GAN Application in Mobile Devices

Gang YU

http://skicyyu.org/

iskicy@gmail.com

Outline

- Basics in GAN
- GAN for Face
- GAN for Human
- GAN for Font
- GAN as A Tool
- Conclusion

Basics in GAN

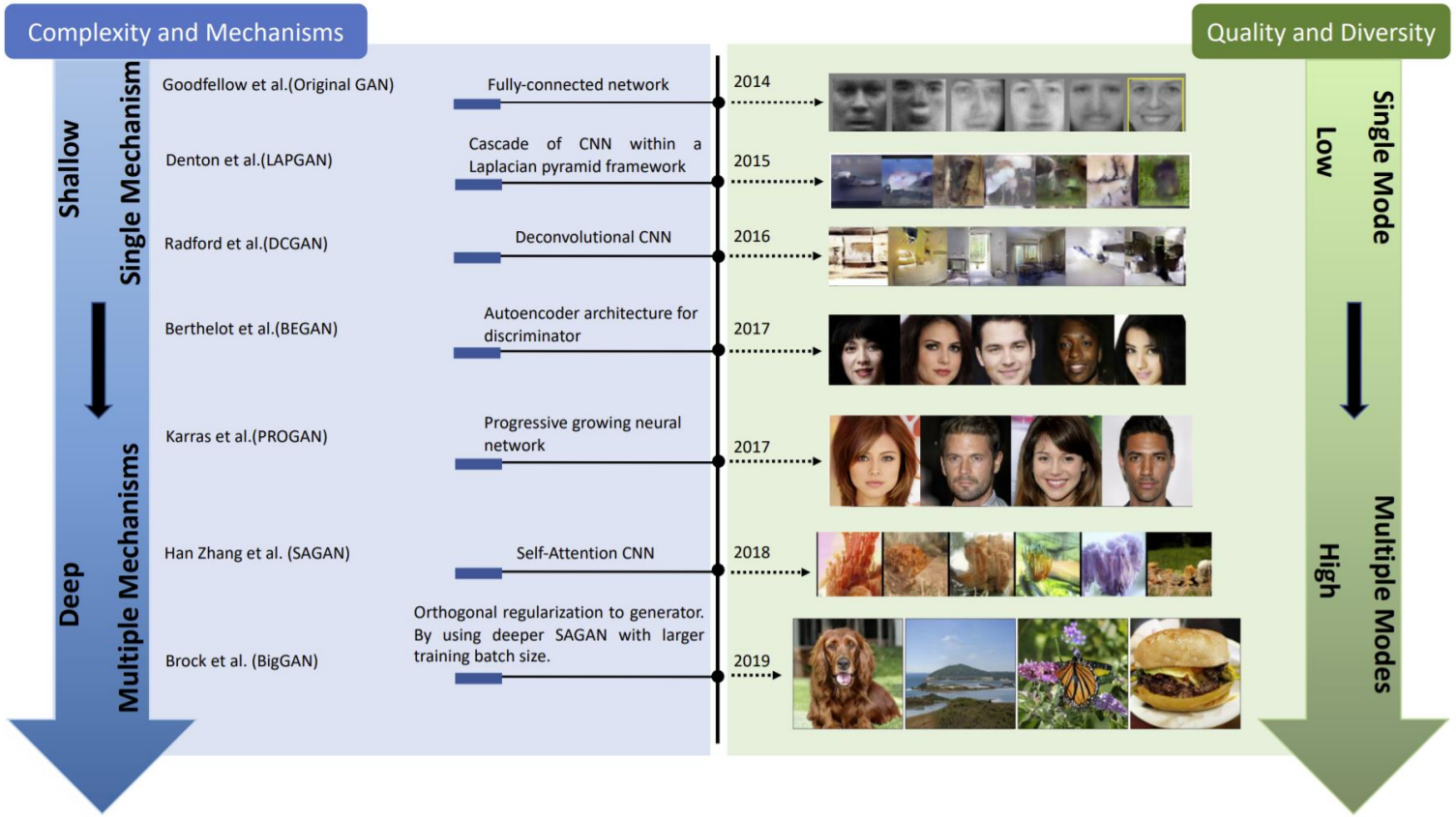
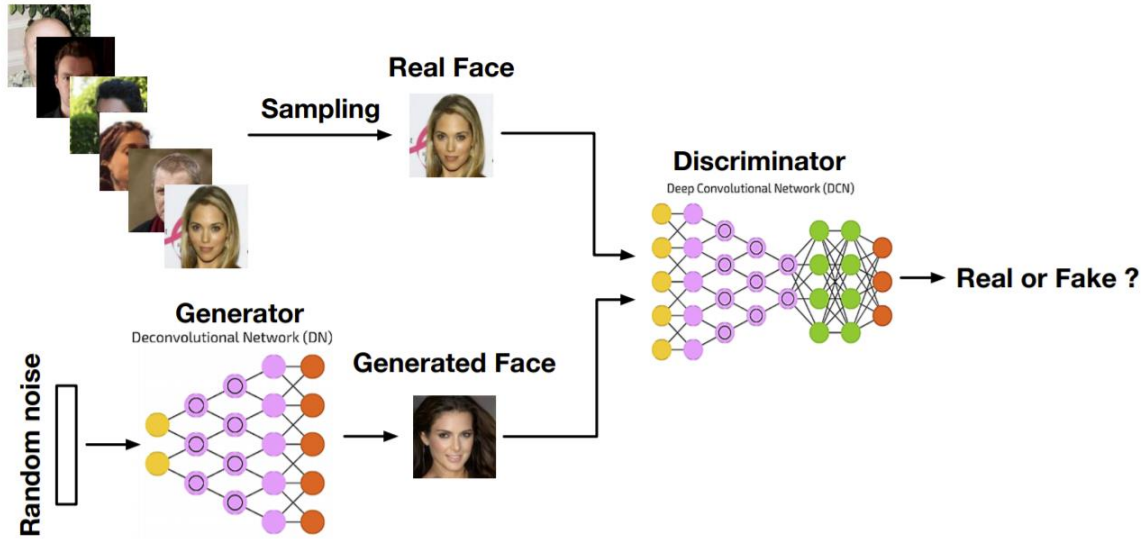


Figure from: Generative Adversarial Networks: A Survey and Taxonomy
https://arxiv.org/pdf/1906.01529

