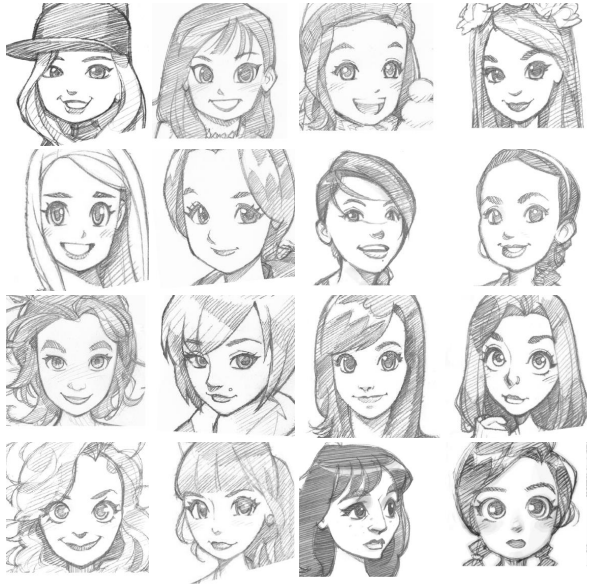
Pencil Cartoonized Portrait by Style Transfering with JoJoGAN

Poompich Eiabtawee
6310412016

Goal

Style dataset



INPUT

Portrait

Style

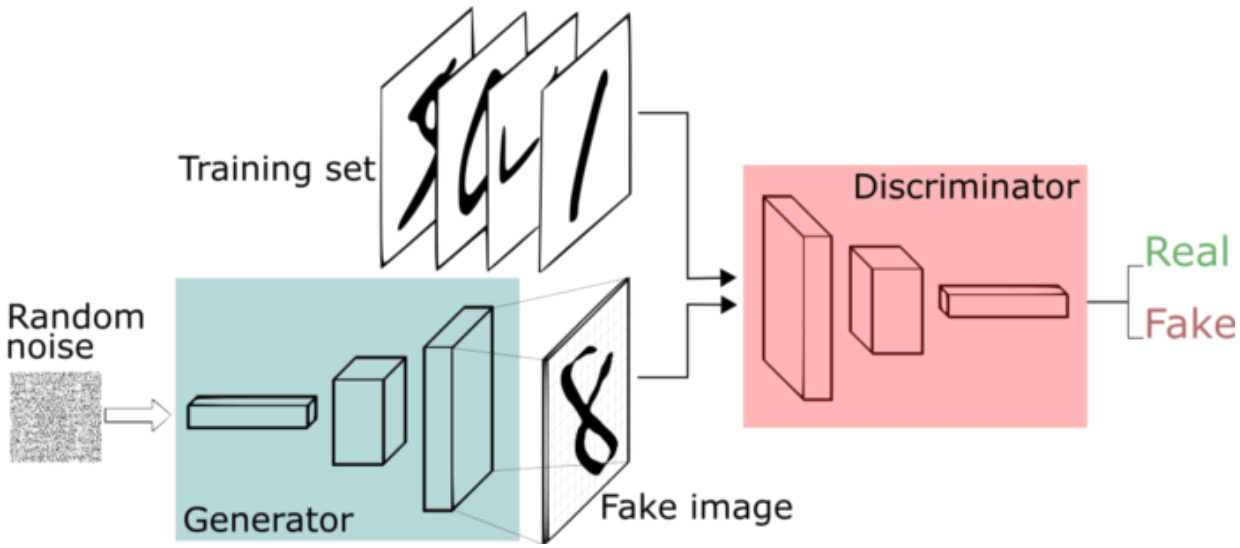
