智能可视建模与仿真实验室

Intelligent Visual Modeling & Simulation (iGame) Lab

Bridging Unpaired Facial Photos and Sketches by Line-drawings

Meimei Shang [1], Fei Gao [1], Xiang Li [1], Jingjie Zhu [2], Lingna Dai [1]

[1] School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China.

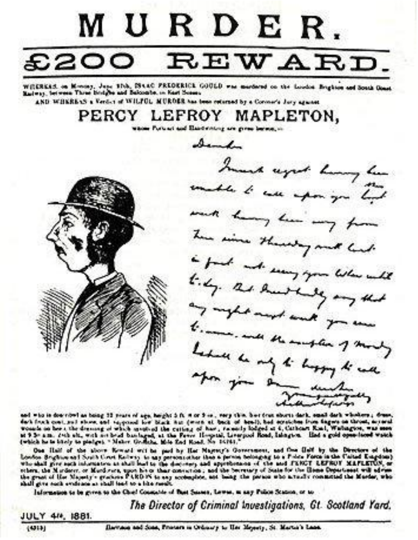
[2] AiSketcher Technology, Hangzhou 311215, China

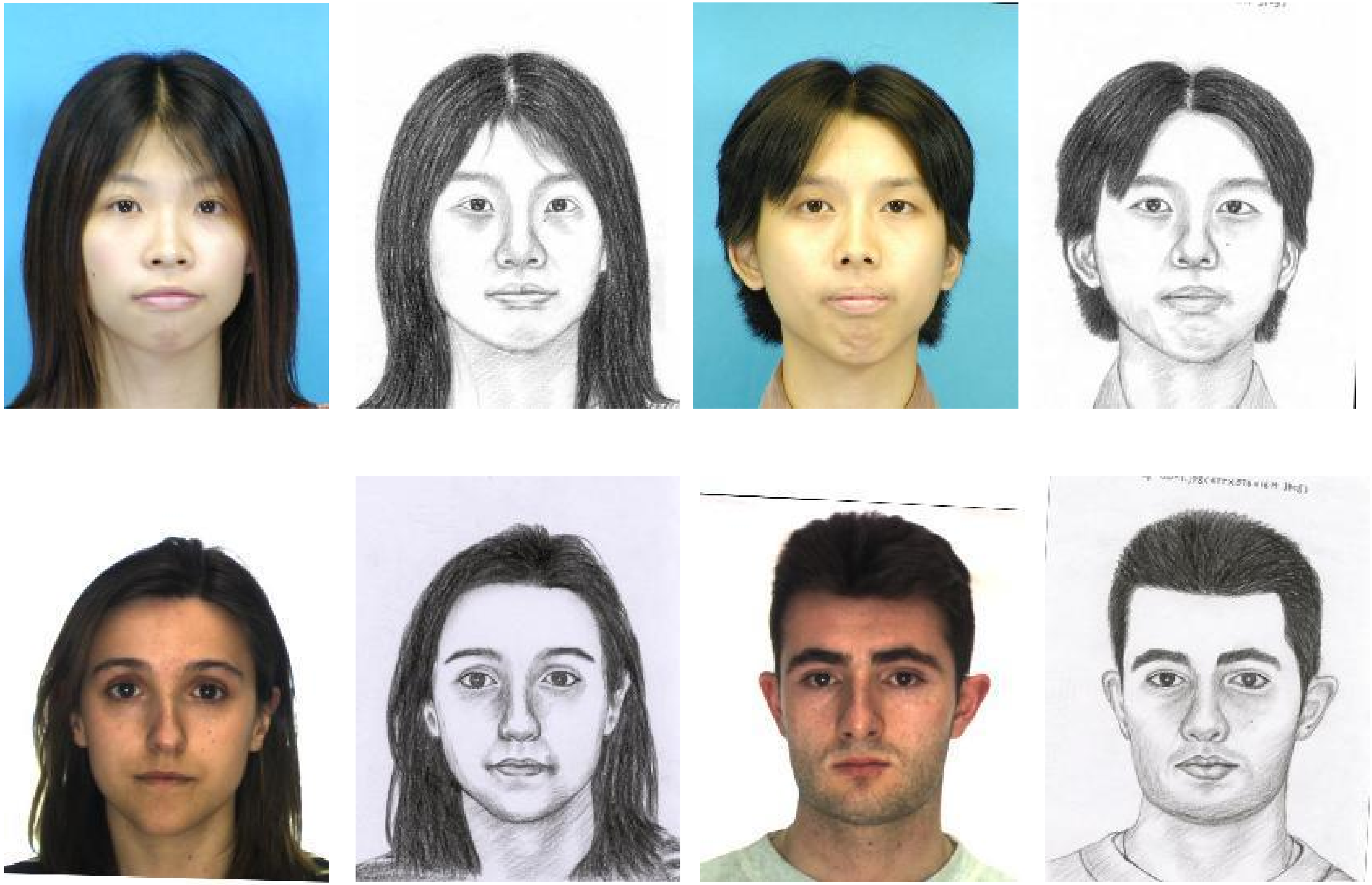
Motivations

Face Sketch Synthesis

- **Facial Photo → Sketch**
- **Applications**
  - Digital entertainment
  - Public Security

Mapleton in 1881
The first *Police composite picture* to appear on a 'wanted' poster and in a newspaper.