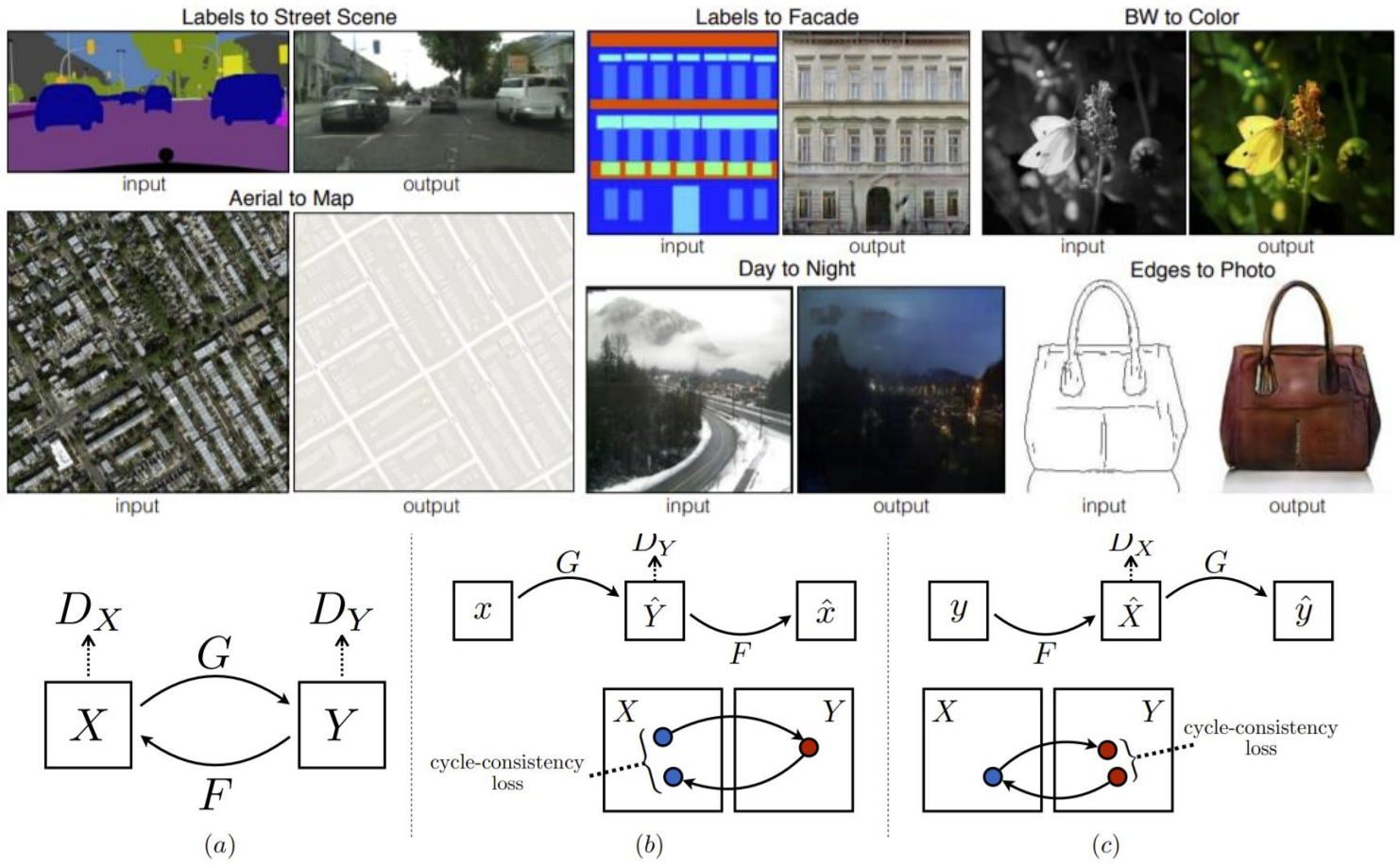
Basics in GAN



$$\min_G \max_D \mathbb{E}_{\mathbf{x}\sim p_r} \log[D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_\mathbf{z}} \log\left[1 - D(G(\mathbf{z}))\right].$$

Figure from: Generative Adversarial Networks: A Survey and Taxonomy
https://arxiv.org/pdf/1906.01529

Basics in GAN:Pix2Pix – CycleGAN - StyleGAN



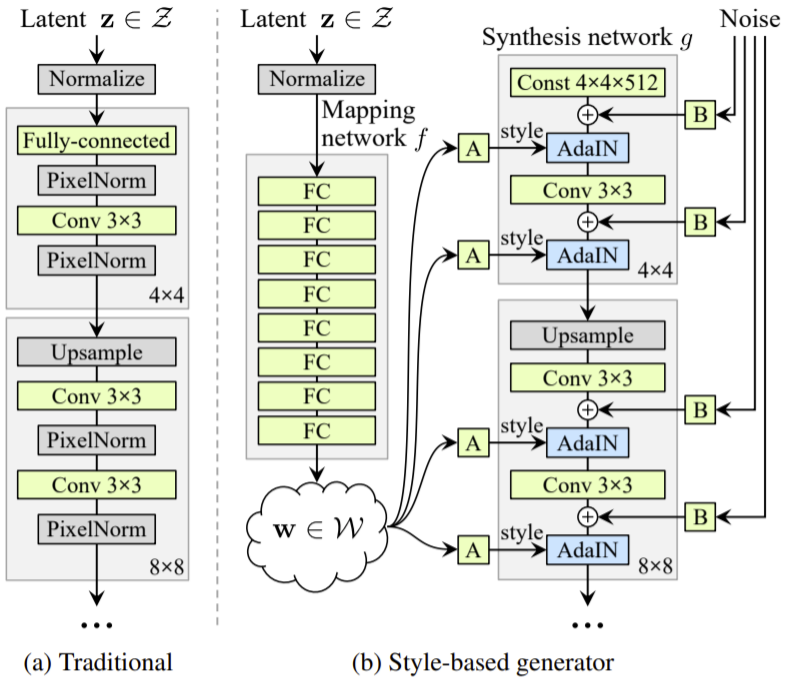


Image-to-Image Translation with Conditional Adversarial Networks https://arxiv.org/abs/1611.07004
Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks https://arxiv.org/abs/1703.10593
A Style-Based Generator Architecture for Generative Adversarial Networks https://arxiv.org/abs/1812.04948

Outline

- Basics in GAN
- GAN for Face
- GAN for Human
- GAN for Font
- GAN as A Tool
- Conclusion

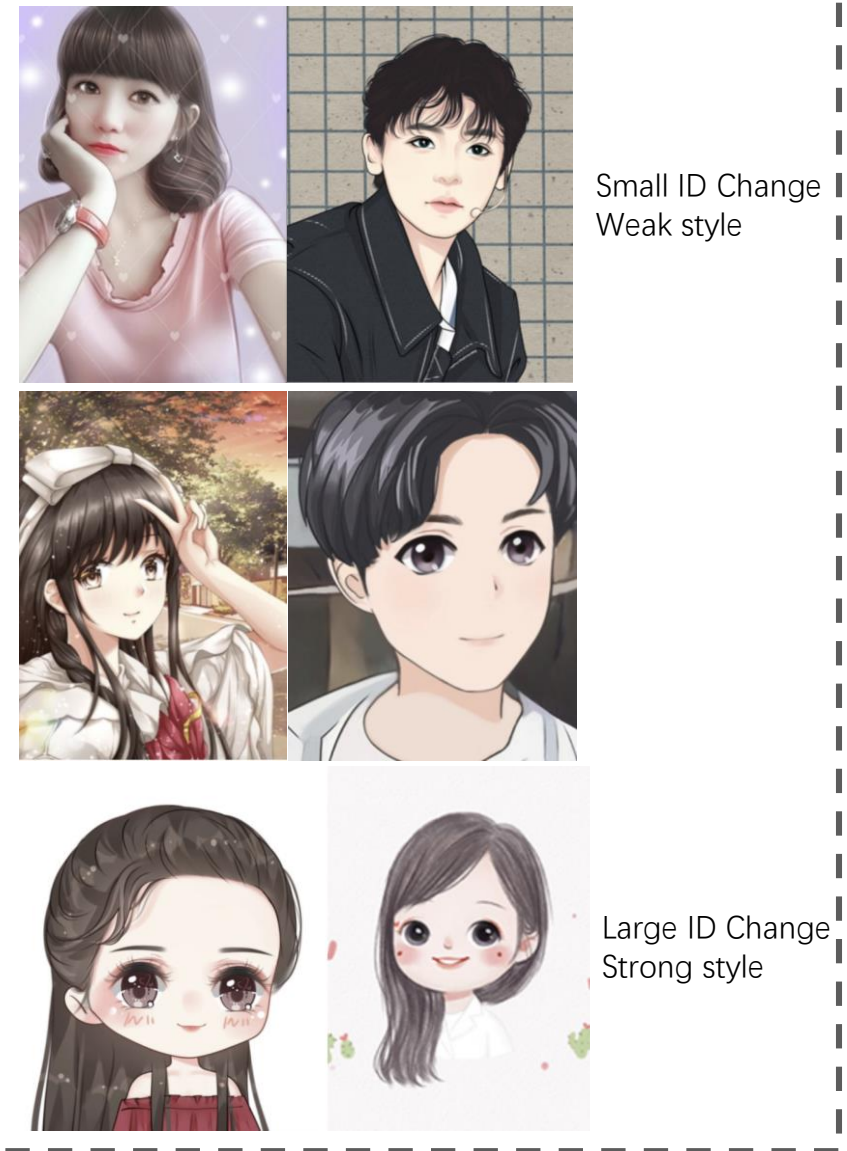
GAN for Face: which one is Generated?