StyleGAN

Expect OUTPUT

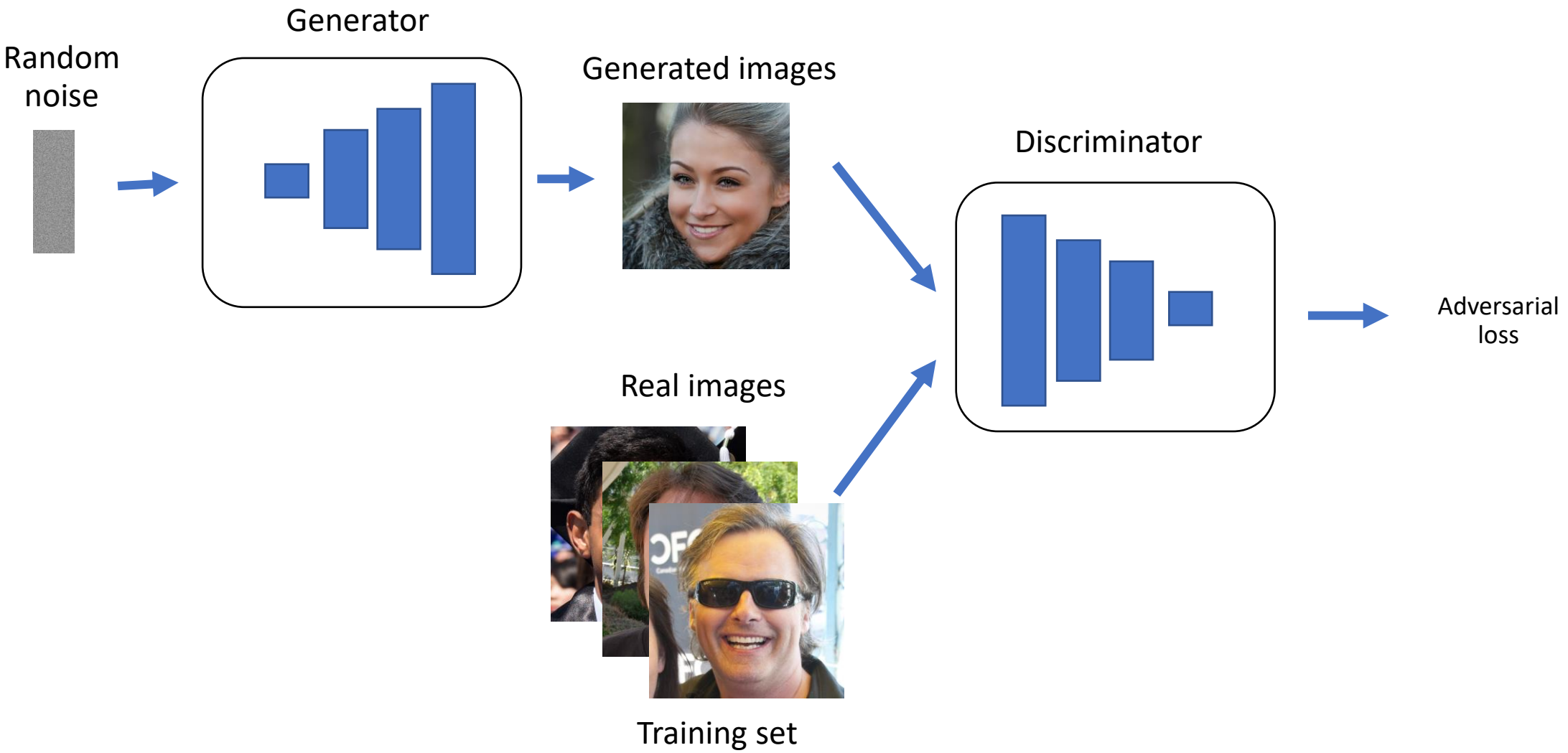
Pencil Cartoonized Portrait

Related Works

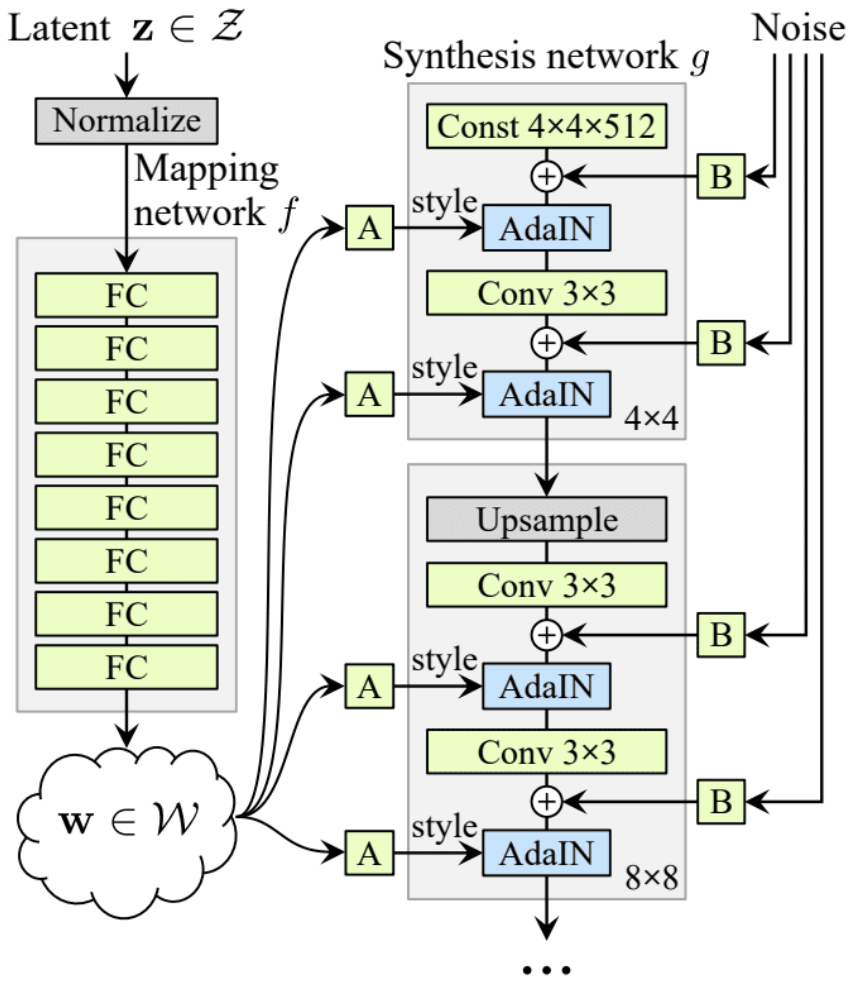
- GANs:Generative Adversarial Networks
- StyleGAN
- Gan Inversion
- JoJoGAN

