

## CS534 Computer Vision

Facial Sketches to Colored Images using  
GANs

Parth Goel

Prashant Kanth

Rishika Bhanushali



# References

## Algorithms

- Pix2pix. [ Phillip Isola et al. ]
- Sketch2face. [ Julia Gong, Matthew Mistele ]
- DeepFaceDrawing [ Shu-Yu Chen, Wanchao Su et al. ]
- Cycle-GAN [ Jun-Yan Zhu et al. ]

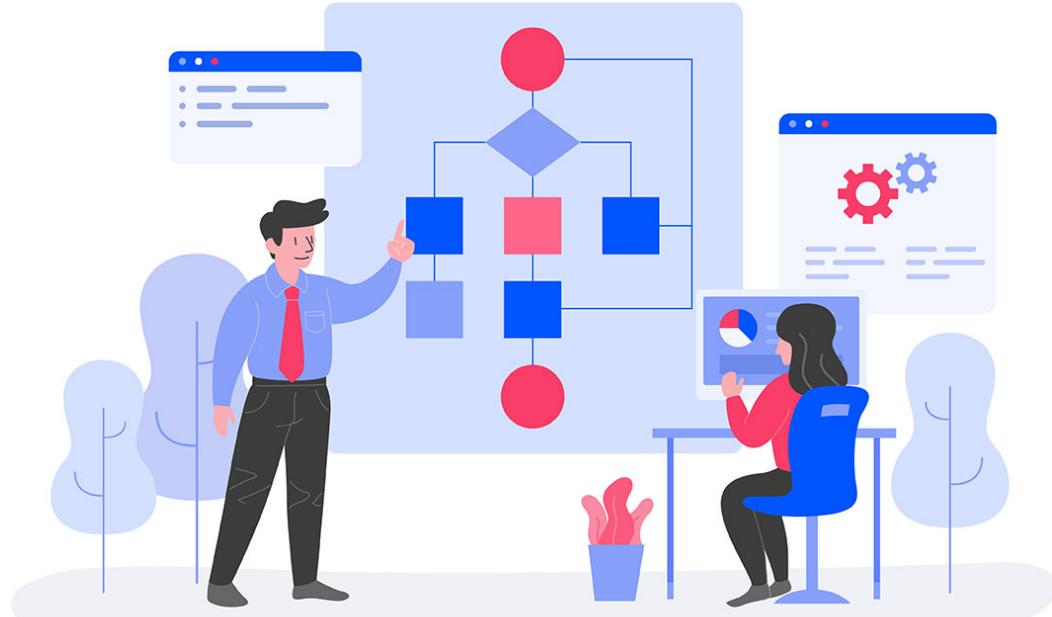
## Dataset

- CUHK Student Dataset