GAN for Face: Overview

- Application
  - Face Cartoon Generation
  - Face Attribute editing
  - Face Reenactment
- Foundations
  - ID vs Attribute
  - Model Compression
  - Data

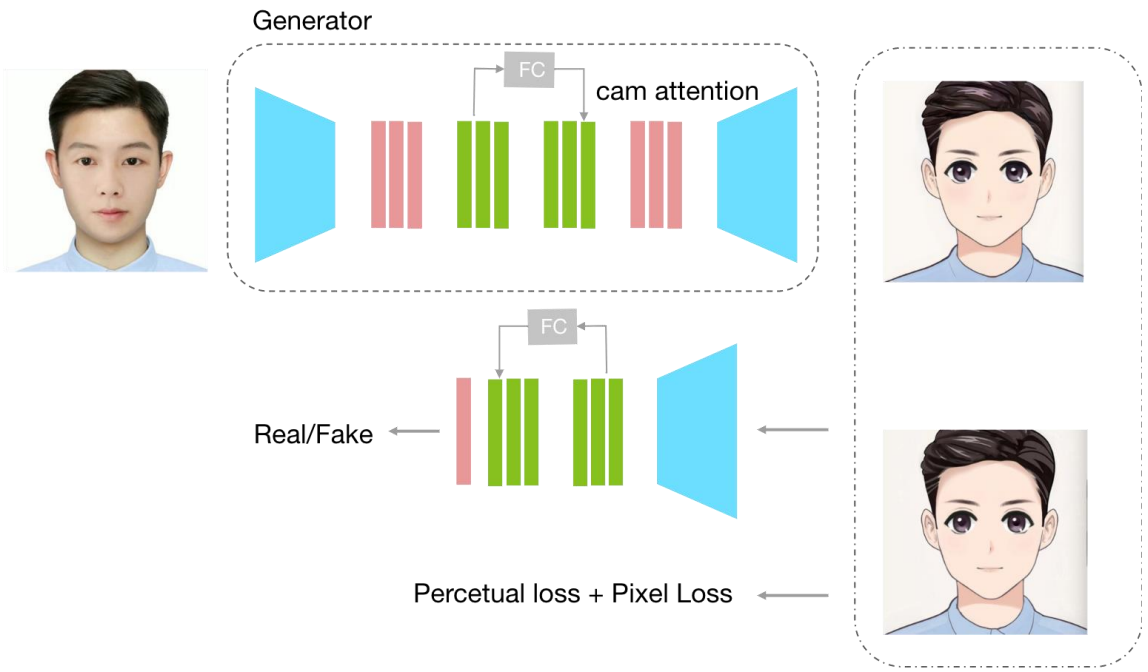
GAN for Face: Face Cartoon Generation

- Face Cartoon
  - Maintain both the cartoon **style** and face **ID** feature
- Challenges
  - Limited training data
  - Robustness for the generation
  - Fast speed for mobile devices