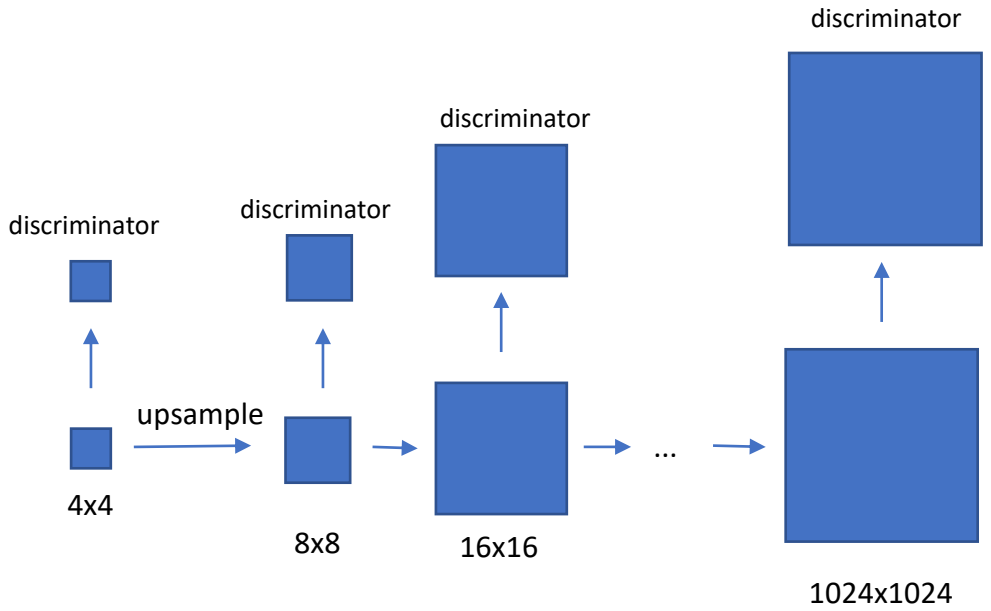


GANs: Generative Adversarial Networks



StyleGAN





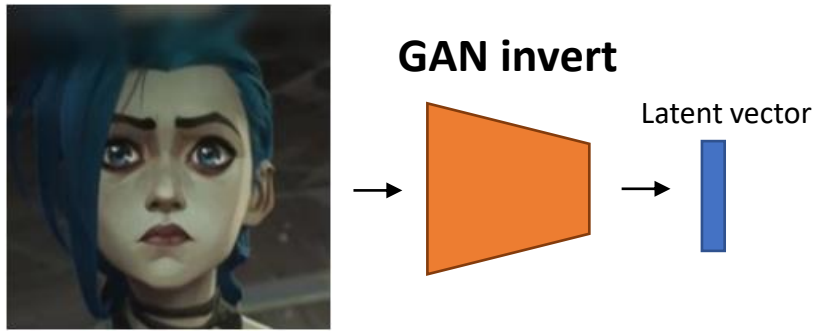
Progressive Growing GAN

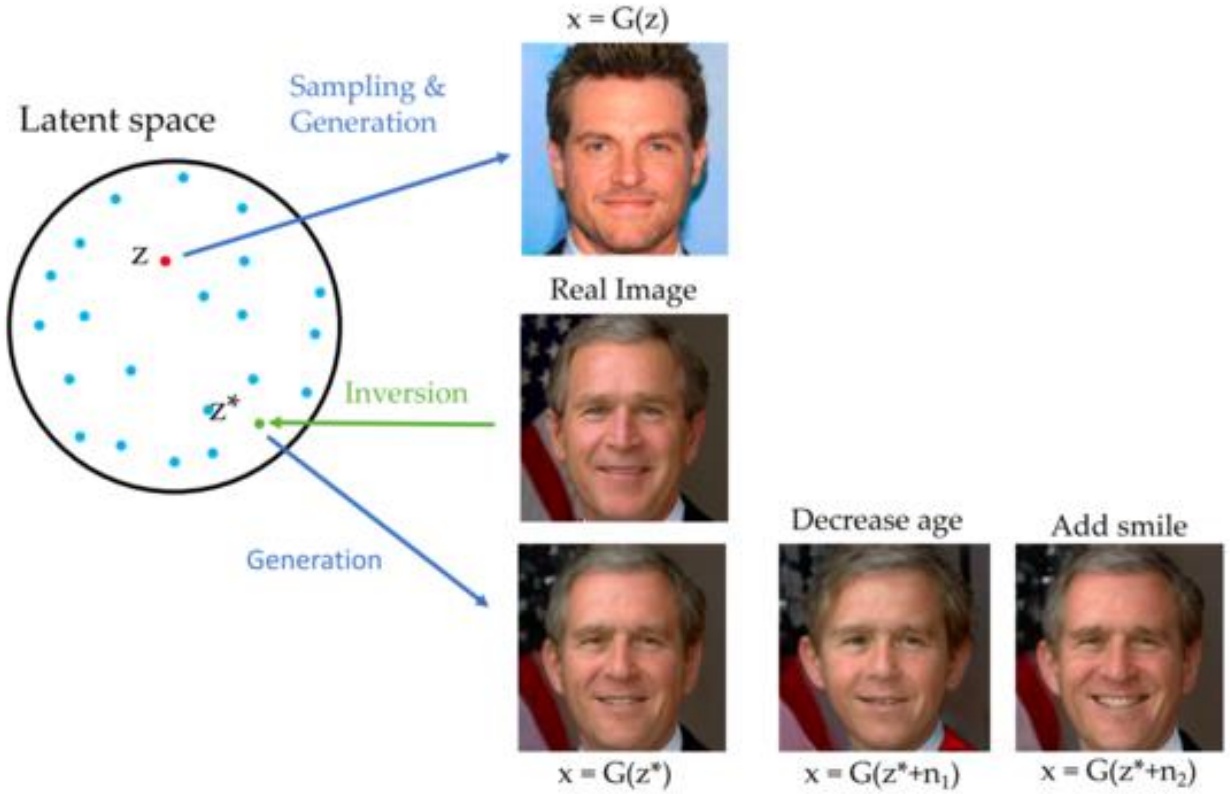
Synthesis Network

- AdaIN(Adaptive Instance Normalization)
- Noise
- Style
  - Latent vector > Mapping Network > Style Vector
  - 2 style vector per 1 leyer > 18 style vector overall
  - Each style vector control style of the input differently

Gan Inversion

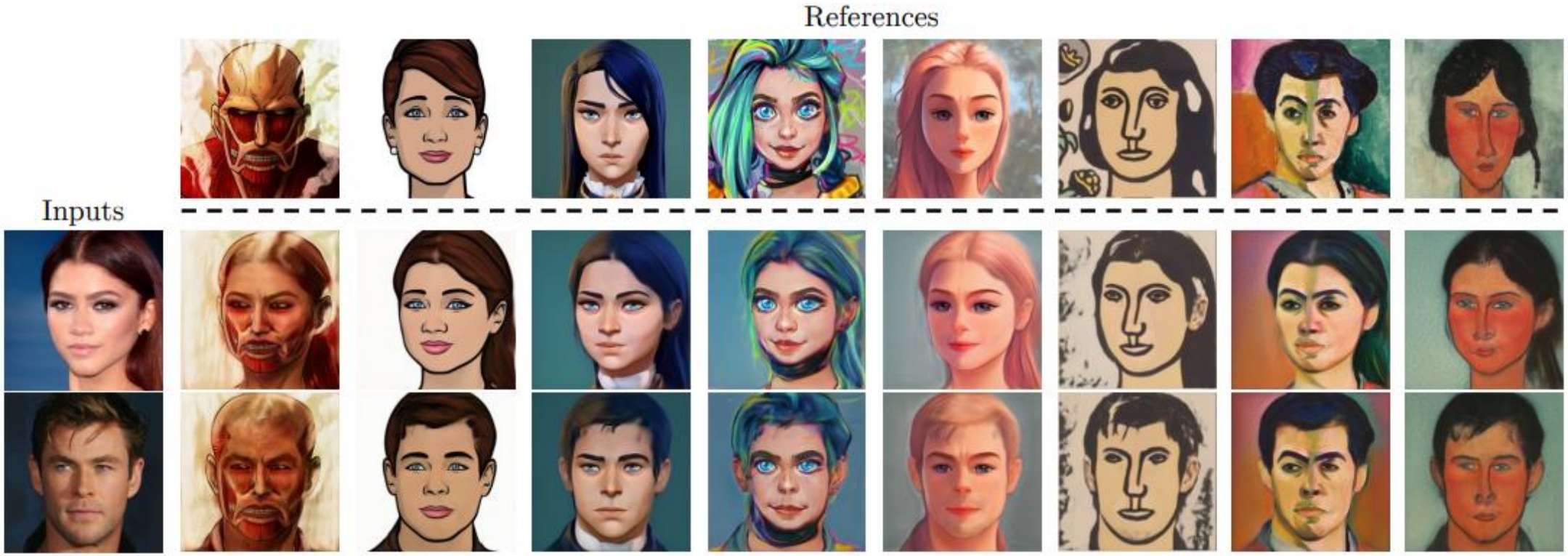
- Get Latent vector form the images





JoJoGAN: One Shot Face Stylization

- Learning a style mapper from a single example of the style
- produce a substantial paired dataset from a single example style