- FS2K Dataset



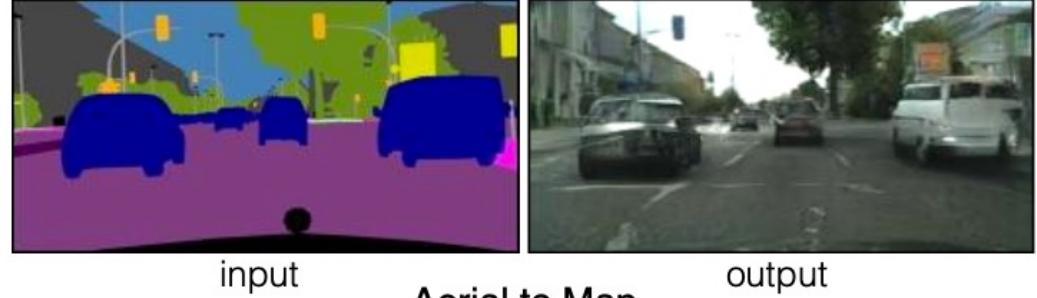
# Introduction

Computer Algorithm



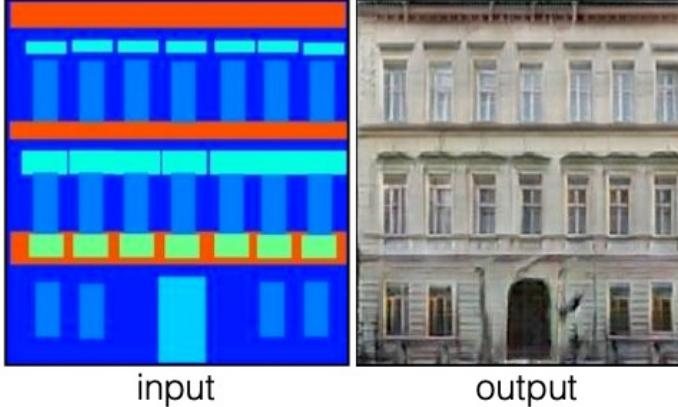
Algorithms

Labels to Street Scene



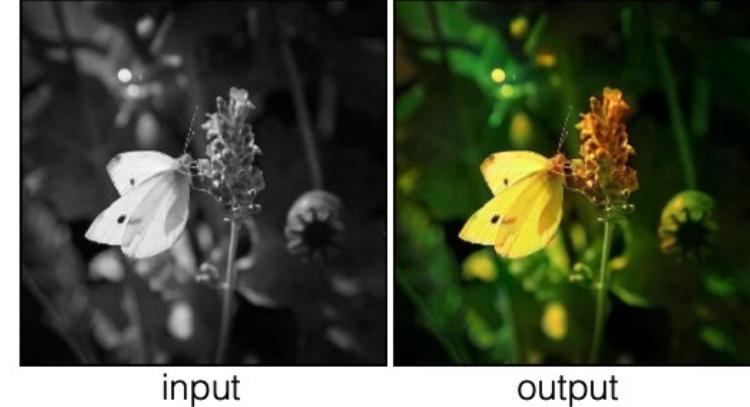
input

Labels to Facade



input

BW to Color



input

output

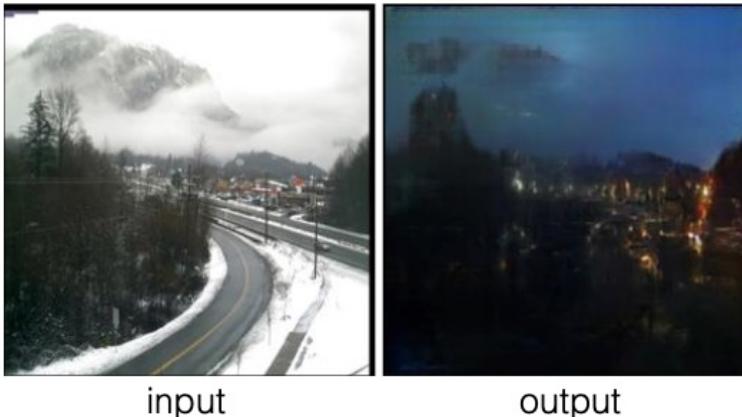
Aerial to Map



input

output

Day to Night



input

output

Edges to Photo