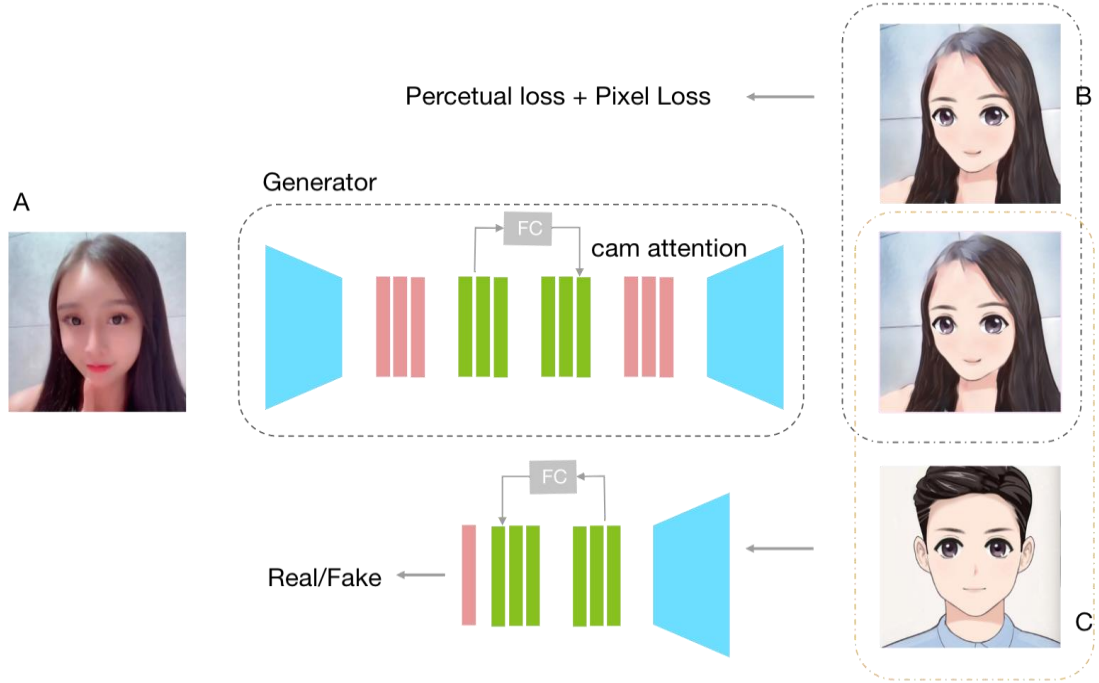


GAN for Face: Face Cartoon Generation

First Stage



Second stage

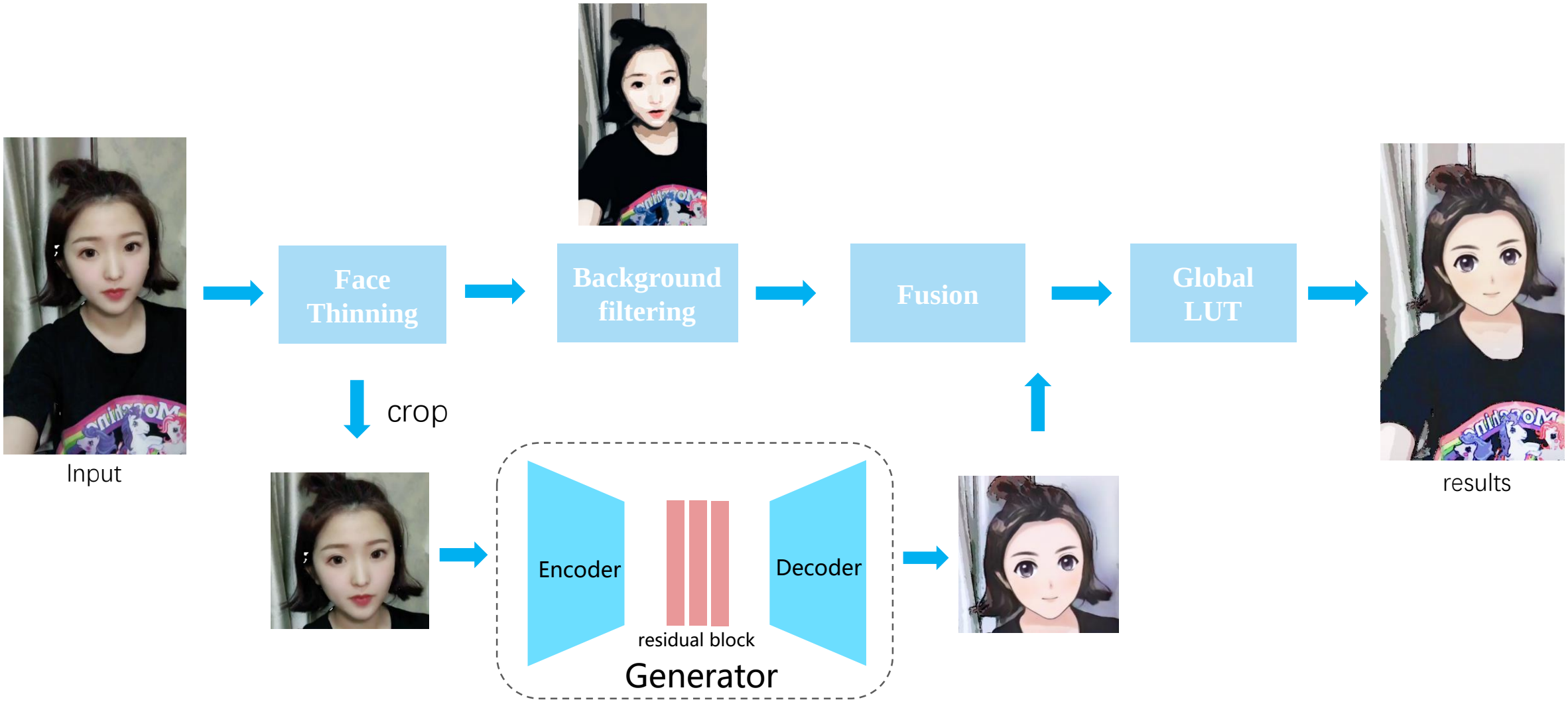


$$Loss = L_{gan}^{xy} + L_{pixel}^{xy} + L_{perceptual}^{xy}$$

$$= (E_{x\sim B}[(D_y(x)] + E_{x\sim C}[(1 - D_y(G_{xy}(x))]) + E_{x\sim A,y\sim B}[|y - G_{xy}(x)|] + E_{x\sim A,y\sim B}[|E_{vgg}(y) - G_{vgg}(G_{xy}(x))|]$$

A is a set of real faces, B is a set of cartoon faces,
A is paired with B, C is a small set of cartoon faces for adversarial training

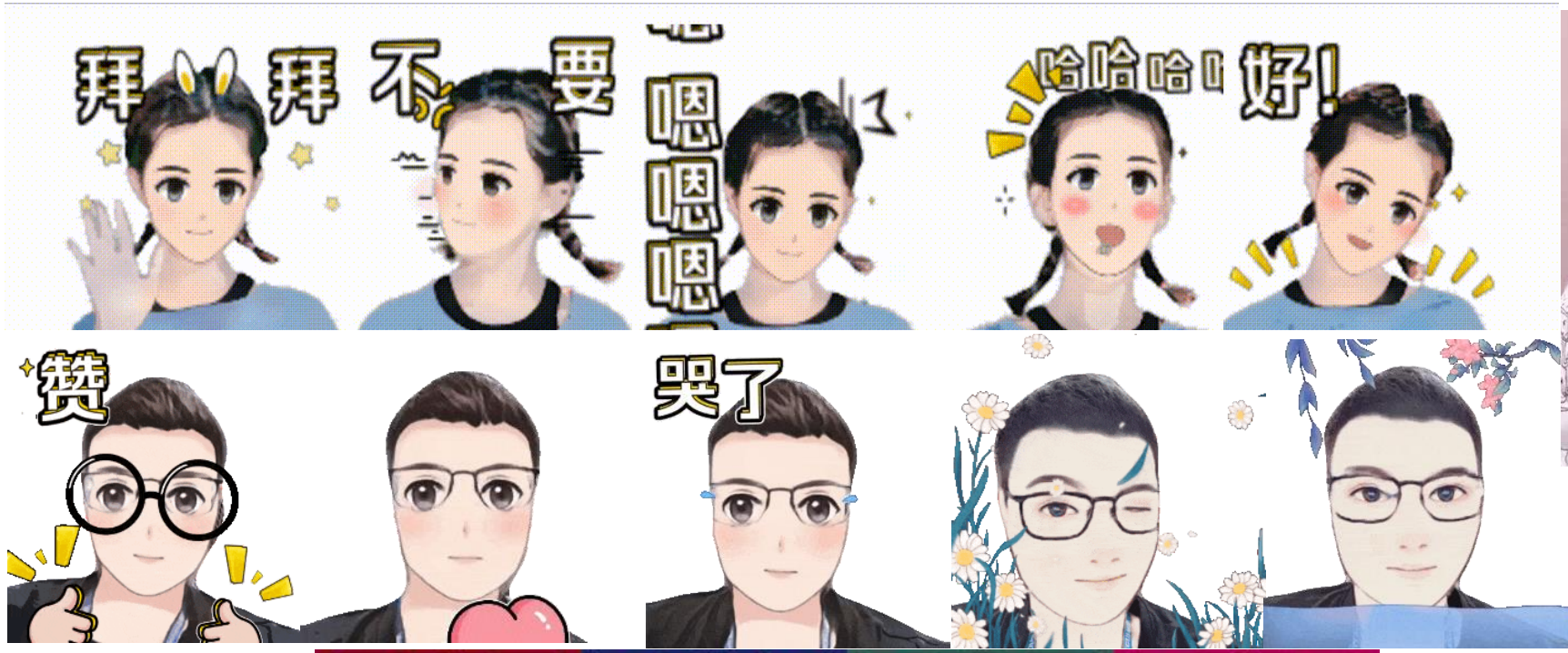
GAN for Face: Face Cartoon Generation Pipeline