MURDER.
£200 REWARD.
PERCY LEFROY MAPLETON,
The Director of Criminal Investigations, Gt. Scotland Yard.
JULY 4th, 1881.



Face Sketch Synthesis

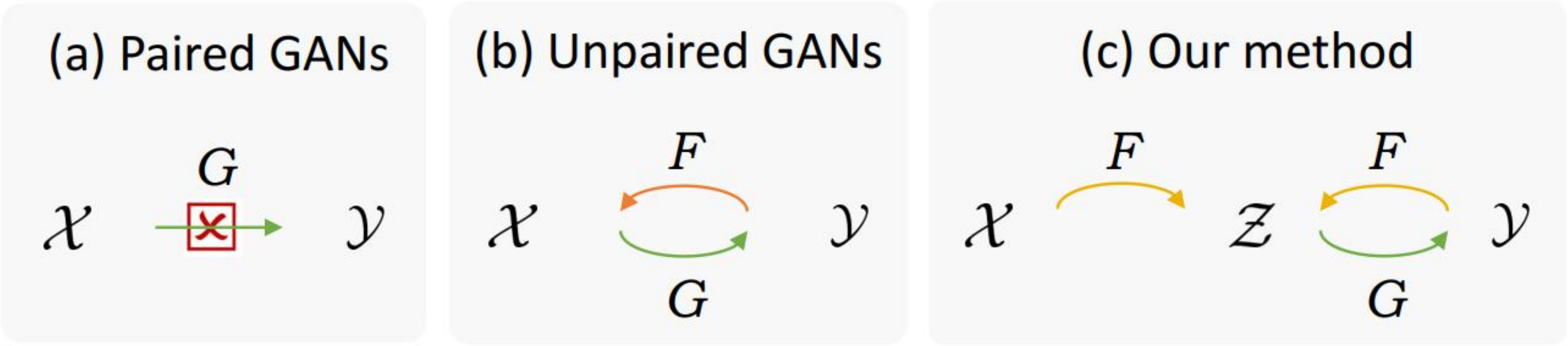


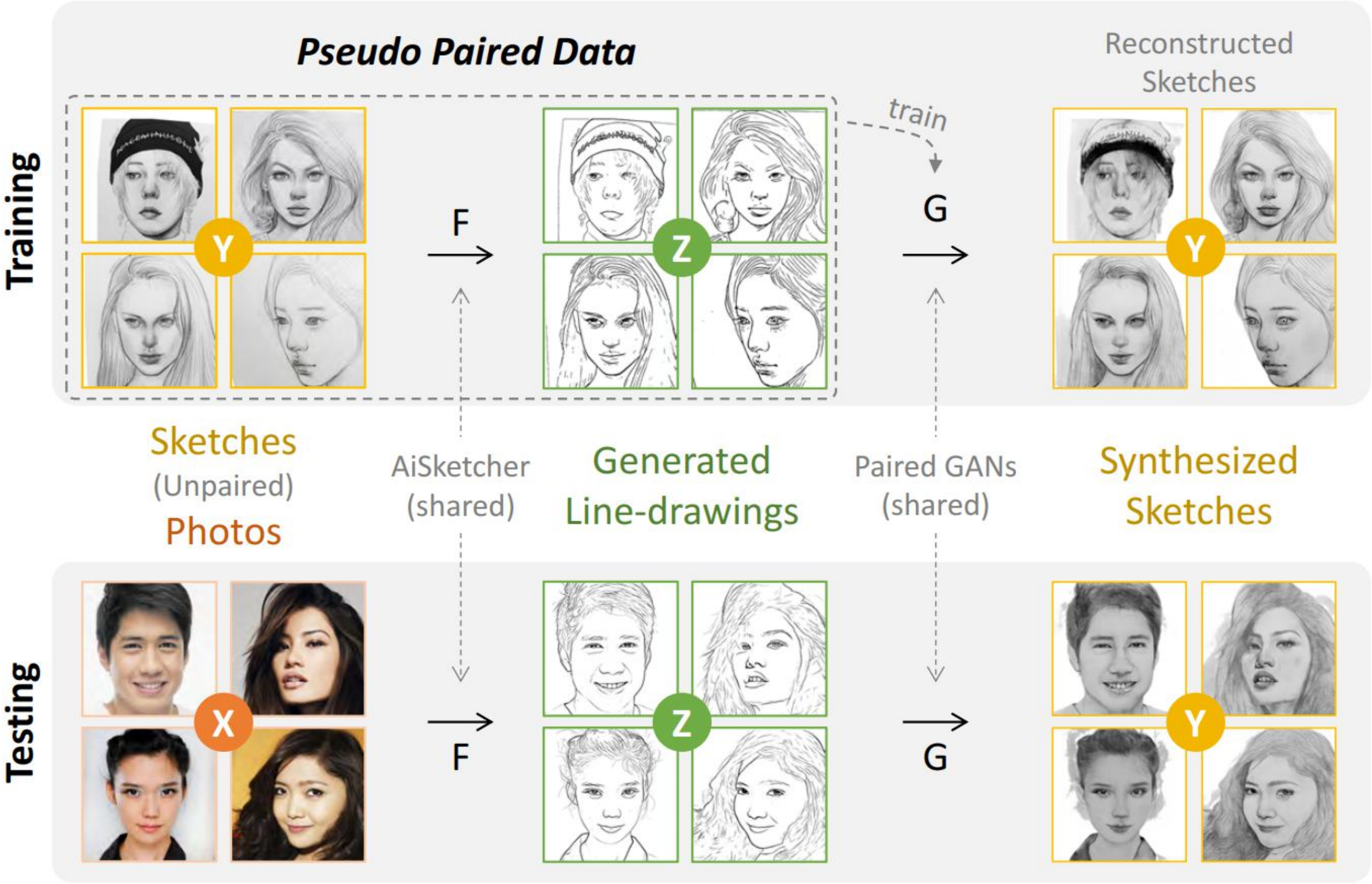
Fig. 1. Illustration of applying paired GANs, unpaired GANs, and our method to unpaired training samples.