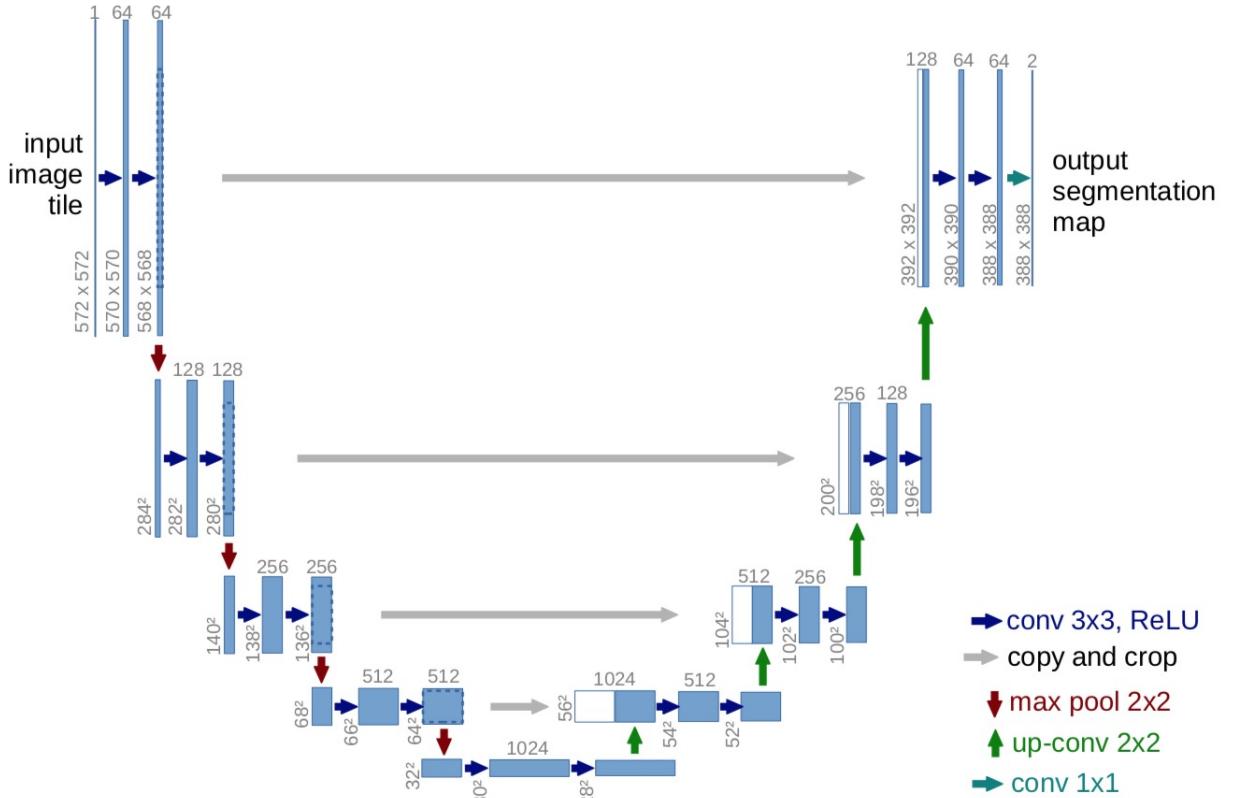
input

output

Pix2pix

■ **Image-to-image Translation with  
Conditional Adversarial Networks**

# Architecture – U-Net

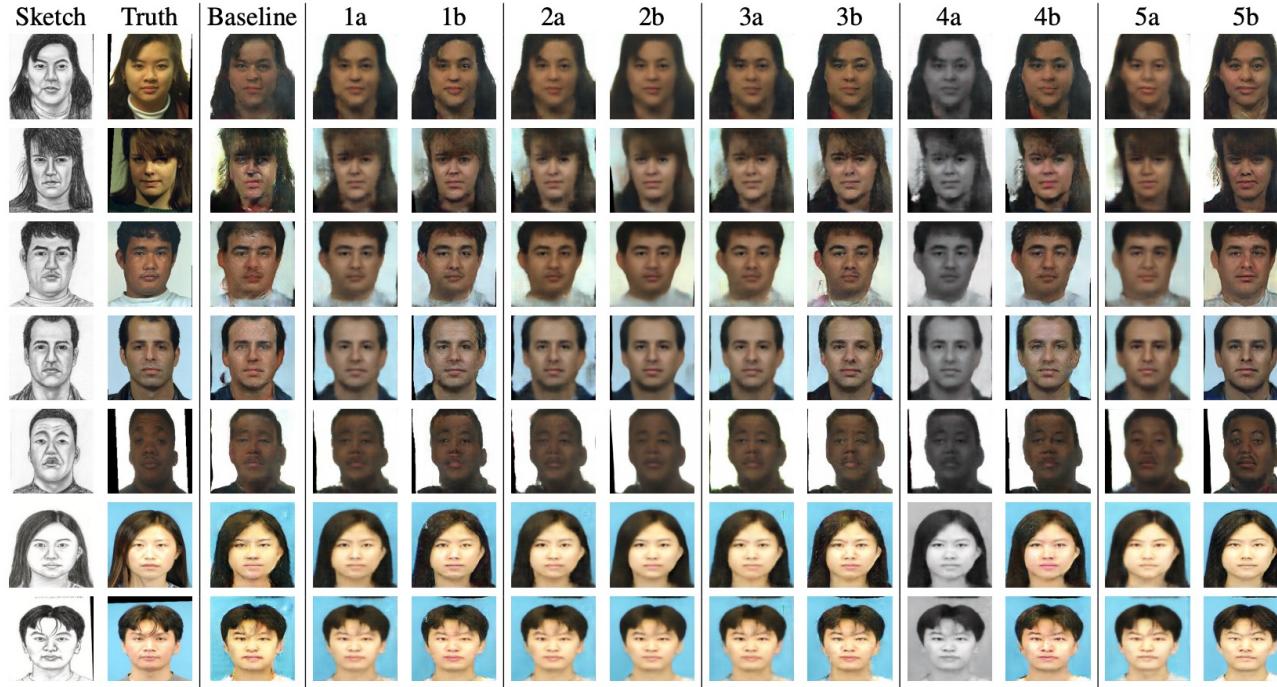


# Methodology

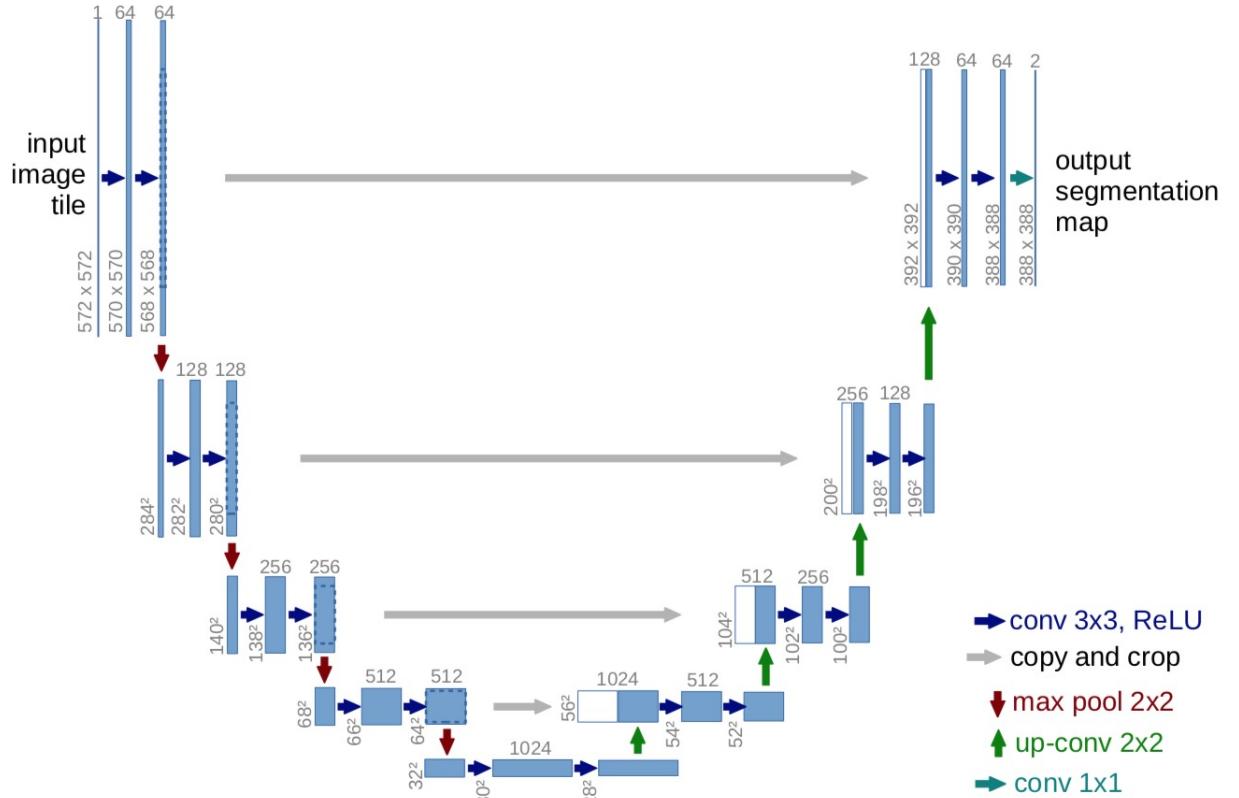
$$\begin{aligned}\mathcal{L}_{cGAN}(G, D) = & \mathbb{E}_{x,y}[\log D(x, y)] + \\ & \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))]\end{aligned}$$