JoJoGAN structure

Reference image, y

Transfer Learning
GAN invert

style vector

w

Transfer Learning
StyleGAN

inverted, x

$S\{ s_i \}$

w s_1 s_2 s_3 ... s_i

Latent vector s_i
close to w

StyleGAN

Images close to x

Latent
space

S

w s_1 s_2 s_3 ... s_i

$S\{w, s_1, s_2, s_3, ..., s_i\}$ -> Training set

Training set
$S\{ s_i \}$

Fine-Tuning
StyleGAN

Generated images

Reference image, y

Transfer Learning
StyleGAN

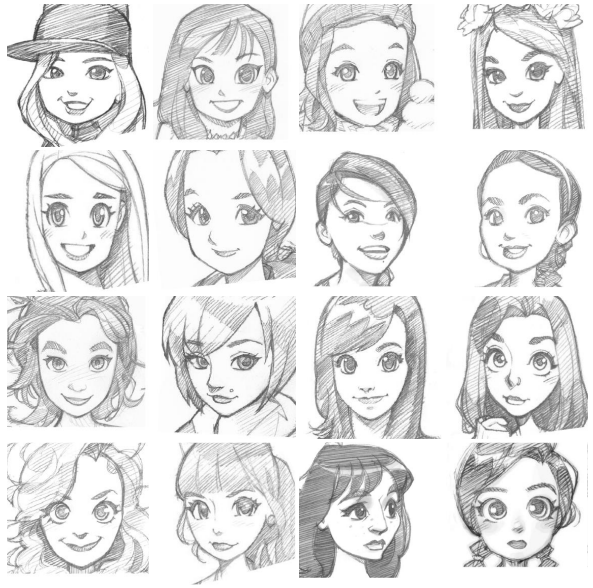
discriminator

perceptual
loss

Experiment

Expect OUTPUT

Reference dataset



INPUT

Condition for input

- Image contain full 1 face

Randomly portrait

Face structure

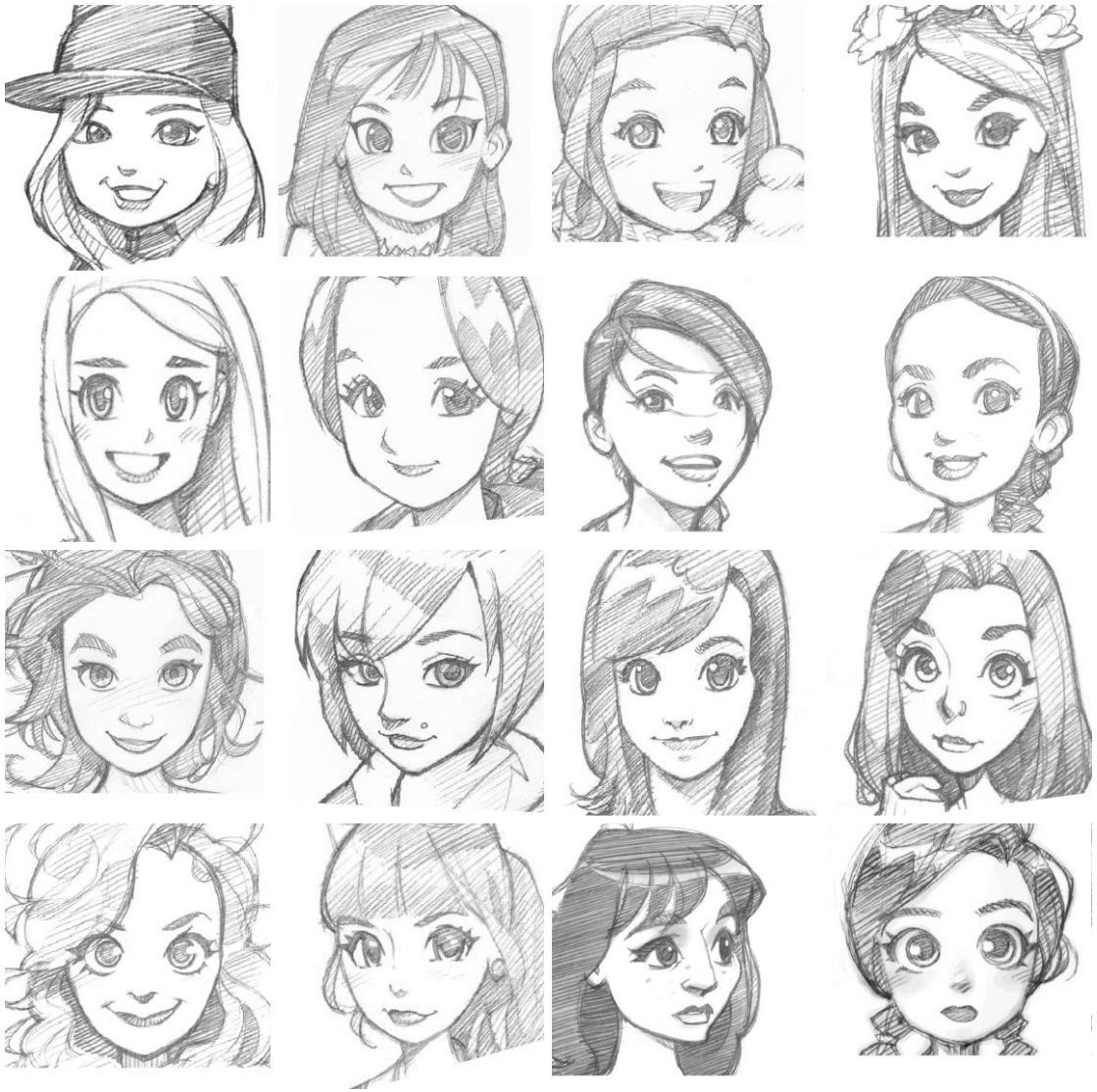
Style

JoJoGAN

Expect
OUTPUT

Pencil Cartoon Portrait

Reference dataset

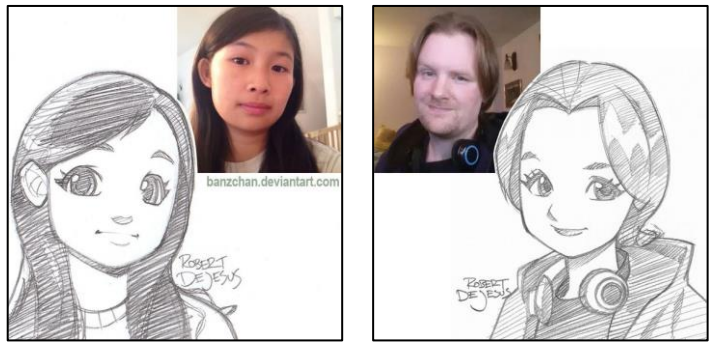


pencil cartoon portrait

18 images

512 x 512, no background

By Robert DeJesus
American artist

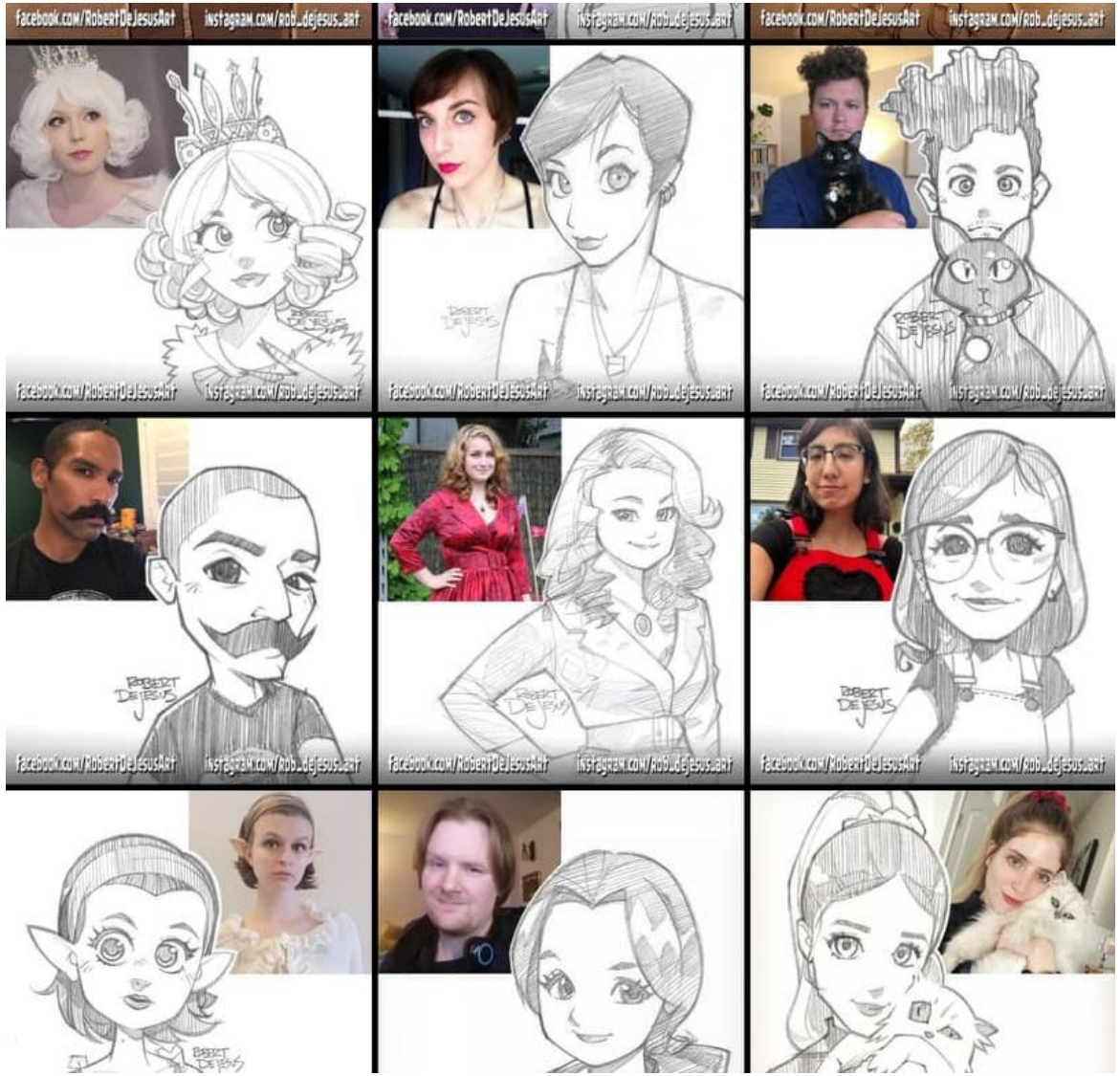


Original image form Robert DeJesus

Robert DeJesus, American artist

instagram.com/rob_dejesus_art/
facebook.com/RobertDeJesusArt/

Random face drawing



StyleGAN Pretrain