Our main idea is

bridging the photo domain and the sketch domain by using the line-drawing domain.

Our work: sRender

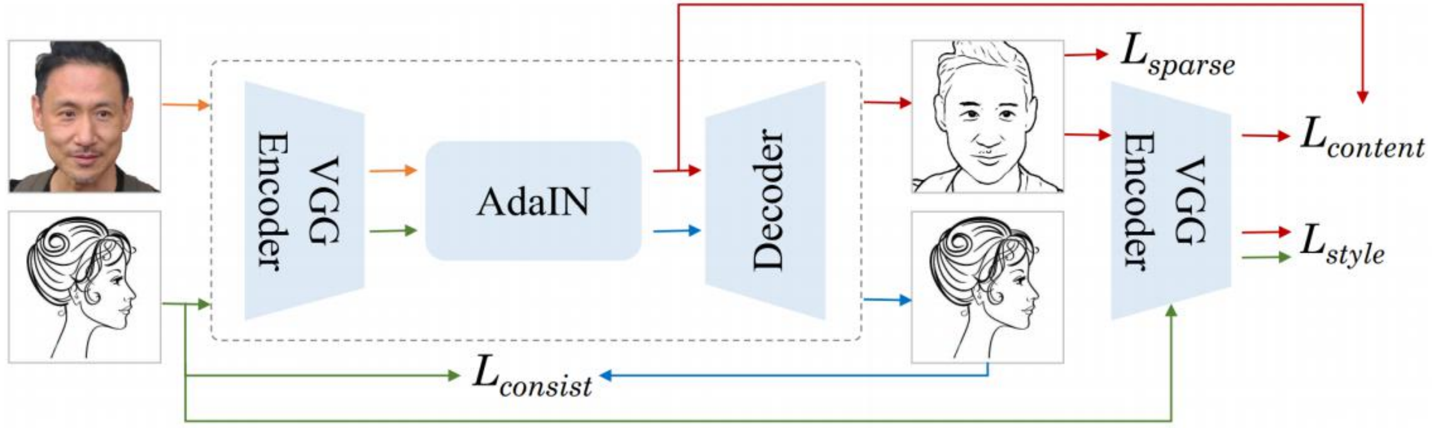
Our method



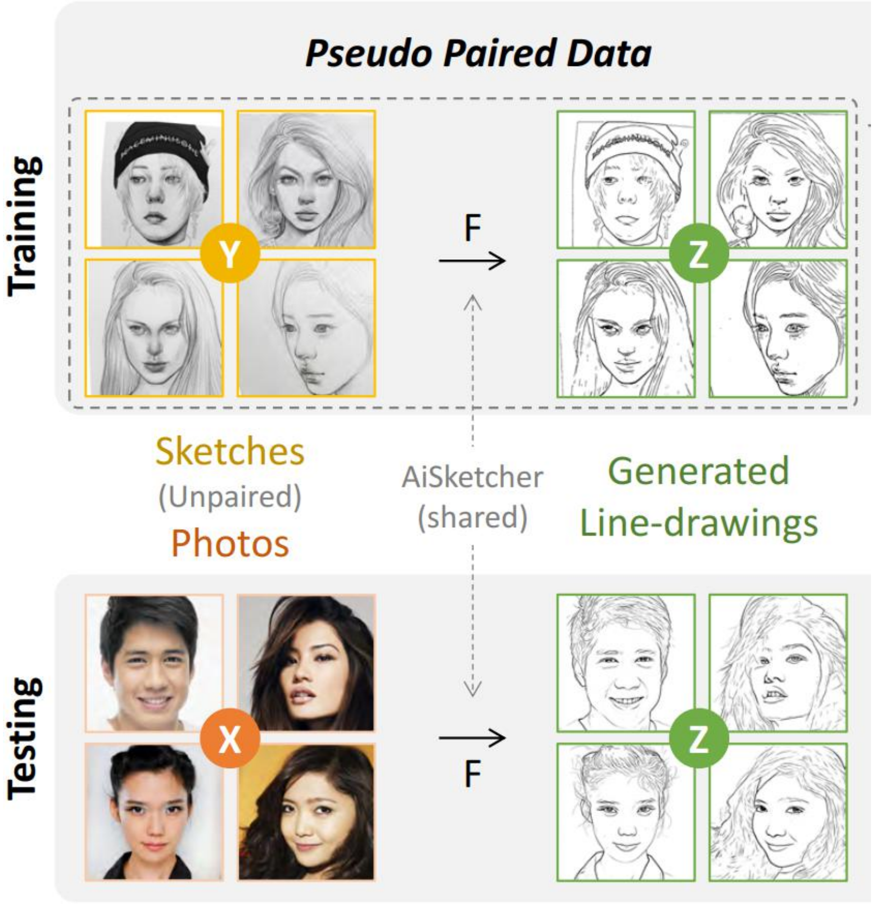
Pipeline of the proposed method

F: Line-Drawing Generation

- **AiSketcher: Neual Style Transfer (NST) based method, need no paired data.**



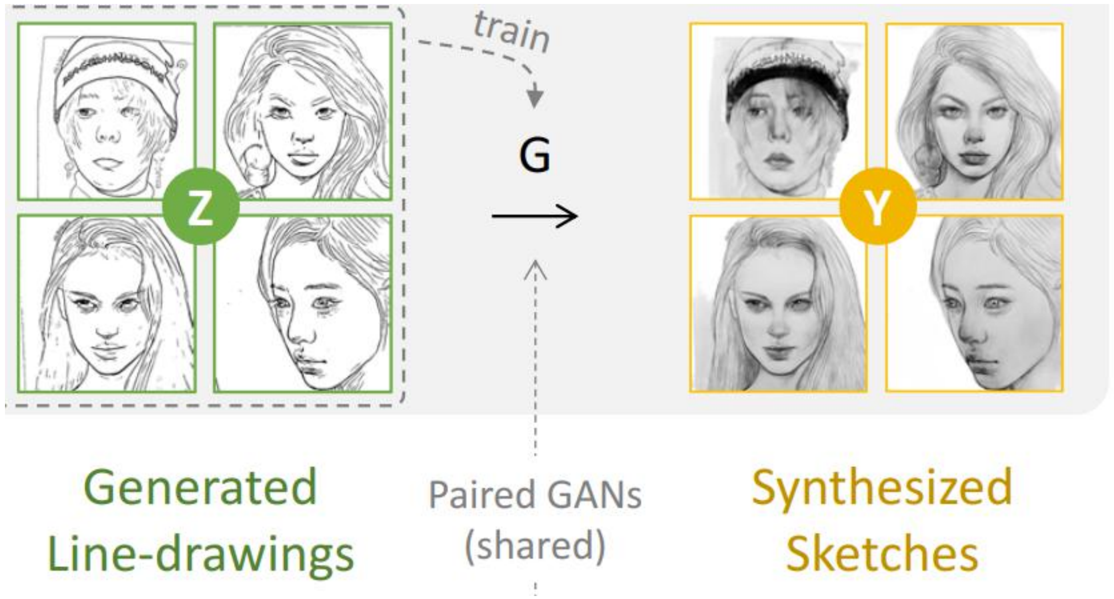
Pipeline of AiSketcher



Fei Gao, Jingjie Zhu, Zeyuan Yu, Peng Li, Tao Wang, "Making Robots Draw A Vivid Portrait In Two Minutes," in the Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2020), pp. 9585-9591, Las Vegas, USA, 2020. http://aiart.live/AiSketcher/