$$\mathcal{L}_{L1}(G) = \mathbb{E}_{x,y,z}[\|y - G(x, z)\|_1].$$

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L1}(G).$$



# Architecture – U-Net

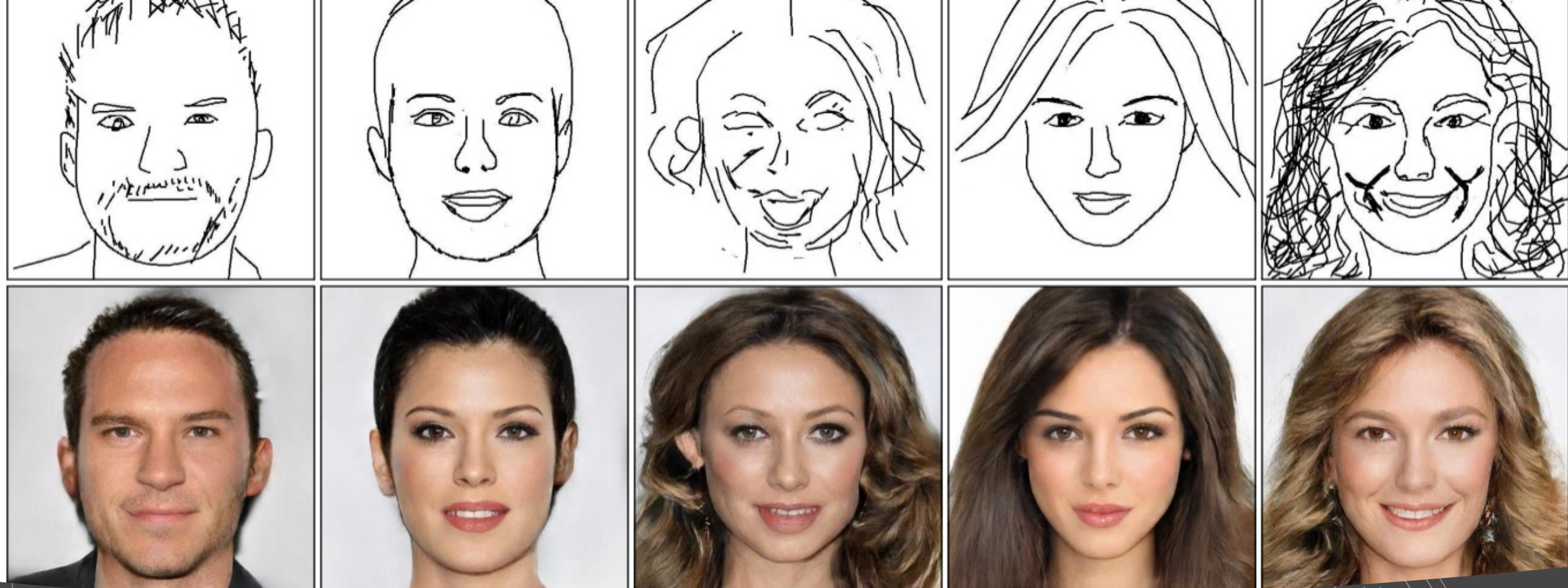


# Methodology

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,z}[\log(1 - D(x, G(x, z)))].$$

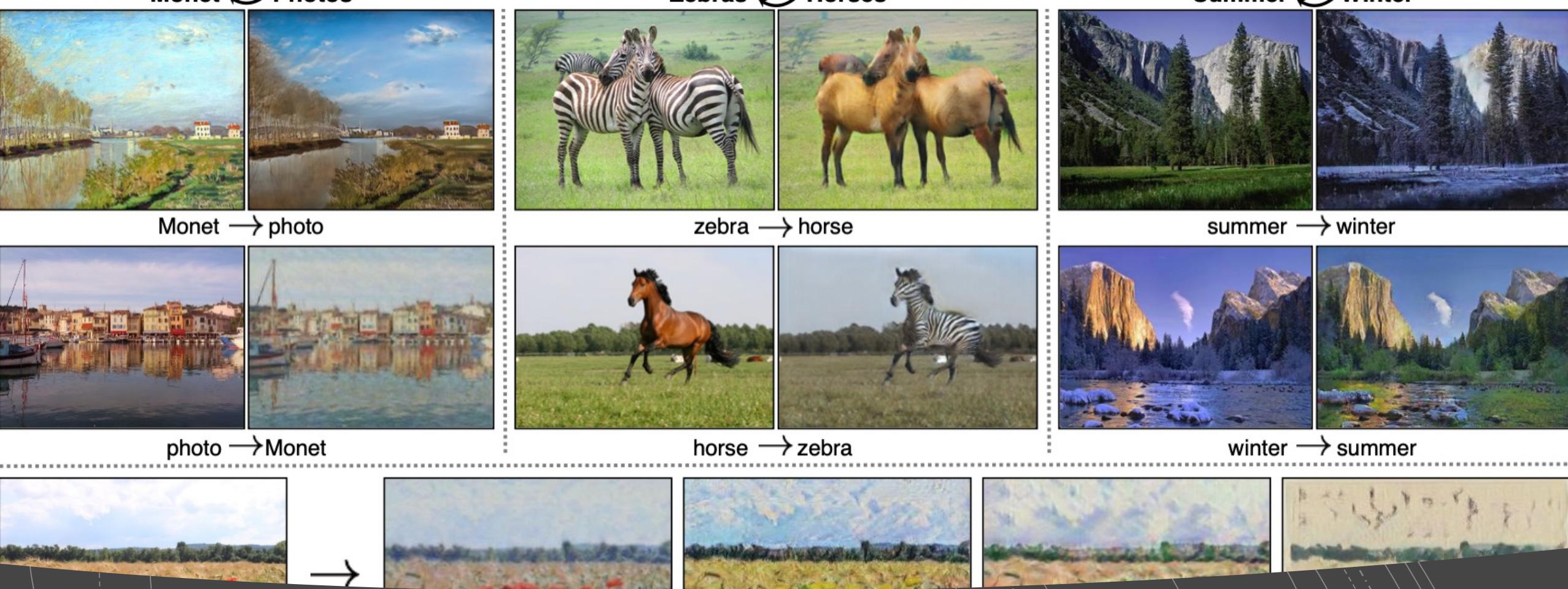
$$\mathcal{L}_{L_1}(G) = \mathbb{E}_{x,y,z}[||y - G(x, z)||_1].$$

$$\mathcal{L}_{cGANFinal} = \mathcal{L}_{cGAN}(G_1, D) + \lambda \mathcal{L}_{L_1}(G_1) + \lambda \mathcal{L}_{L_1}(G_2).$$



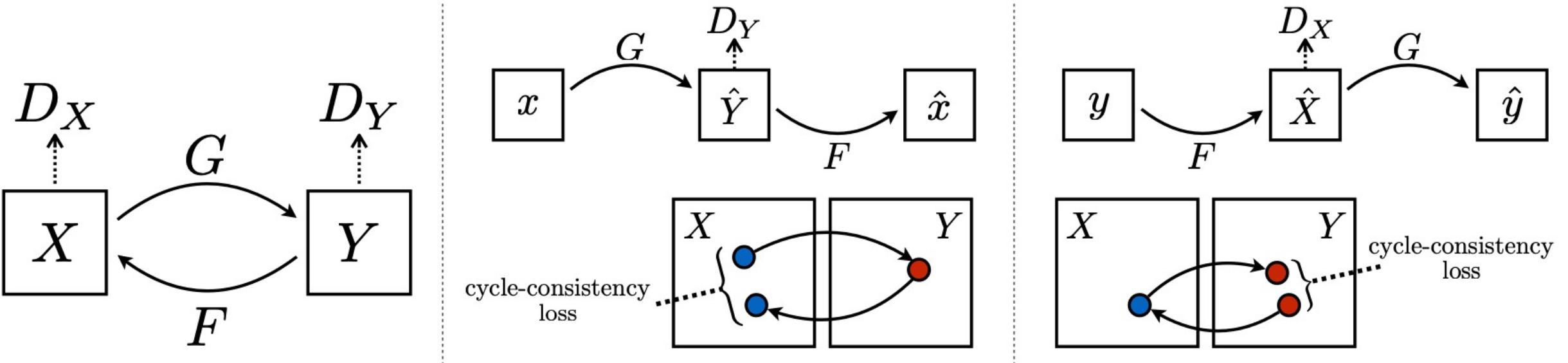
# DeepFaceDrawing

- Deep Generation of Face Images from Sketches



Cycle-GAN

- Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks



# Architecture

# Methodology

$$\begin{aligned}\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = & \mathbb{E}_{y \sim p_{\text{data}}(y)} [\log D_Y(y)] \\ & + \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log(1 - D_Y(G(x)))]\end{aligned}$$