GAN for Face: Demo Results

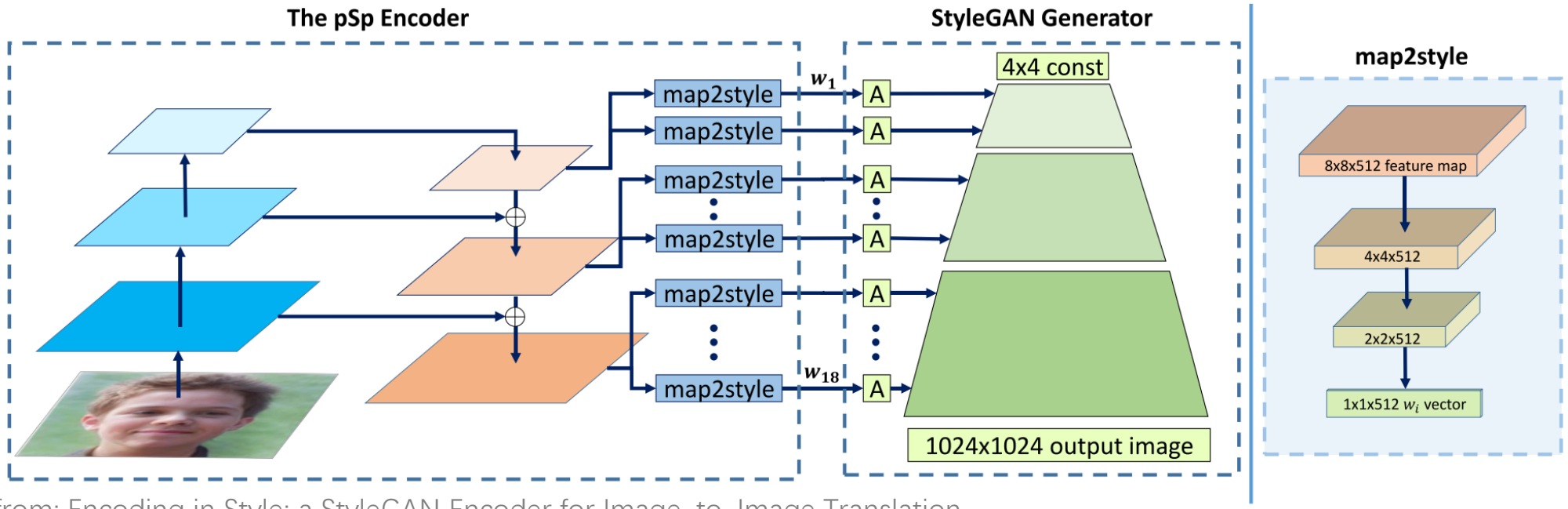
GAN for Face: Demo results

GAN for Face: More results



Face Attribute Editing： Pipeline

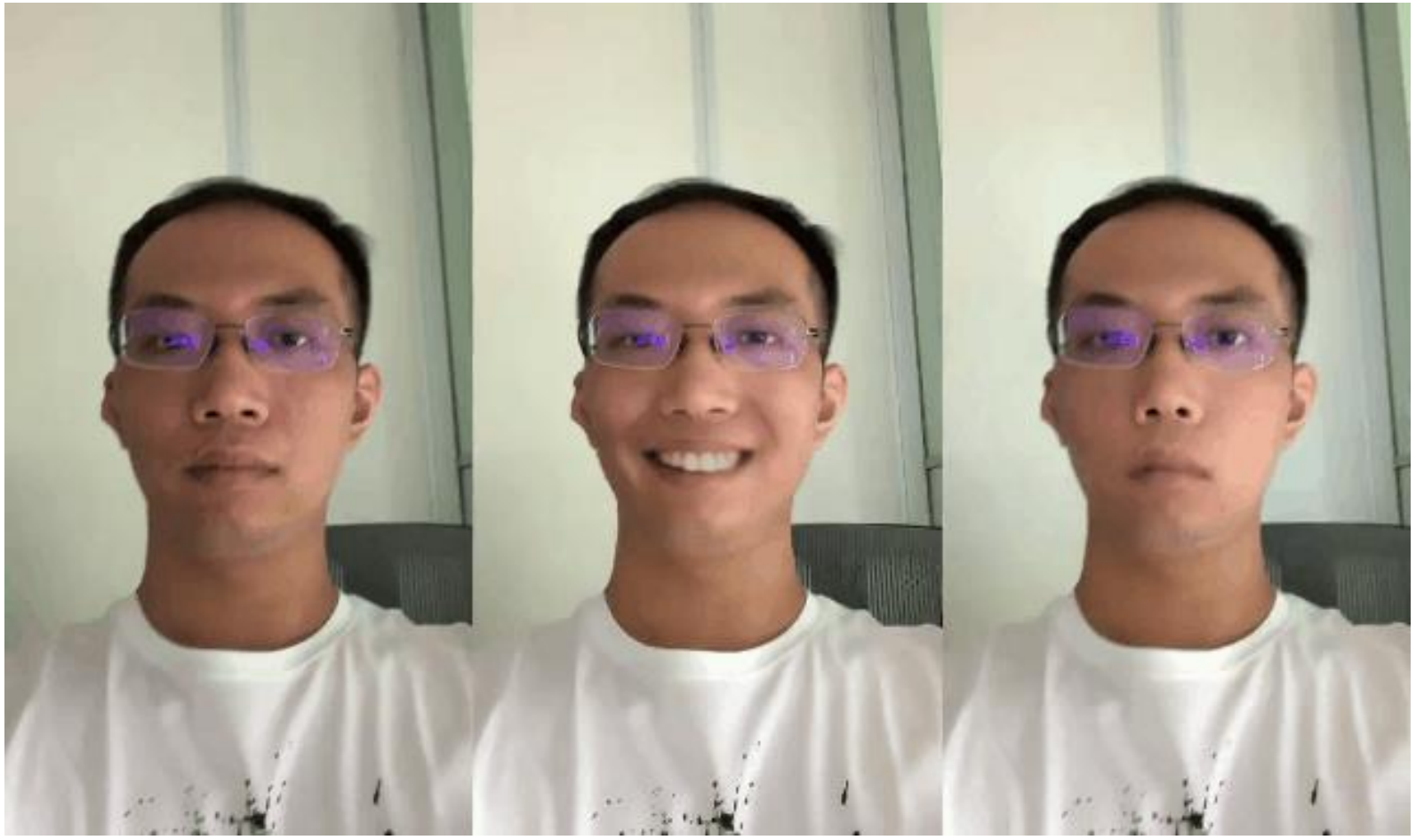
- Following StyleGAN
  - ID + Attribute
- StyleGAN pipeline: GAN inversion + attribute disentangle



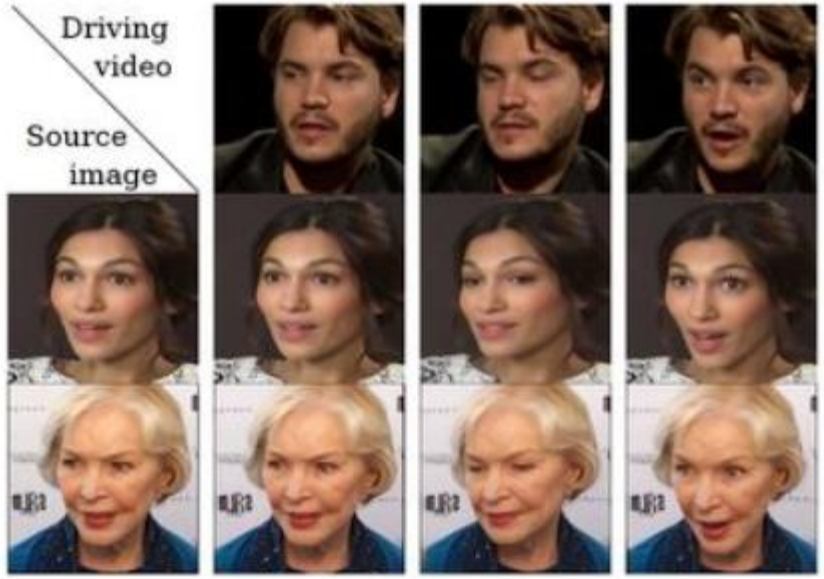
One example from: Encoding in Style: a StyleGAN Encoder for Image-to-Image Translation
https://arxiv.org/abs/2008.00951