Face Alignment
FFHQ dataset at 1024×1024

example

JoJoGAN with My Target

Reference image

Transfer Learning
GAN invert

style vector

Transfer Learning
StyleGAN

inverted, x

$w_1, w_2, w_3, \ldots, w_k$

Latent space

S_k

$w_k\ s_{k,1} s_{k,2} s_{k,3} \ldots s_{k,i}$

Training set
$S_k\{ s_{k,i} \}$

StyleGAN
Fine-Tuning

Generated images

Reference image

Transfer Learning
StyleGAN

discriminator

perceptual loss

$S_k\{w_k, s_{k,1}, s_{k,2}, s_{k,3}, \ldots, s_{k,i}\}$ -> Training set

Step1: Gan Inversion

Reference image

Face Alignment

DLIB → **crop**

Detecting Facial Features

5-point landmarking

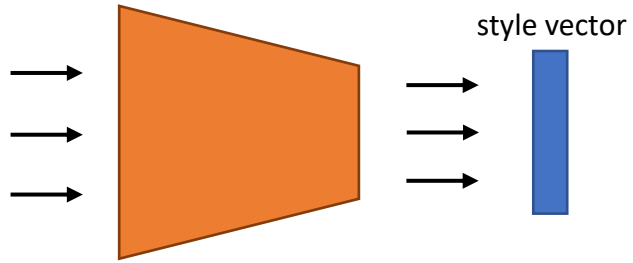
face alignment reference

1024x1024

Transfer Learning

e4e network

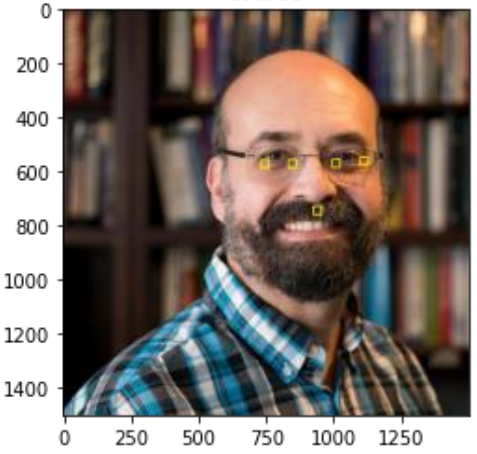
GAN inversion



example

DLIB

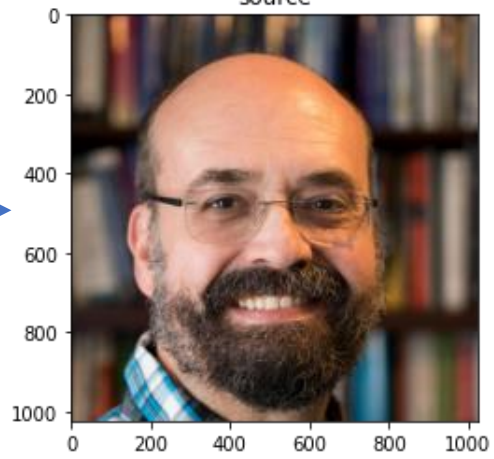
source



example

crop

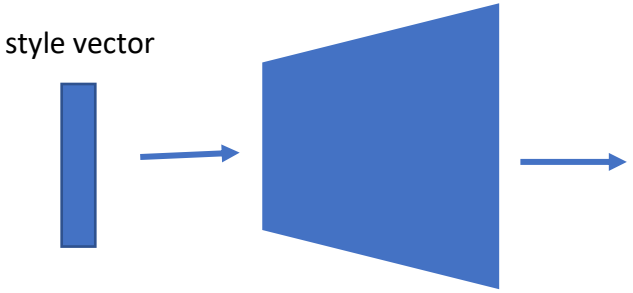
source



Transfer Learning

StyleGAN

style vector



Generated images

Step 2: Create Style Vectors set

Find a set of style vectors that are close to s(w)

style vectors from step1

s(w)