$$\begin{aligned}\mathcal{L}_{\text{cyc}}(G, F) = & \mathbb{E}_{x \sim p_{\text{data}}(x)} [\|F(G(x)) - x\|_1] \\ & + \mathbb{E}_{y \sim p_{\text{data}}(y)} [\|G(F(y)) - y\|_1].\end{aligned}$$

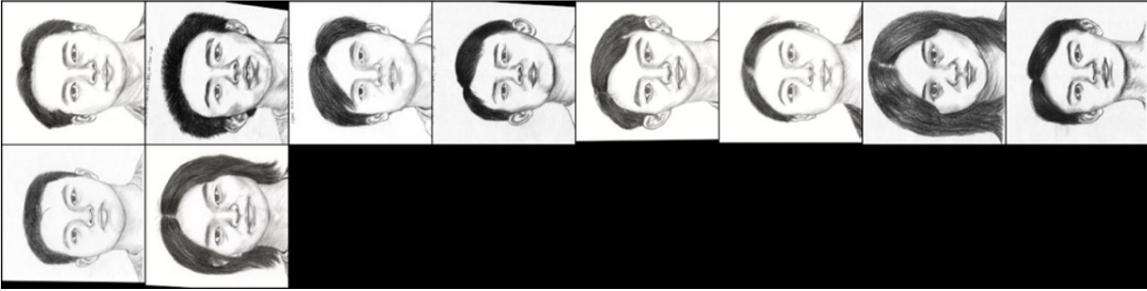
$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) = & \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ & + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ & + \lambda \mathcal{L}_{\text{cyc}}(G, F),\end{aligned}$$

# Improvements

- Dataset
- Weight Initialization
- Architecture
- Background Removal
- Input Size

The background of the image consists of a grid of faces, likely from a video game or movie, arranged in three rows. The faces are mostly dark and shadowed, with some highlights on the forehead and nose. The grid is composed of approximately 15-20 individual portraits.

Results



Input Sketch



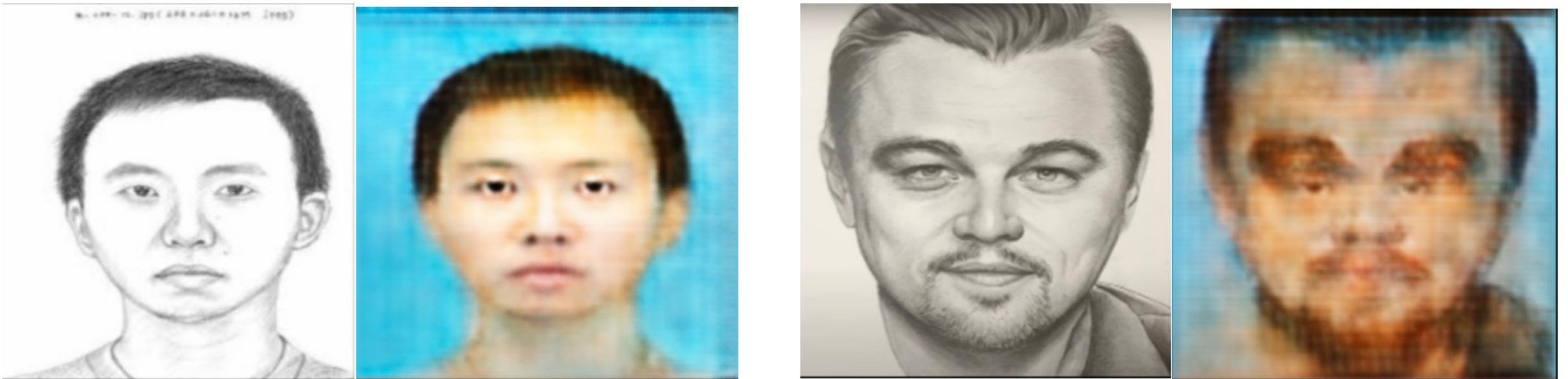
Source Truth



Generated Result

## Results

[256x256 input dim] weight initialized  $\mu = 1$  and  $\sigma = 0.02$



## Results

*[256x256 input dim] Test Result on sketches from same dataset and outside dataset*