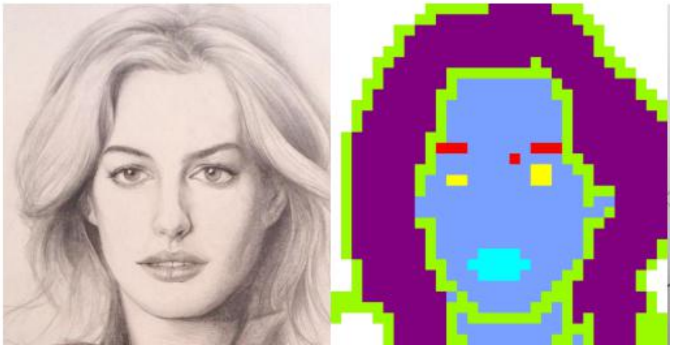
G: Sketch Synthesis Network

- **Generator**
  - 5 Covolutional layers, 9 residual blocks, and 5 Transposed Convolutional layers
- **Discriminator**
  - two discriminators, the same architecture with different scales of images
- **Activation function**
  - ReLU and leaky ReLU in the generator and discriminators, respectively
- **Normalization**
  - Instance normalization is used in all networks



Objectives

- **Stroke Loss**
  - human artists present different facial areas by using diverse strokes.
    - 7 types of strokes according to facial areas:
      skin, hair, boundary, eye brow, eye, clips, and ear.