Face Attribute Editing： Demo results

Face Attribute Editing： Demo results



Face Reenactment: Pipeline



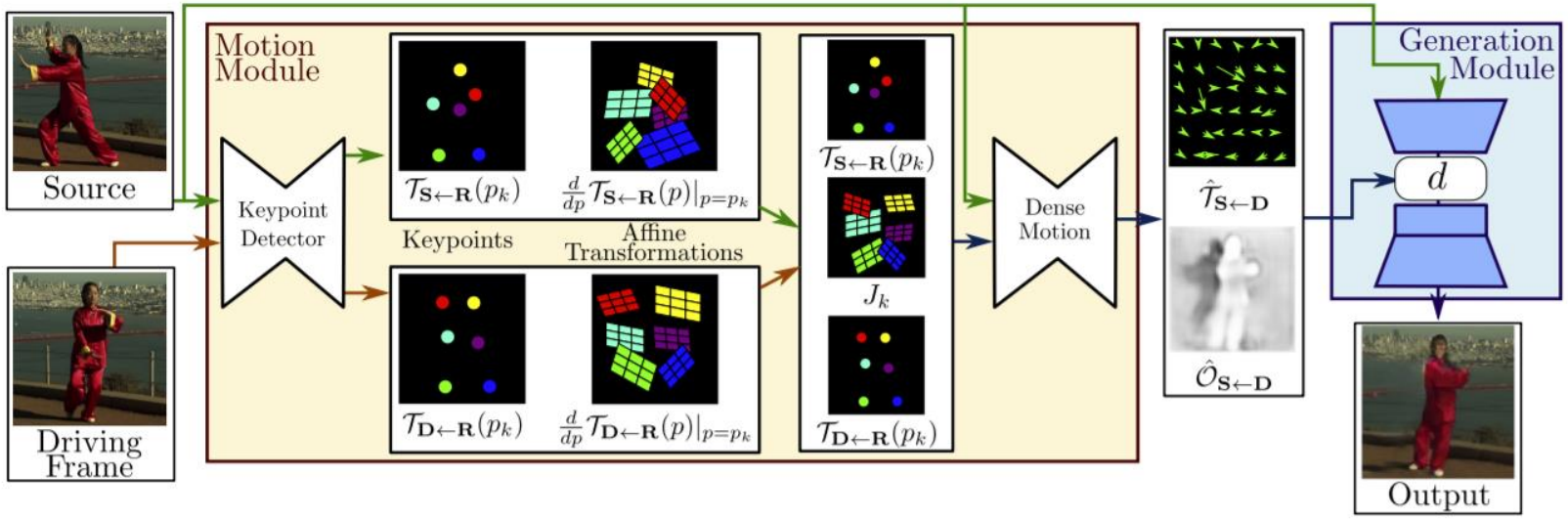


Figure from: First Order Motion Model for Image Animation
https://arxiv.org/abs/2003.00196

Face Reenactment： Results

GAN for Face： Model Compression

- Mobile Inference framework： TNN (https://github.com/Tencent/TNN)
- Following the structure based Depth-wise convs
  - 100M – 1G FLOPs
  - OP optimization for mobile devices
  - INT8 quantization
  - NAS model search

GAN for Face：Data

- Pix2Pix rely on paired data. But it is difficult to obtain the paired data
  - CycleGAN: training with unpaired data
  - StyleGAN
    - Generating data based on random noise
    - Following GAN Inversion and edit the face images
    - StyleGAN Fusion
  - Self-supervised training
  - Manual collection & labeling

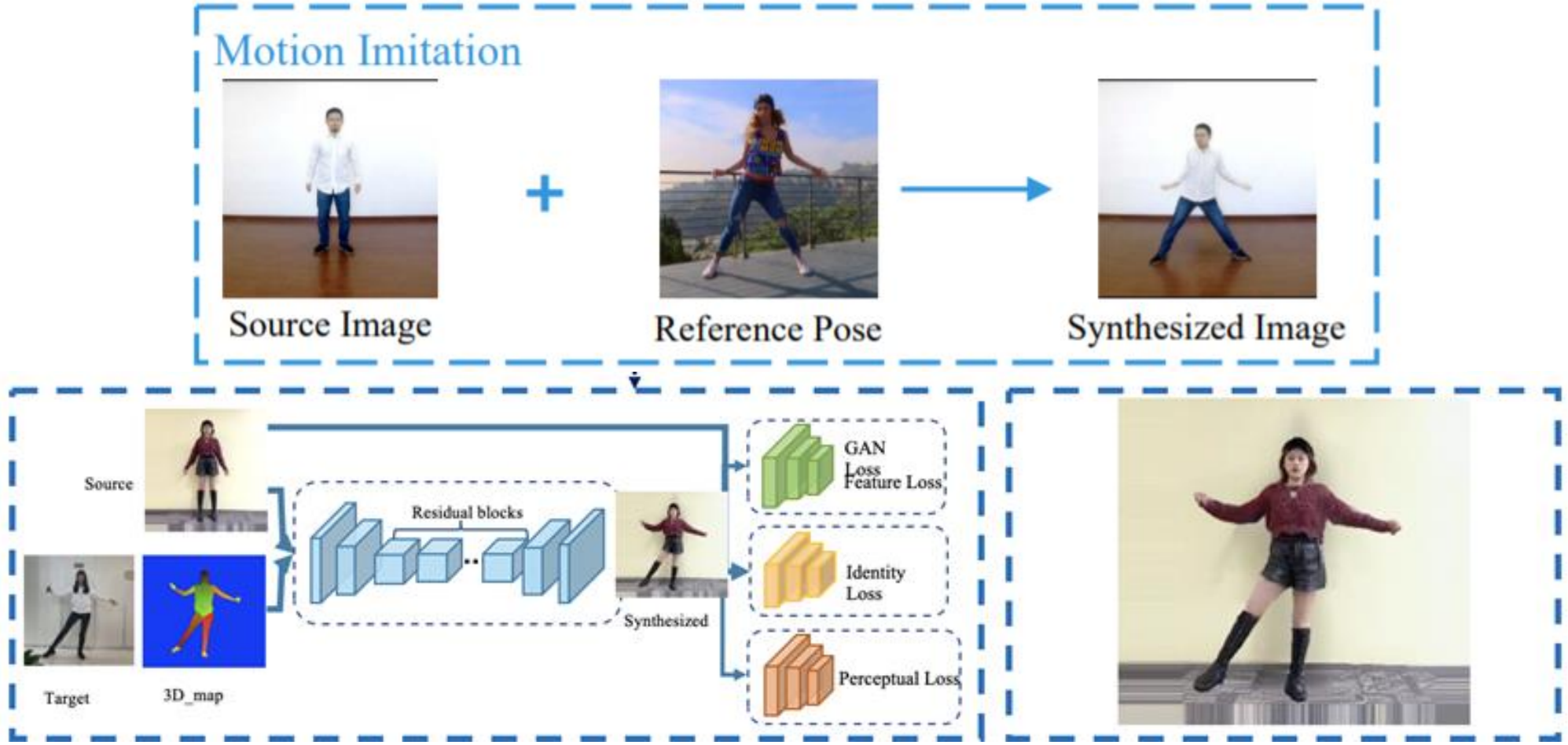
GAN for Face： Robustness

- Optimizing the Badcase
- Consistence for the data

Outline

- Basics in GAN
- GAN for Face
- GAN for Human
- GAN for Font
- GAN as A Tool
- Conclusion