Input Sketch



Generated Result



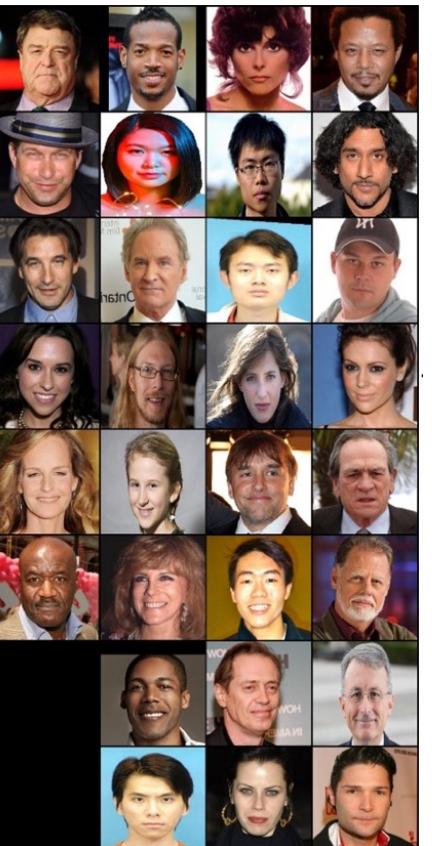
Source Truth

## Results

*[256x256 input dim] merged dataset CUHK and FS2K, removed background*



Input Sketch



Source Truth



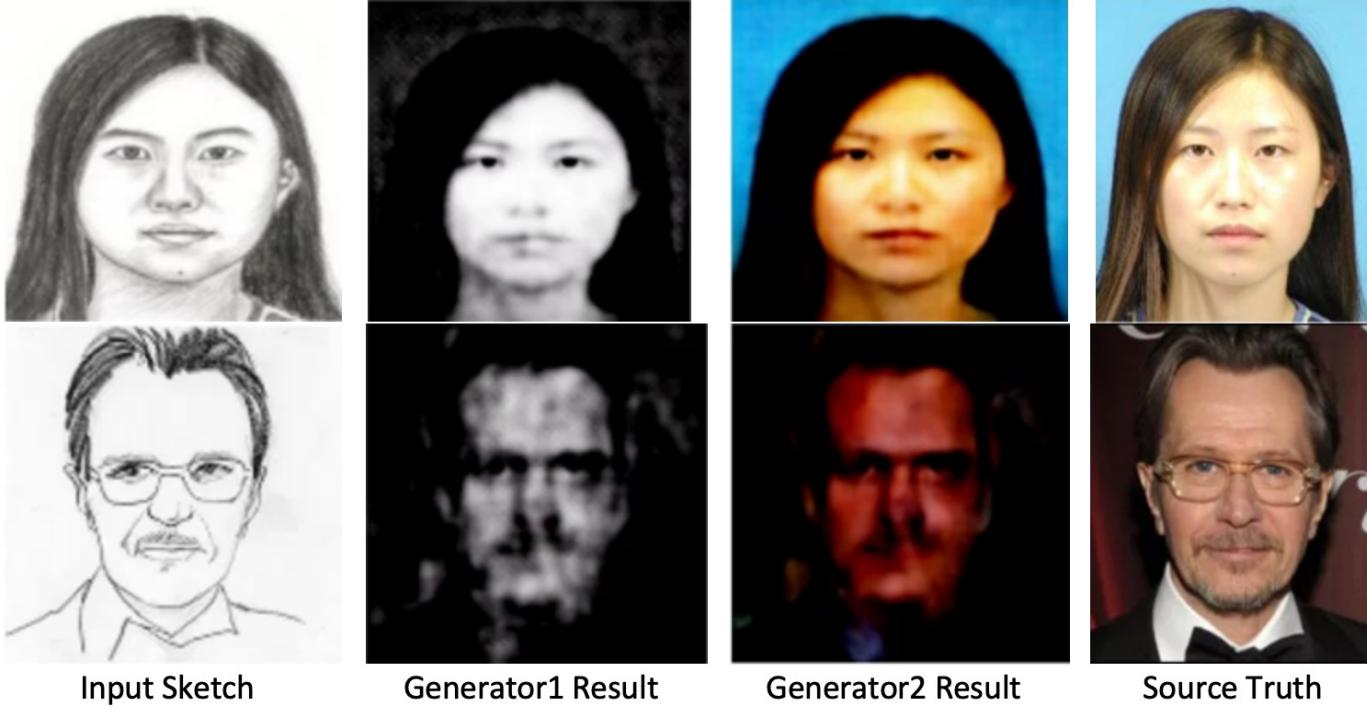
Generator1 Output



Generator2 Output

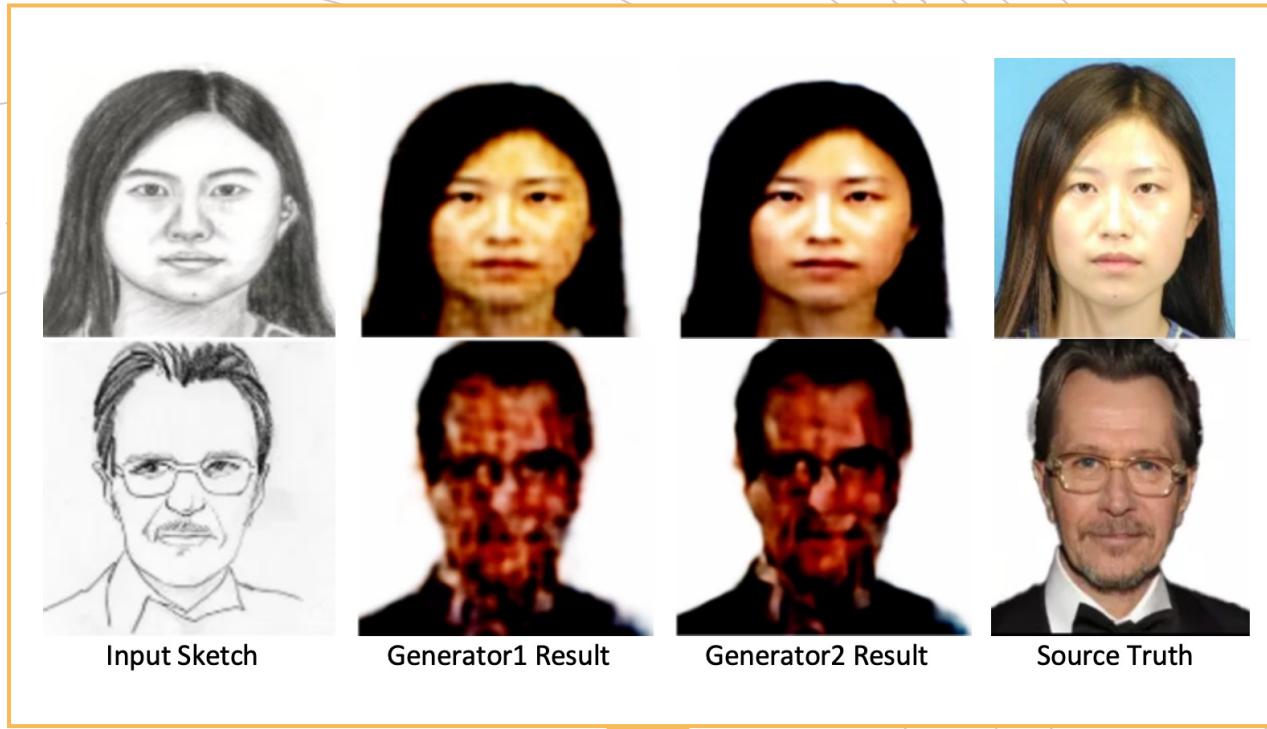