  - train a CNN to correctly classify the stroke type of a given patch, and calculate the loss by

$$\mathcal{L}_{str} = \sum_{i=1}^{n} \sum_{j} \|\psi^j(y_i) - \psi^j(G(z_i))\|_2$$

- **Full Objective**
  - We additionally use perceptual loss, adversarial loss, the feature matching loss:

$$(G^* D_k^*) = \min_G \max_{D_k} \mathcal{L}_{adv} + \lambda_1 \mathcal{L}_{FM} + \lambda_2 \mathcal{L}_{VGG} + \lambda_3 \mathcal{L}_{str}$$

we set $\lambda_1 = 100$, $\lambda_2 = 10$ and $\lambda_3 = 0.002$

Experiments

Dataset

- **Collect from Web: about 300 sketches each type.**

croquis @HYEJUNG

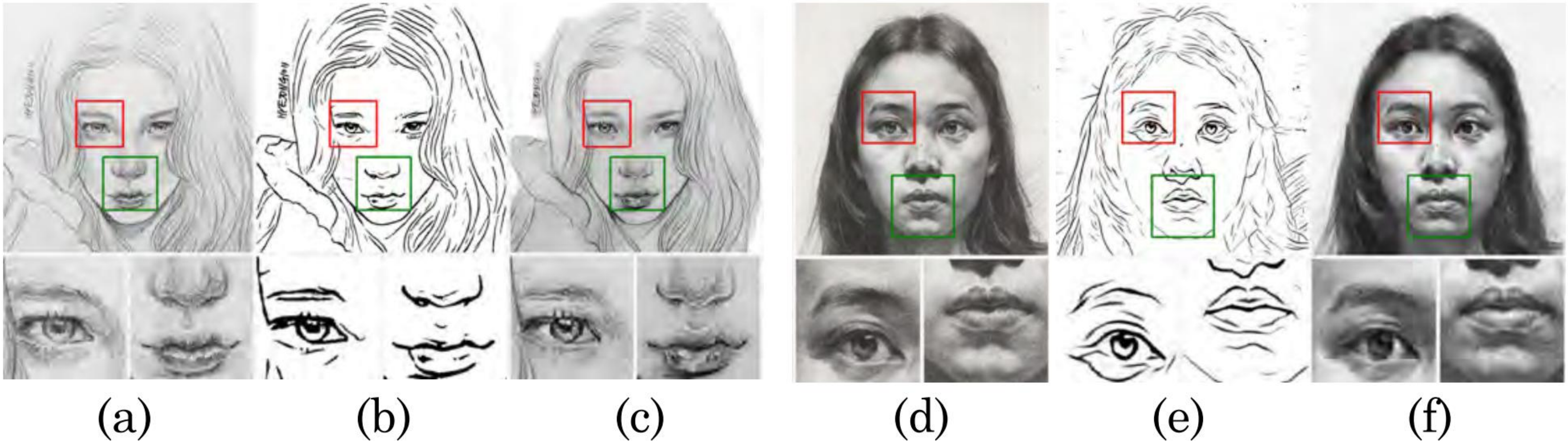
charcoal

Dataset

- **Photos： about 500 facial photos**

example of facial photos select from CelebA-HQ

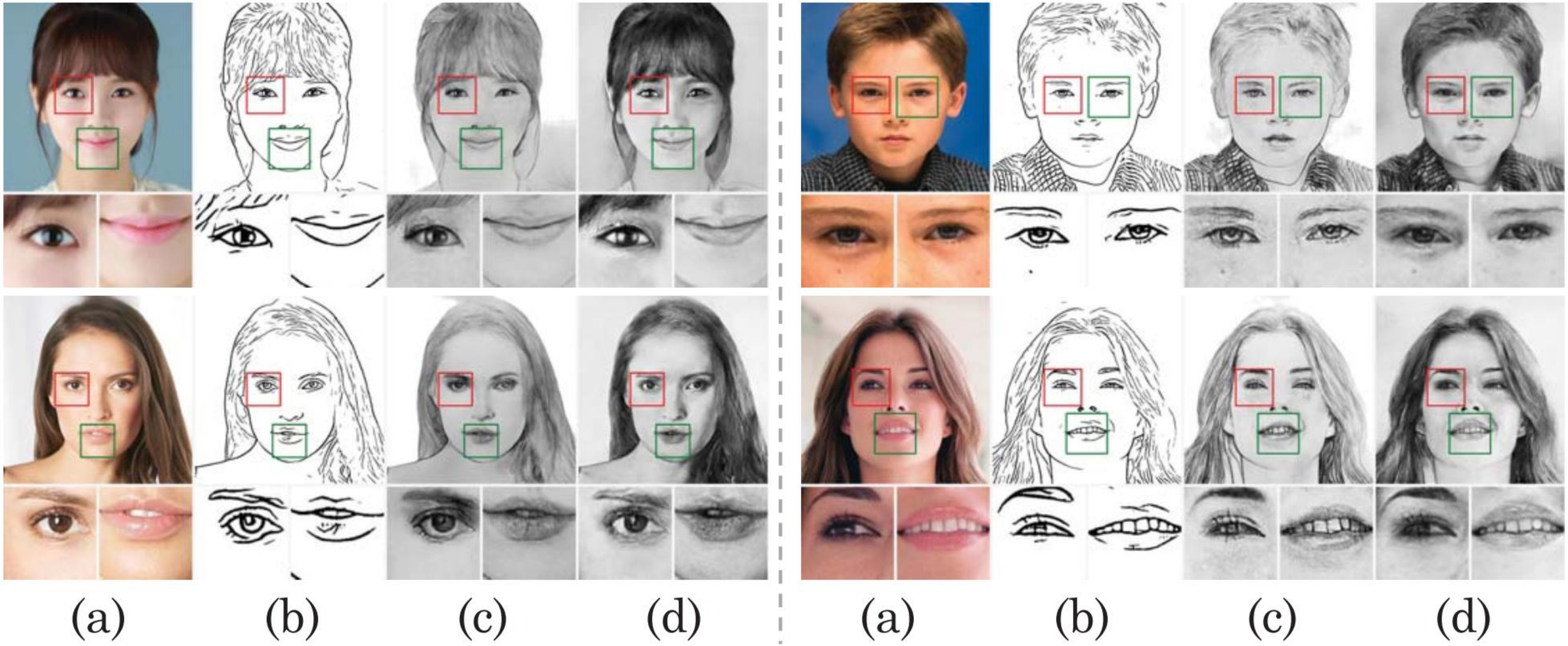
Sketch Reconstruction



(a)Real croquis sketch, (b) synthesised line-drawing, (c) reconstructed sketch;
(d) real charcoal sketch, (e) synthesised line-drawing, and (f) reconstructed charcoal sketch

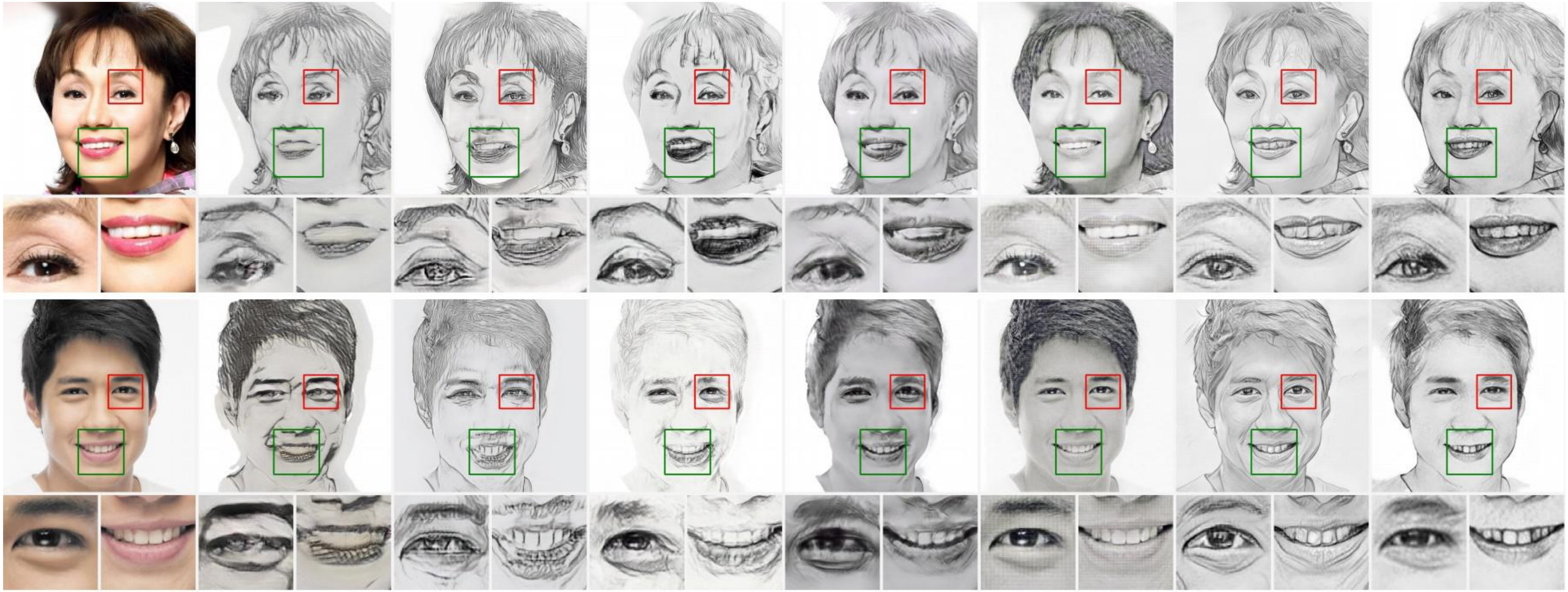
FID of **22.92** on the croquis sketches, and **12.30** on the charcoal sketches