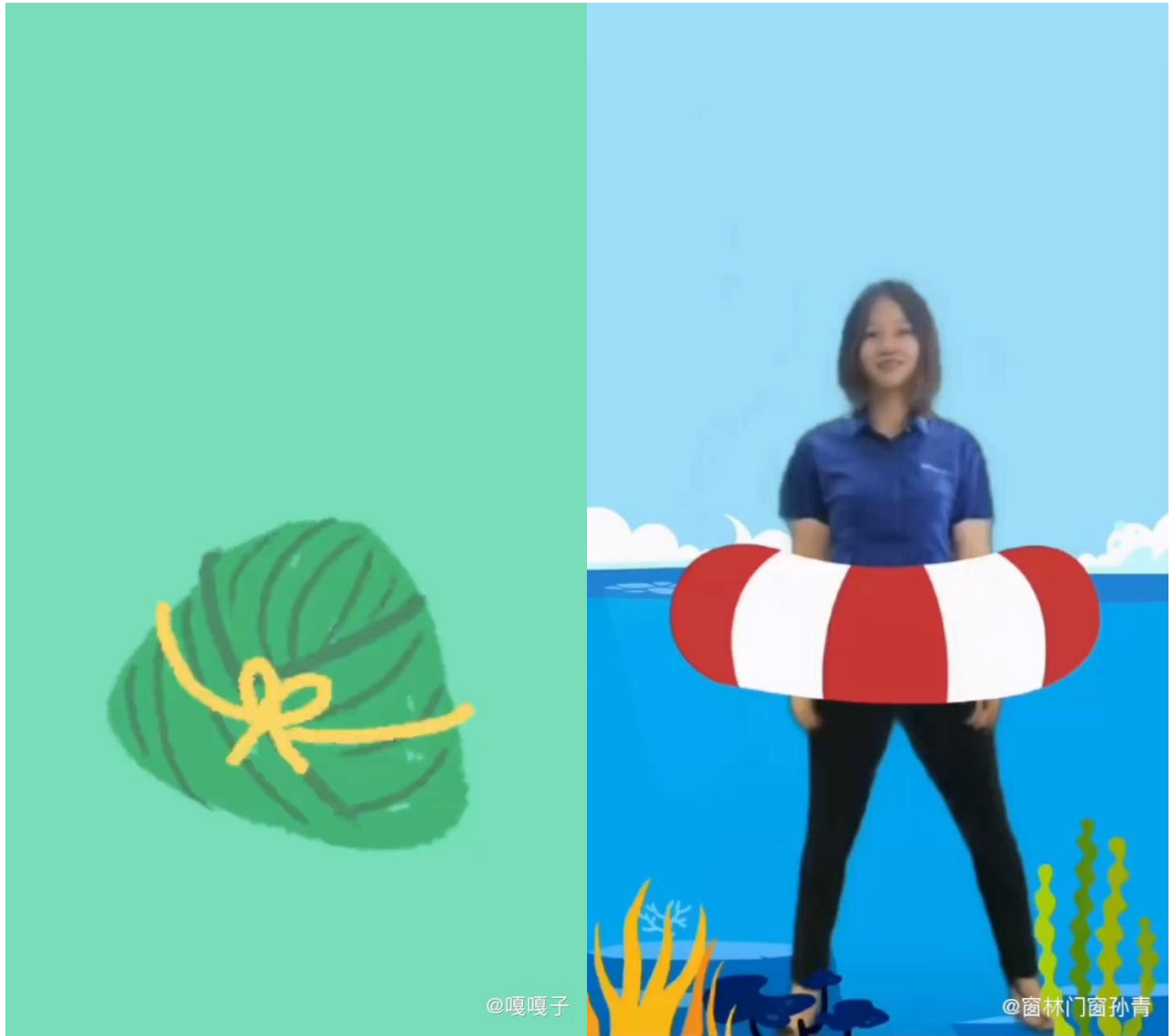
Human pose Transfer： Pipeline



Liquid Warping GAN: A Unified Framework for Human Motion Imitation, Appearance Transfer and Novel View Synthesis
https://arxiv.org/abs/1909.12224

Human pose Transfer： Results

Human pose Transfer： Results

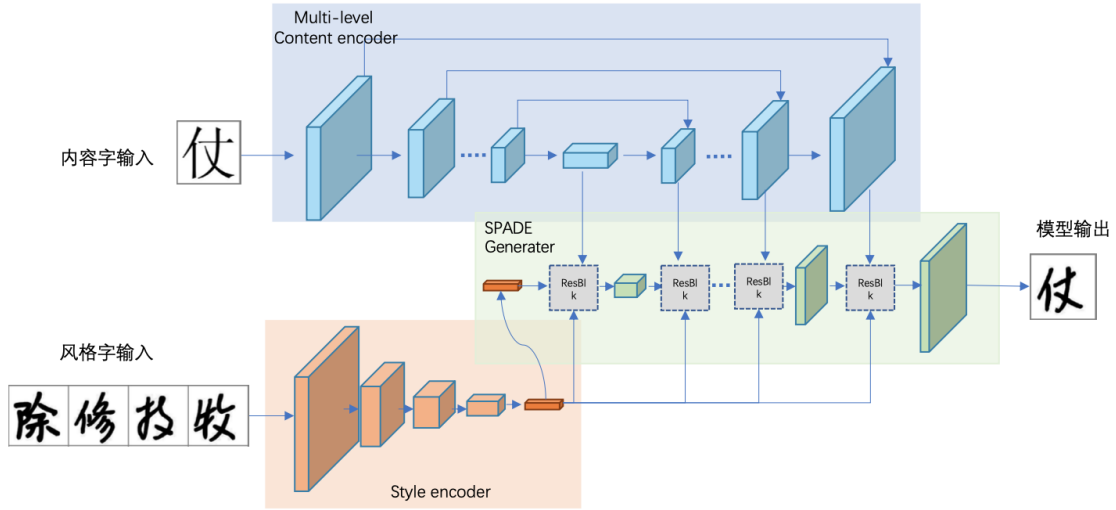


Outline

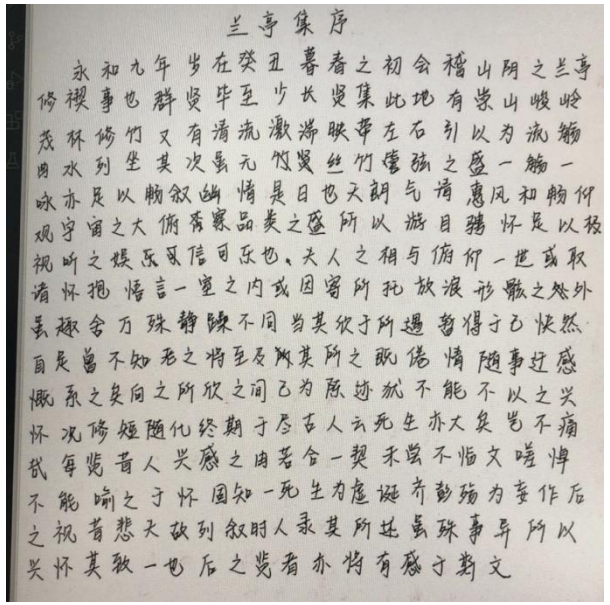
- Basics in GAN
- GAN for Face
- GAN for Human
- GAN for Font
- GAN as A Tool
- Conclusion

