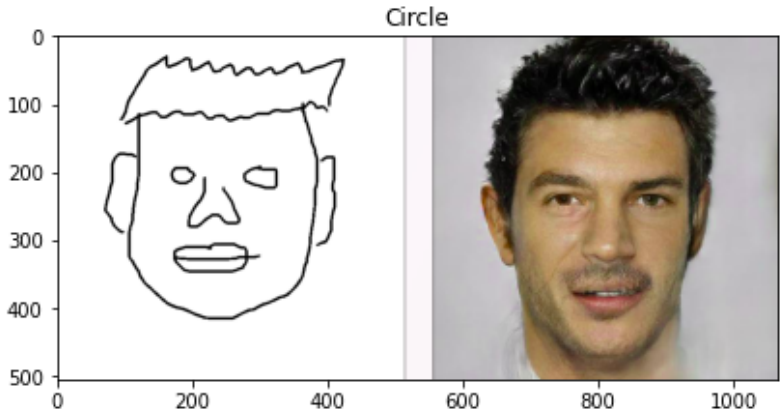
Result comparision

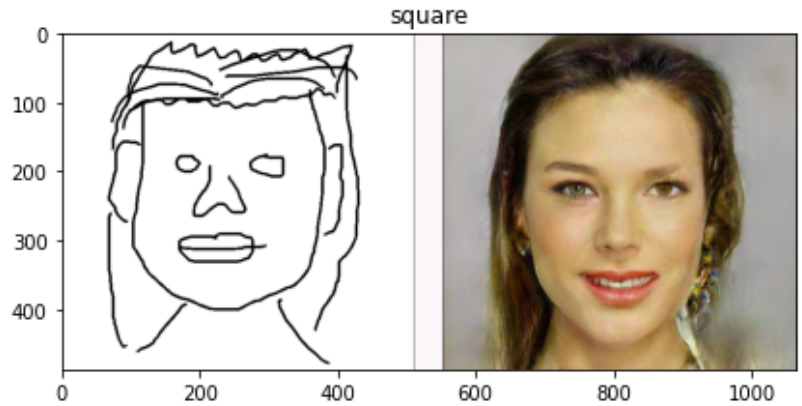


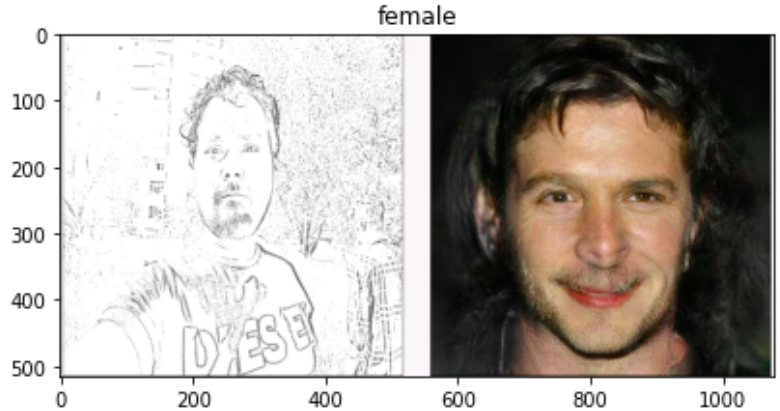
My Result

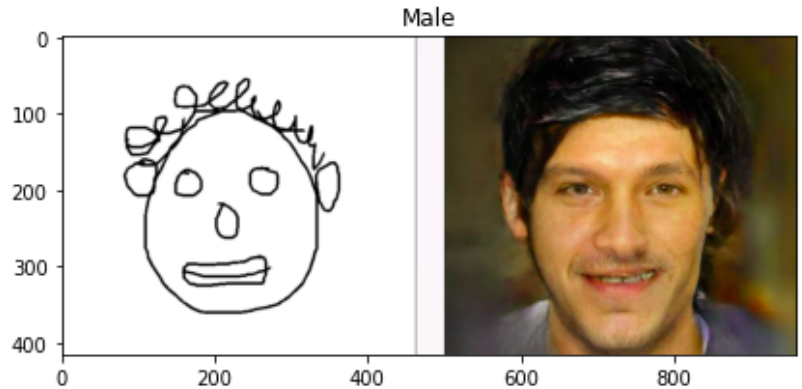
```
plt.imshow(img4)
plt.title('Result3')
```

```
Out[7]: Text(0.5, 1.0, 'Result3')
```