Sketch Generation



(a) Input photo, (b)synthesised line-drawing, (c) generated croquis sketch, and(d) generated charcoal sketch

Comparison with SOTA unsupervised I2I translation methods



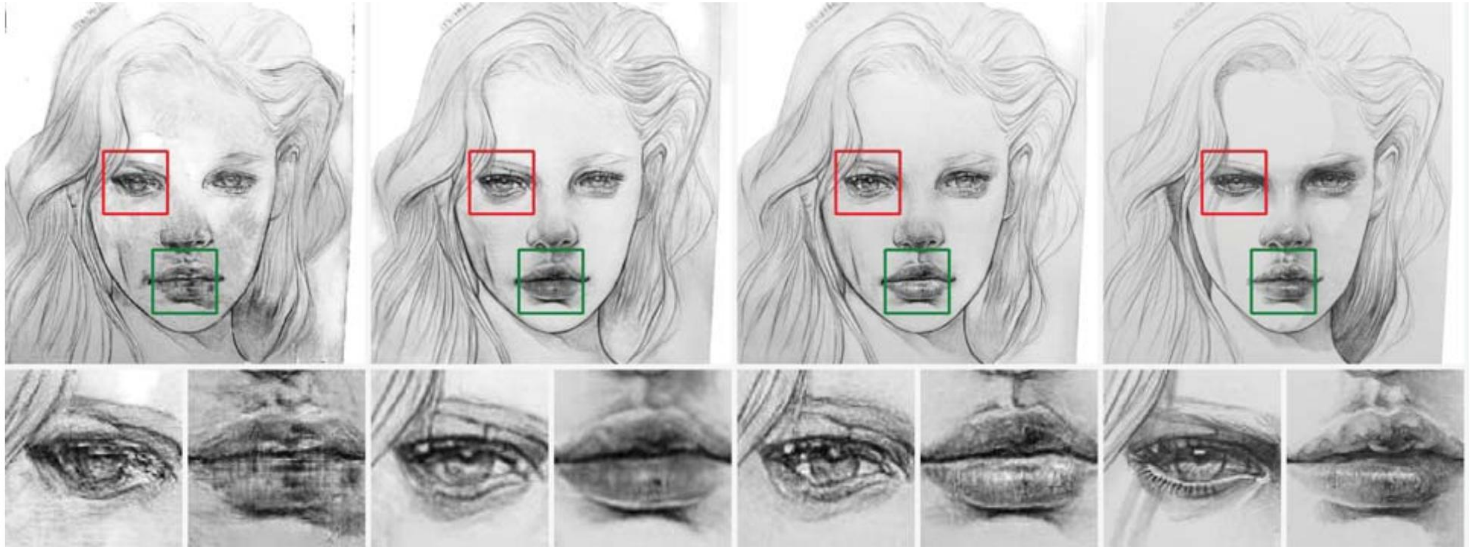
Input MUNIT DRIT NICE-GAN U-GAT-IT CycleGAN AdaIN sRender

	AdaIN [19]	CycleGAN [8]	MUNIT [15]	DRIT [16]	NICE-GAN [26]	U-GAT-IT [14]	sRender (Ours)
FID	49.43	45.51	46.35	42.80	<u>39.71</u>	48.26	**30.35**

Ablation Study

- **Croquis sketch Reconstruction**
  - (a) $\text{sRender}_{\text{Pix2Pix}}$,
  - (b)sRender w/o L_{str},
  - (c) sRender, and
  - (d) the ground truth.

- **Superiority**
  - Pix2PixHD > Pix2Pix
  - L_{str} leads to realistic textures