Font Generation

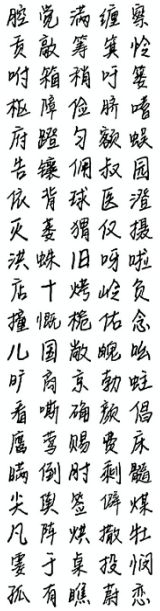
FontShifter



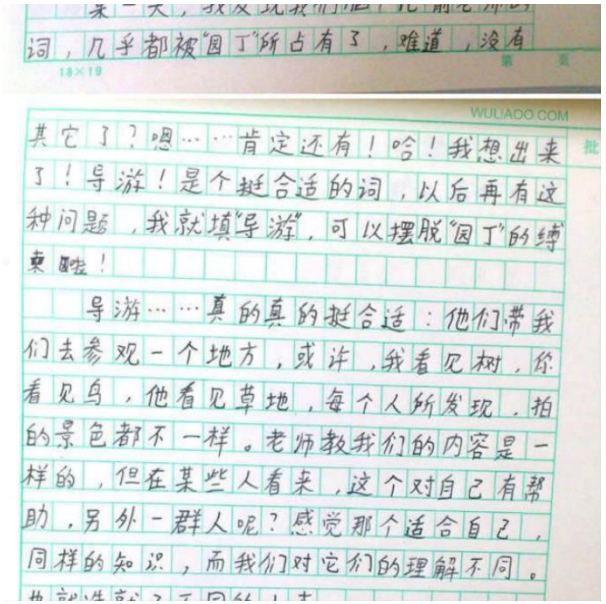
Input



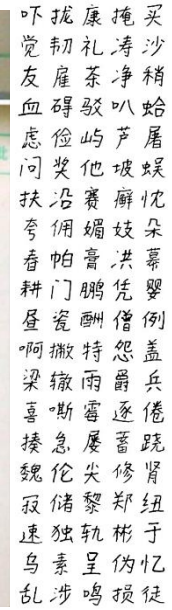
results



Input



Results



Outline

- Basics in GAN
- GAN for Face
- GAN for Human
- GAN for Font
- GAN as A Tool
- Conclusion

GAN as A Tool

- 3D Reconstruction

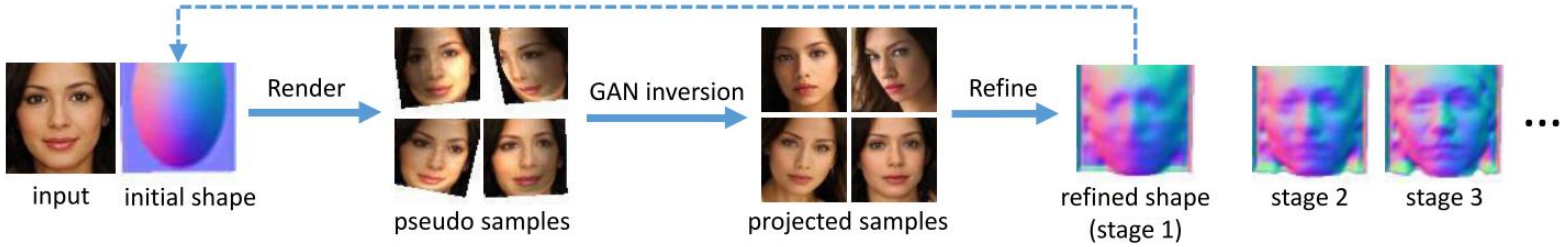


Figure 2: **Framework outline.** Starting with an initial ellipsoid 3D shape (viewed in surface normal), our approach renders various *'pseudo samples'* with different viewpoints and lighting conditions. GAN inversion is applied to these samples to obtain the *'projected samples'*, which are used as the ground truth of the rendering process to refine the initial 3D shape. This process is repeated until more precise results are obtained.

Do 2D GANs Know 3D Shape? Unsupervised 3D shape reconstruction from 2D Image GANs
https://arxiv.org/abs/2011.00844

- Dataset-Labeling

DatasetGAN: Efficient Labeled Data Factory with Minimal Human Effort
https://arxiv.org/abs/2104.06490

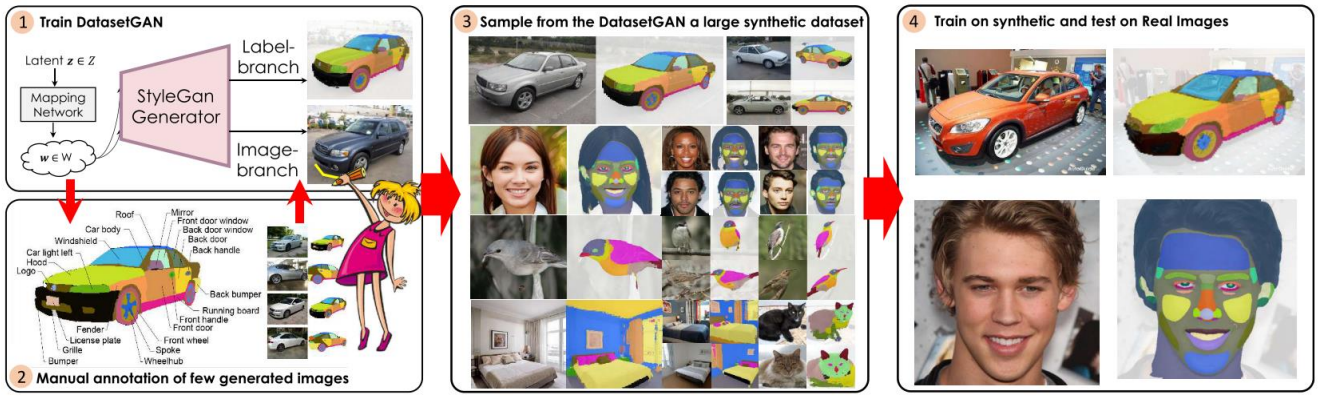


Figure 1: **DATASETGAN** synthesizes image-annotation pairs, and can produce large high-quality datasets with detailed pixel-wise labels. Figure illustrates the 4 steps. *(1 & 2).* Leverage StyleGAN and annotate only a handful of synthesized images. Train a highly effective branch to generate labels. *(3).* Generate a huge synthetic dataset of annotated images authomatically. *(4).* Train your favorite approach with the synthetic dataset and test on real images.

Outline

- Basics in GAN
- GAN for Face
- GAN for Human
- GAN for Font
- GAN as A Tool
- Conclusion

Conclusion

- GAN has been widely used in a lot for mobile applications
  - Face, Human, Font, …
  - A large potential and opportunities for the future
- ID vs Attribute
  - How to disentangle and how to fusion
- Challenges in the future
  - Generation quality： High resolution with details， temporal coherency，robustness to light, pose, …
  - Fine-grained generation
  - Multi-modality fusion