# Results

*[128x128 input dim] merged dataset CUHK and FS2K, with background*



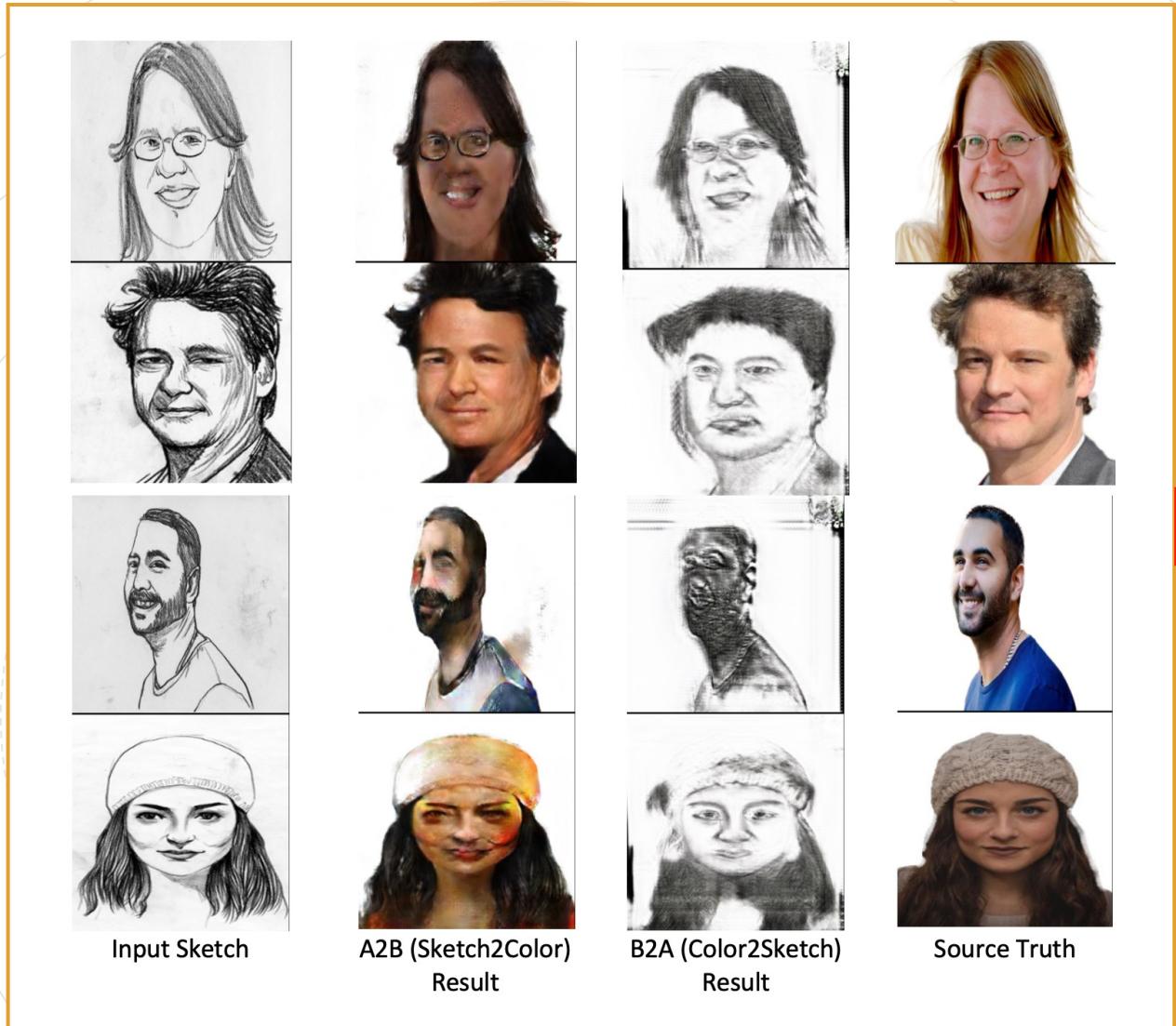
# Results

*Test result [128x128 input dim] merged dataset CUHK and FS2K, with background*



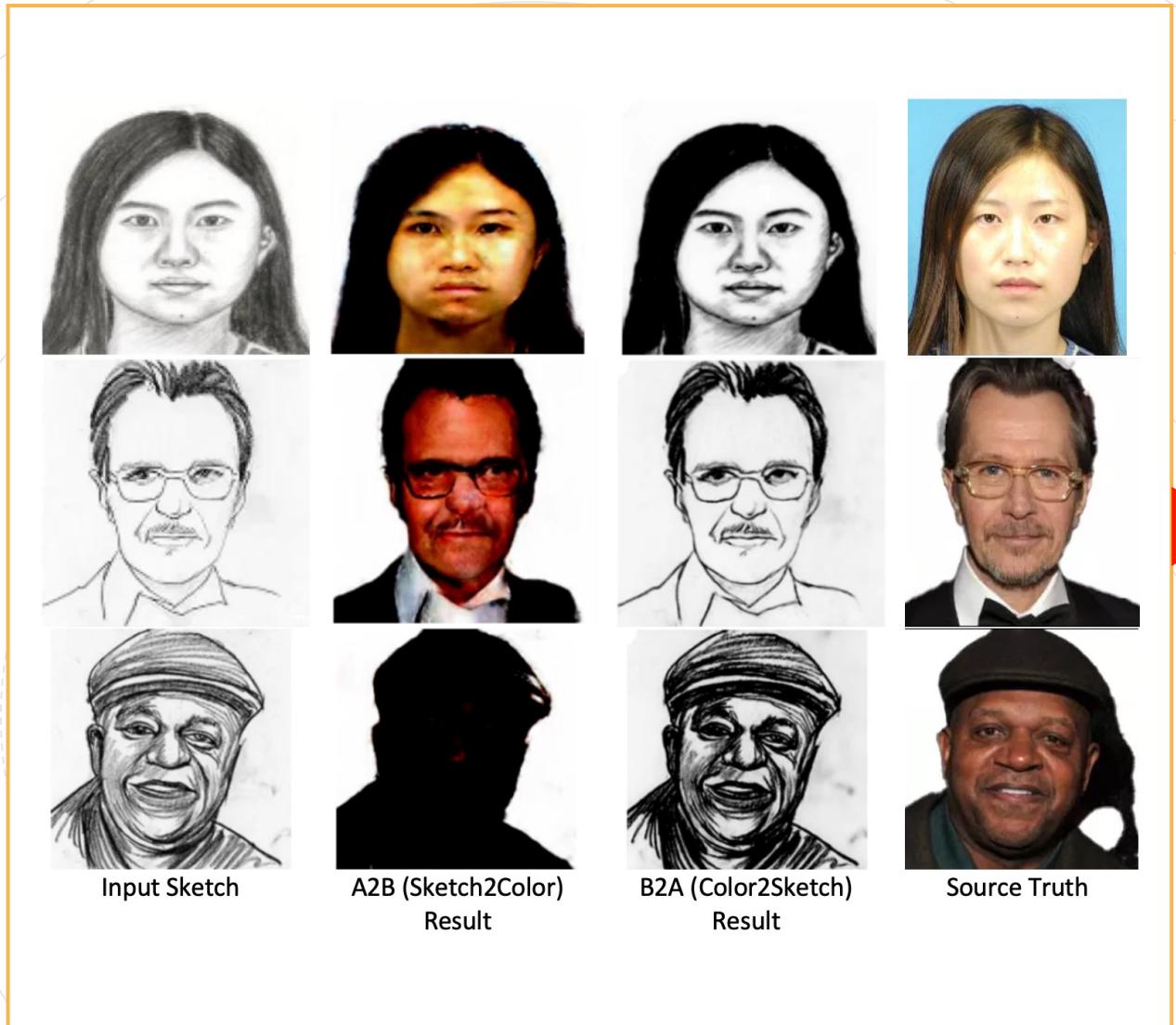
## Results

*[256x256 input dim] merged dataset CUHK and FS2K, without background and colored generator 1*