StyleGAN's style mixing mechanic

$$s_i = M \cdot s + (1 - M) \cdot s(FC(z_i))$$

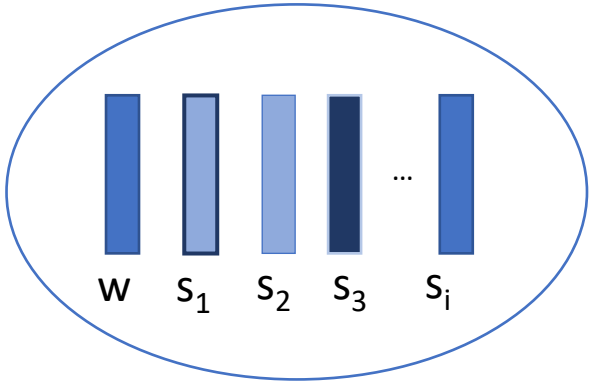
FC: style mapping layer of the StyleGAN

M: {0,1}

-> M is large -> close to style

-> M is low -> close to real face

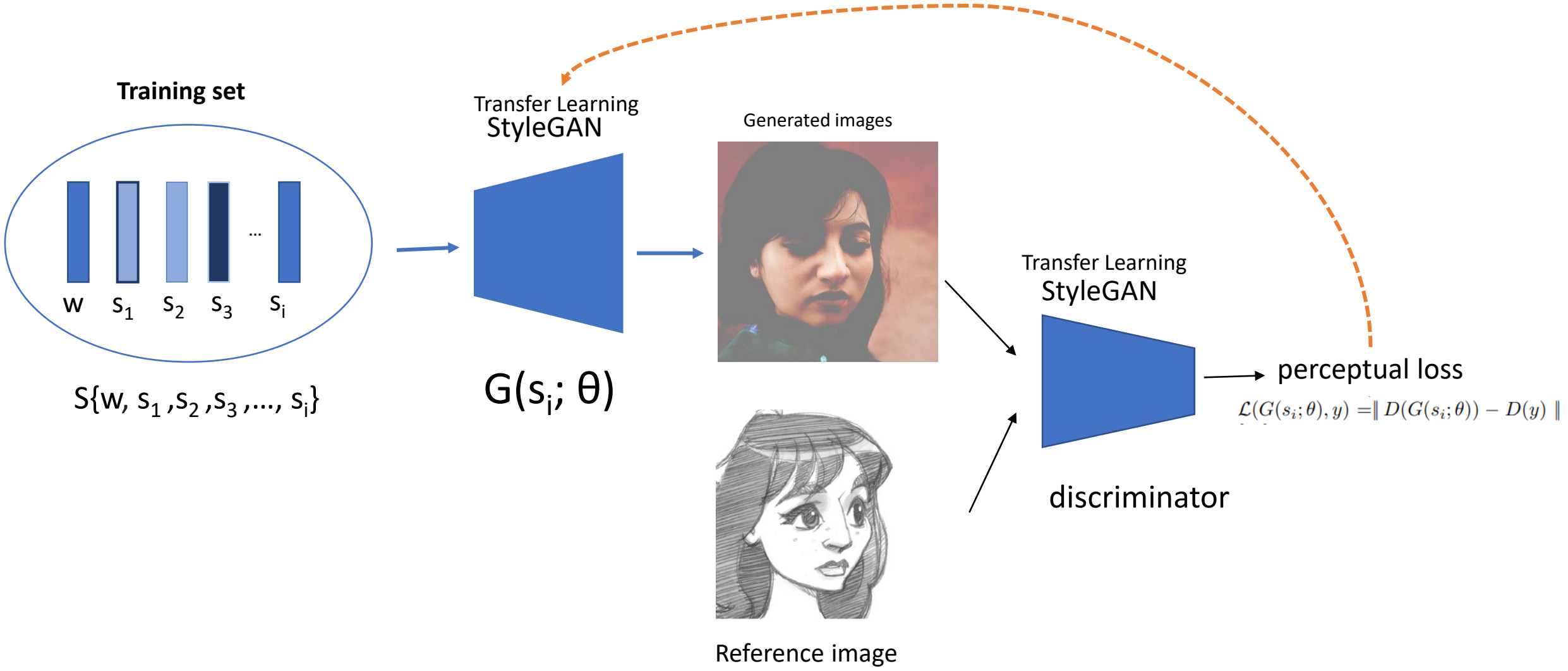
Training set



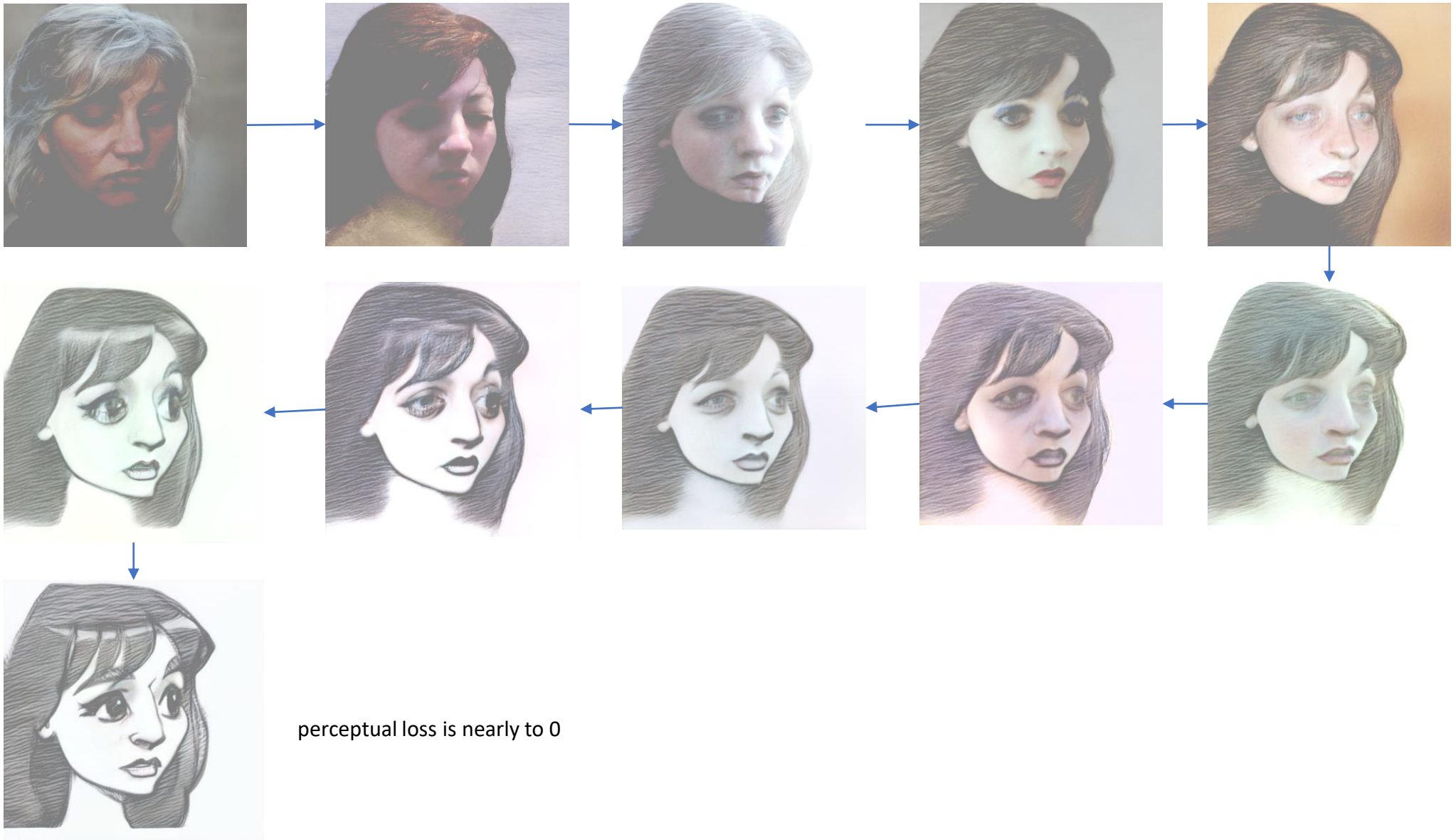
S{w, s_1 ,s_2 ,s_3 ,..., s_i}

This is output of s_i if put it on Generator

Step 3: Fine-Tuning StyleGAN

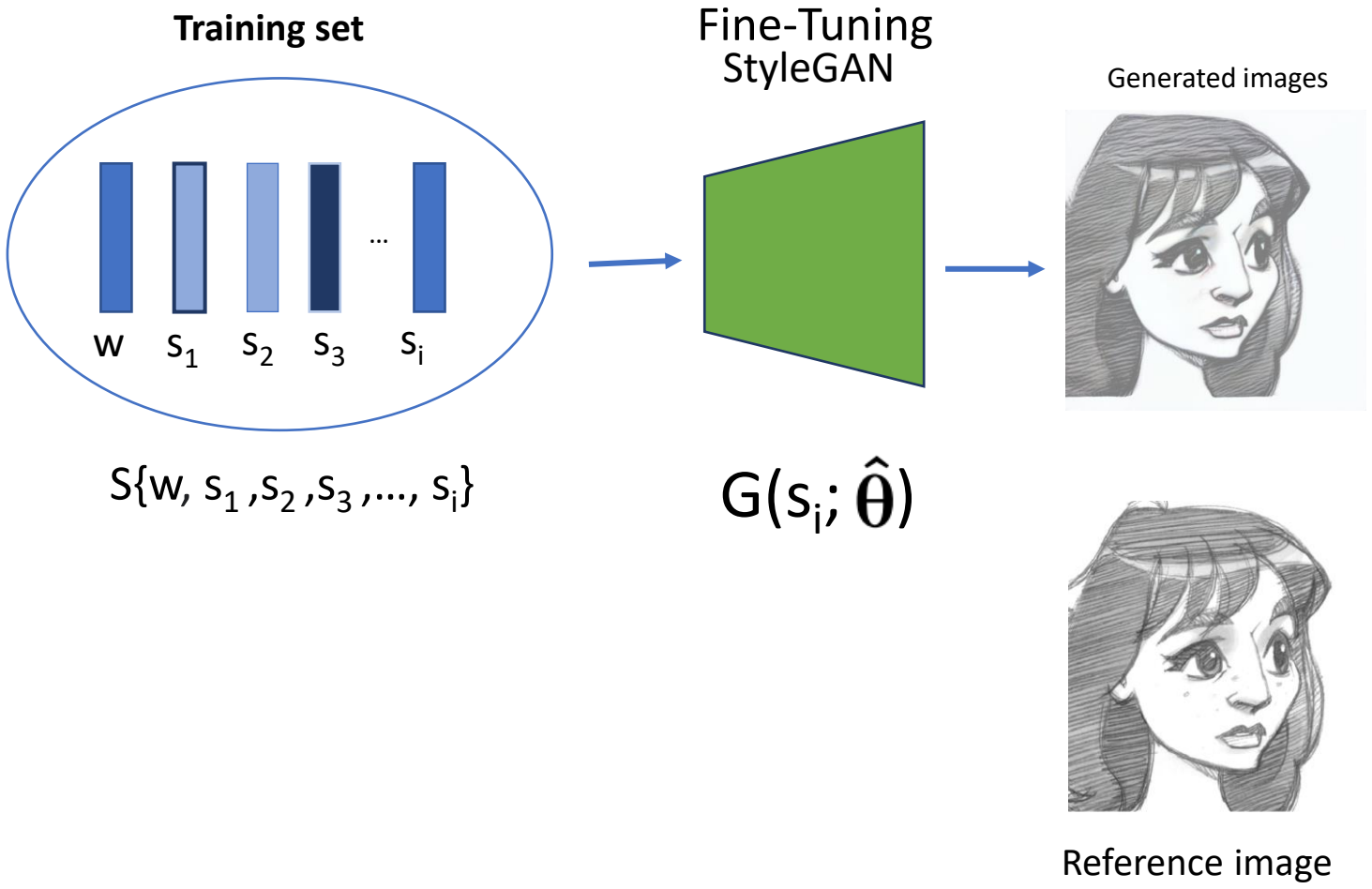


While tuning



perceptual loss is nearly to 0

Step 3: Fine-Tuning Stylegan



Result

Input
Randomly portrait

Face Alignment
DLIB → CROP

Face Alignment

1024x1024

Transfer Learning
GAN invert

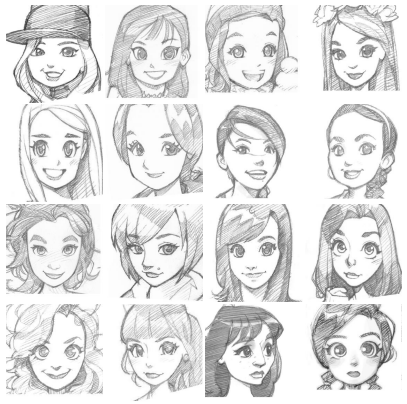
Style vector

Fine-Tuning
StyleGAN

output

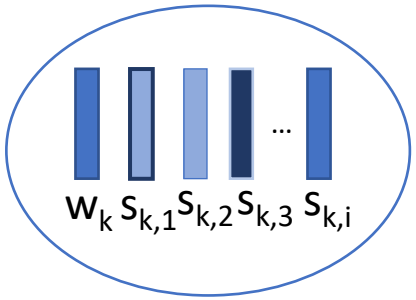
Pencil Cartoon Portrait
1024x1024

Multi Reference Images



N=18