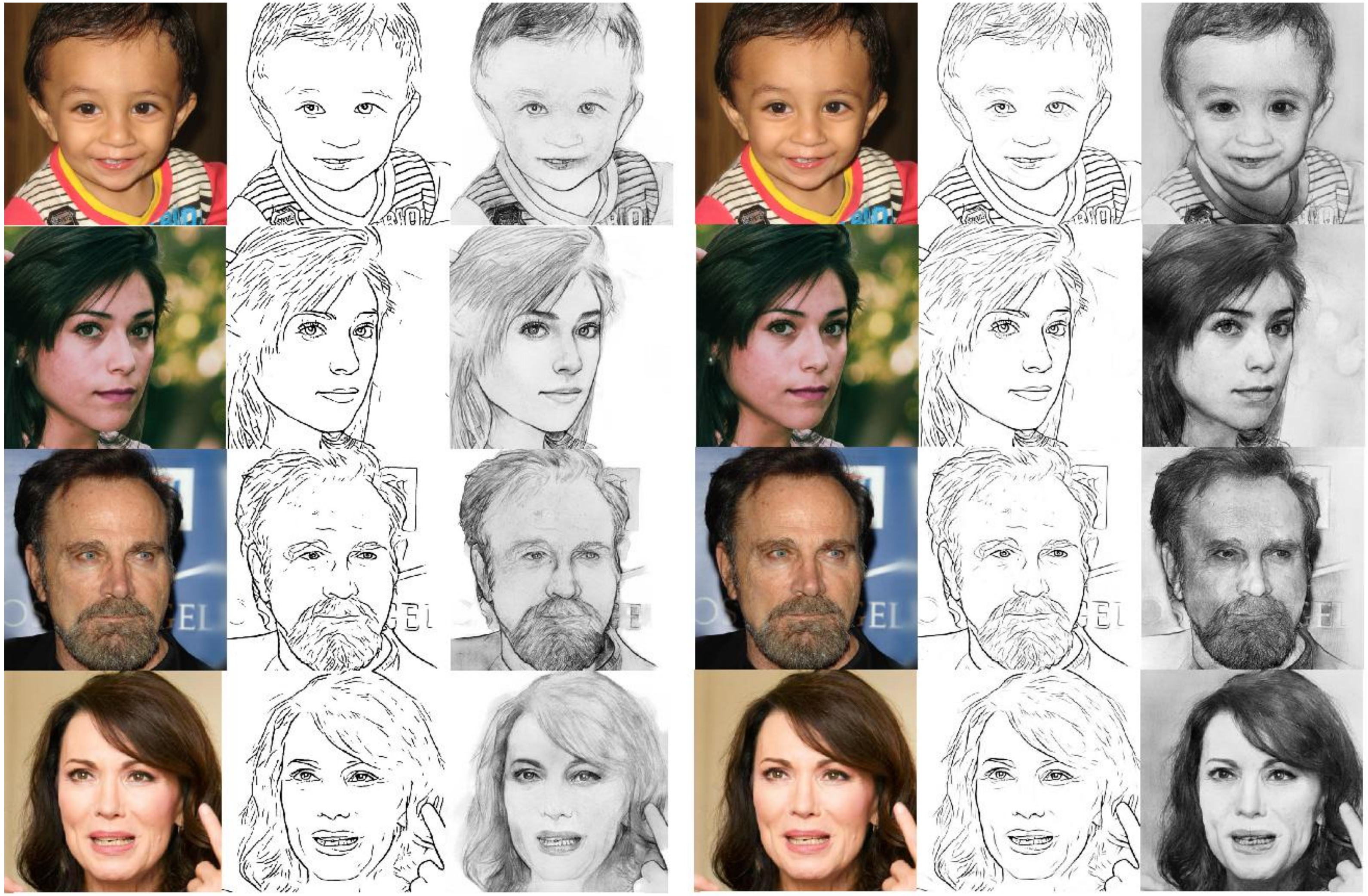
(a) (b) (c) (d)

	$\text{sRender}_{\text{Pix2Pix}}$	sRender w/o $\mathcal{L}_{str}$	sRender
FID	37.49	22.97	**22.92**
Scoot	0.557	0.570	**0.587**
Acc.	0.672	0.739	**0.750**

Summary

Conclusions

- **Contributions**
  - A novel Framework
  - Stroke Loss
  - SOTA results
  - https://aiart.live/sRender
- **Future Work**
  - Quality improvement
  - Various styles of artistic portraits, e.g. oil paintigs
  - Semi-supervised learning
  - Few-shot learning



(a) (b) (c) (a) (b) (c)

left: croquis style; right: charcoal style
(a) photos, (b) line-drawings, (c)generated sketches

Demonstration

- **You can try extensions of this work in WeChat or Browser:**

WeiXin (WeChat)

WEB API

References (selected)

- Fei Gao, Jingjie Zhu, Zeyuan Yu, Peng Li, Tao Wang, "Making Robots Draw A Vivid Portrait In Two Minutes," in the Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2020), pp. 9585-9591, Las Vegas, USA, 2020.
- Ting-Chun Wang and Ming-Yu Liu and Jun-Yan Zhu and Andrew Tao and Jan Kautz and Bryan Catanzaro, High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018.
- Huang, Xun and Belongie, Serge, Arbitrary Style Transfer in Real-time with Adaptive Instance Normalization, ICCV, 2017.
- Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks. In CVPR 2017.
- Jun-Yan Zhu*, Taesung Park*, Phillip Isola, Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In ICCV 2017. (* equal contributions)

智能可视建模与仿真实验室

Intelligent Visual Modeling & Simulation (iGame) Lab

Thank You !

Q & A