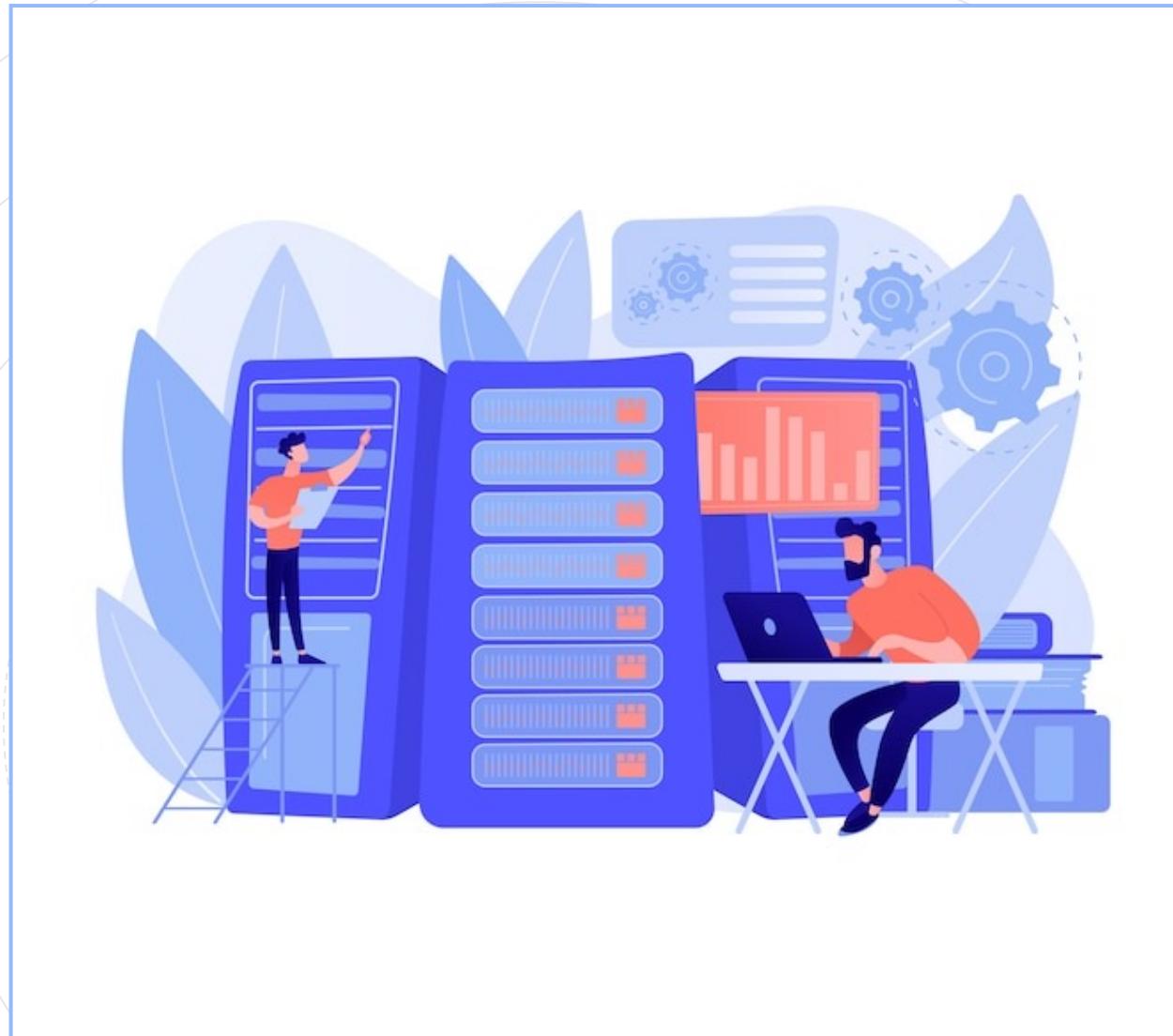
## Results

[256x256 input dim] merged dataset CUHK and FS2K, without background



## Results

[256x256 input dim] merged dataset CUHK and FS2K, without background, also showing bad result due to dark shading and partial removal of background



Dataset



CUHK Student Database

# FS2K Database





Thank You