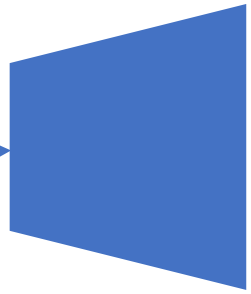
S_k



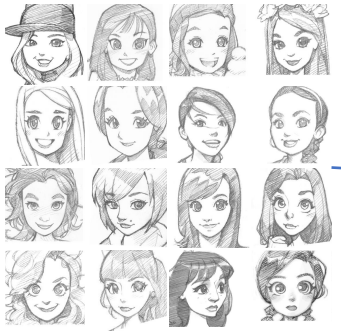
$S_k\{w_k, s_{k,1}, s_{k,2}, s_{k,3}, \ldots, s_{k,i}\}$

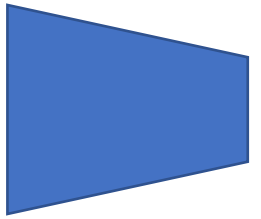
Training set

StyleGAN



Pair



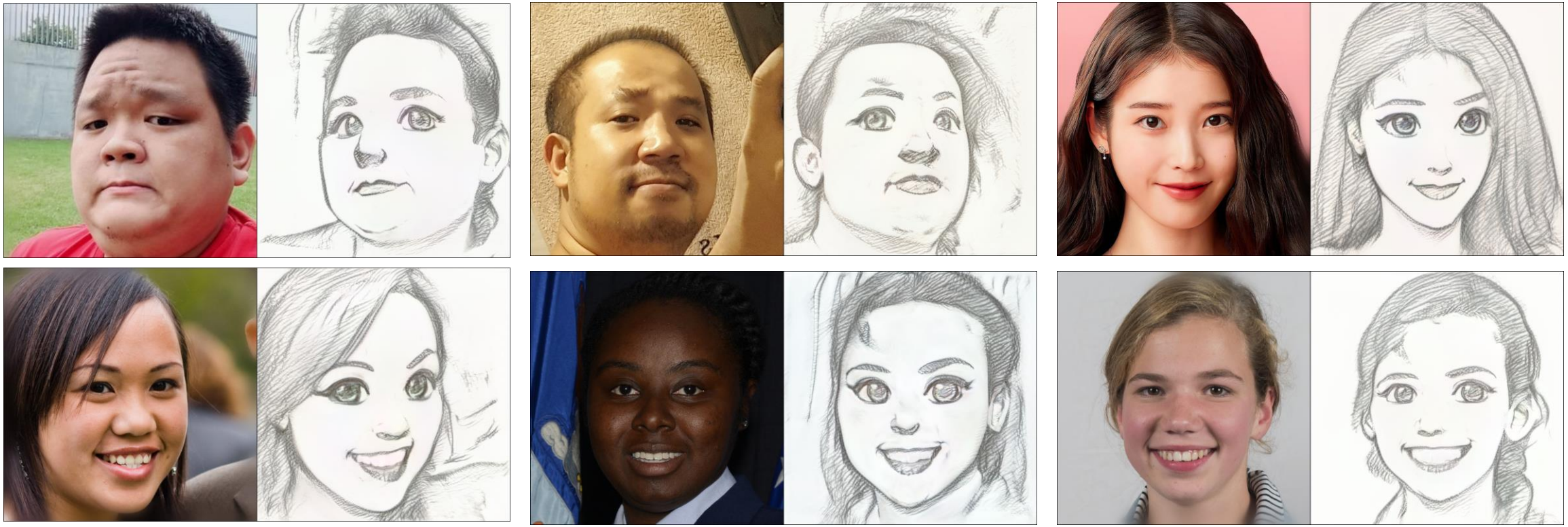


discriminator

perceptual loss

Output with our Multi Reference Images

input output



output with different style

Conclusion

- The portrait used as reference in the JoJoGAN pre-trained model must possess complete elements of the face and the position of the facial features must be related to the face alignment.
- The JoJoGAN only utilizes a single image as reference, thus, the output variations are limited to the style of the reference image.
- The JoJoGAN can generate a restyle of the input while following the reference style photo, thus, it is more efficient in terms of computation time when compared to other GANs.
- The output of the JoJoGAN can be utilized as a reference image for artistic works.

reference ที่ไม่ได้ทำ Face Alignment

Target reference images มากเกินไป

Reference

- JoJoGAN (2021), https://arxiv.org/abs/2112.11641
- CartoonGAN (2018), https://openaccess.thecvf.com/content_cvpr_2018/papers/Chen_CartoonGAN_Generative_Adversarial_CVPR_2018_paper.pdf
- ProGAN (2017), https://arxiv.org/abs/1710.10196
- StyleGAN (2018), https://arxiv.org/abs/1812.04948
- GANs (2014), https://arxiv.org/abs/1406.2661
- e4e network, https://arxiv.org/abs/2102.02766
- Generative Adversarial Networks — Explained, https://towardsdatascience.com/generative-adversarial-networks-explained-34472718707a
- StyleGAN Explained, https://machinelearningmastery.com/introduction-to-style-generative-adversarial-network-stylegan/
- StyleGanComponents, https://isaac-flath.github.io/fastblog/computer%20vision/gan/2021/03/01/StyleGanComponents.html