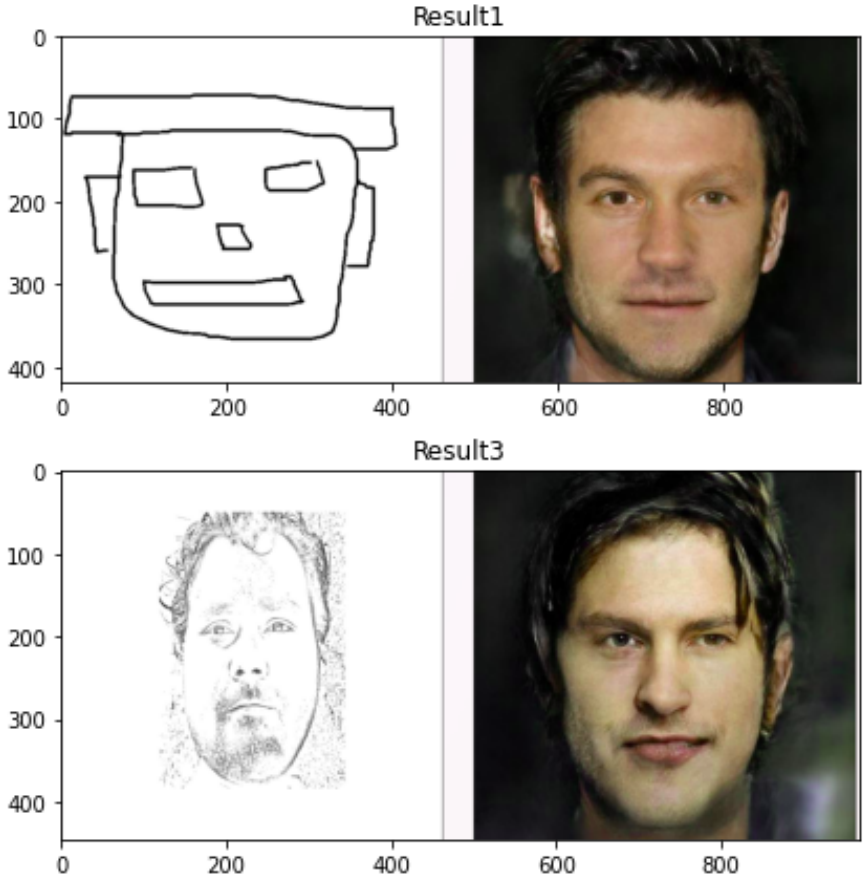


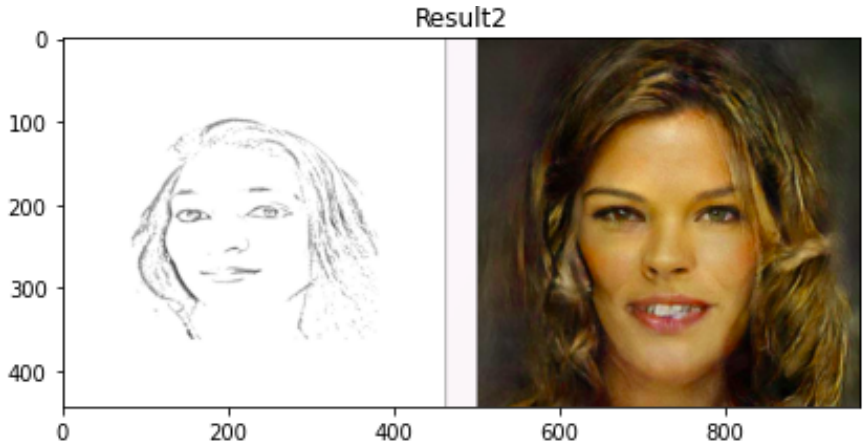






My Result


Result1
0
100
200
300
400
0
200
400
600
800


Result2
0
100
200
300
400
0
200
400
600
800

Result3
0
100
200
300
400
0
200
400
600
800

Thank you for your time.