High Fidelity Face Swapping using Generative Adversarial Network

Nguyen Tien Manh

Supervisor: Dr. Do Thai Giang

Outline

1. Introduction
2. Neural network
3. Generative Adversarial Network(GAN) in Face Manipulation
4. High Fidelity Face Swapping
5. Experimental Result
6. Conclusion and Future Work

Introduction

Face swapping is a common method of creating false content that involves replace a target face with a source face while preserving the target's facial attribute and identity information.

Image: Facebook

Introduction

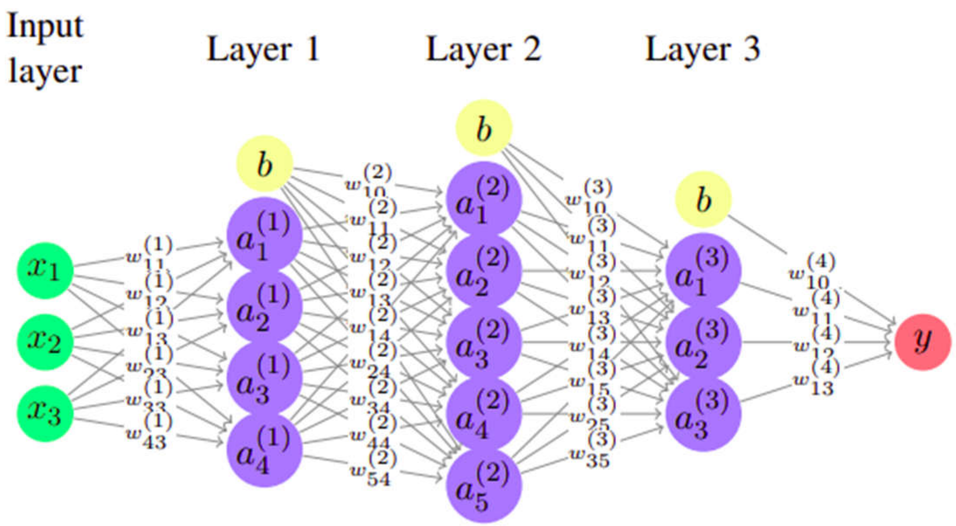
- Recently, GANs (Generative Adversarial Networks) is the driving force behind the progression of face synthesis and manipulation task.

- The aim of this work is to focus on the fidelity of face swapping method. Specifically, to get more perceptually appealing results, we use GANs to synthesized swapped face such that it should be seamlessly blended into the target image with stable quality and follow the target scene’s lighting conditions.

Background Knowledge

Feed Forward Neural Network

We define feedforward neural network as a function approximation that learn a mapping y = (x; θ) to best approximate the desired output.



For each layer, given the n-dimensional input $x = (x_1, x_2, ..., x_n)$, the output computed as:

$$y = f(x) = \sigma(\sum_{i=1}^{n}(w_i \cdot x_i) + b)$$

Background Knowledge

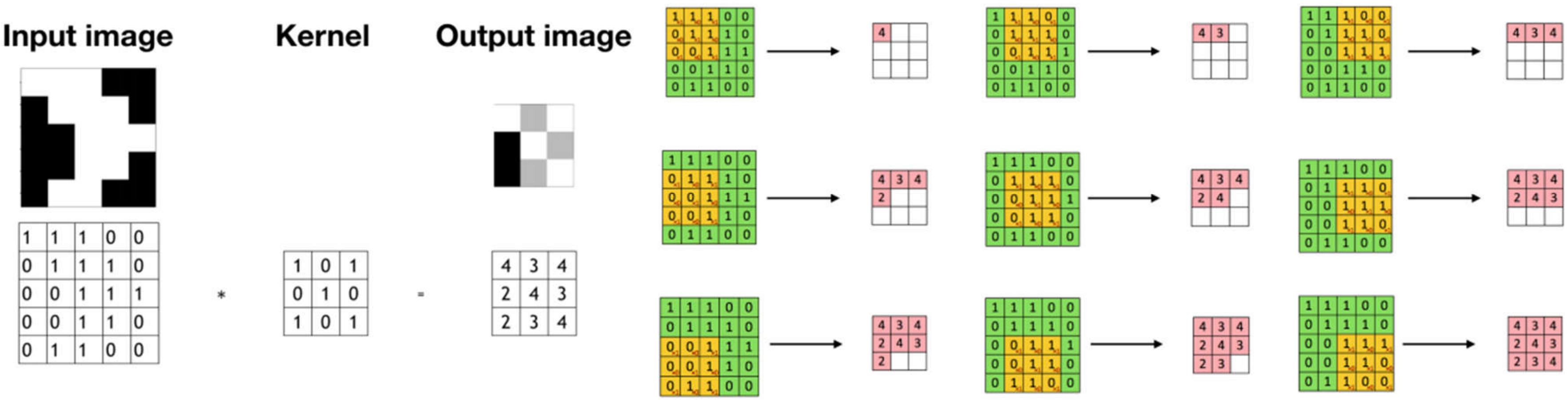
Gradient-based learning

To gradually approach the global optima, the target is optimizing each layer's weight by a factor proportional to the cost C and to the input (x):

$$w' = w^L - \frac{\partial C}{\partial w^L}$$

Background Knowledge

Convolutional Neural Network



Background Knowledge

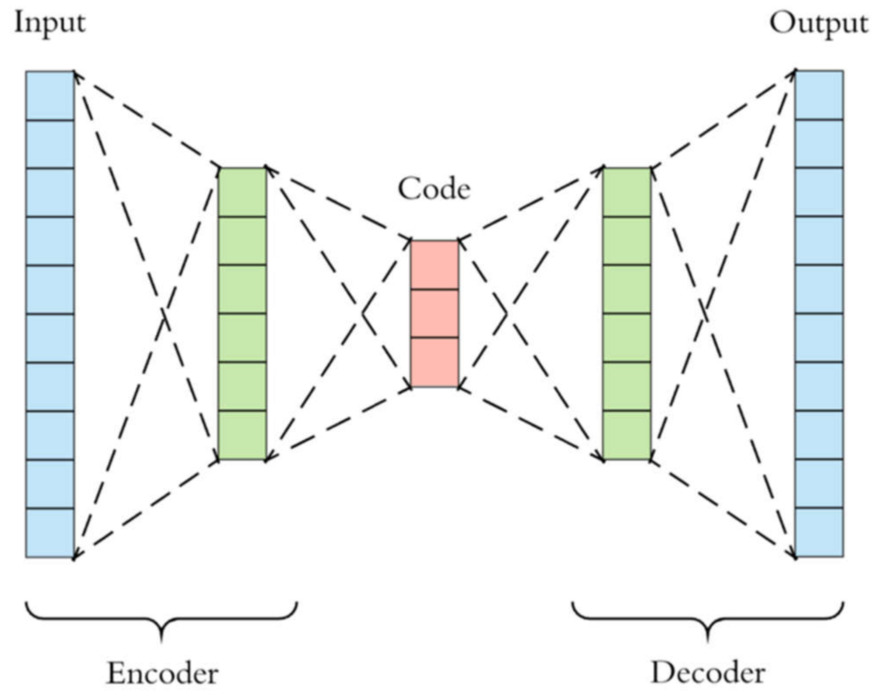
Generative Model

- Discriminative Model
- Generative Model

Encoder-Decoder Network

(ED Network)

$$D(E(x)) = x$$



Background Knowledge

Generative Adversarial Network(GAN)

- Consisted of 2 independent neural network: a generator G and a discriminator D
- The generator’s job is to fool the discriminator such that it can not distinguish generated image and real image.
- D and G play the following two-player minimax game with value function V (G, D):

$$\min_G \max_D V(G, D) = E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(x)))]$$

Background Knowledge

Evolution of generated images from GAN.

Figure taken from [1]

GAN for Face Manipulation

Face Manipulation:

- Face Synthesis
- Face Swapping(*)
- Face Editing
- Face Reenactment

GAN for Face Manipulation

Batch Normalization

$$BN(x) = \gamma(\frac{x - \mu(x)}{\sigma(x)}) + \beta$$

$$\mu_c(x) = \frac{1}{NHW} \sum_{n=1}^{N} \sum_{h=1}^{H} \sum_{w=1}^{W} x_{nchw}$$

$$\sigma_c(x) = \sqrt{\frac{1}{NHW} \sum_{n=1}^{N} \sum_{h=1}^{H} \sum_{w=1}^{W} (x_{nchw} - \mu_c(x))^2 + \epsilon}$$

GAN for Face Manipulation

Instance Normalization

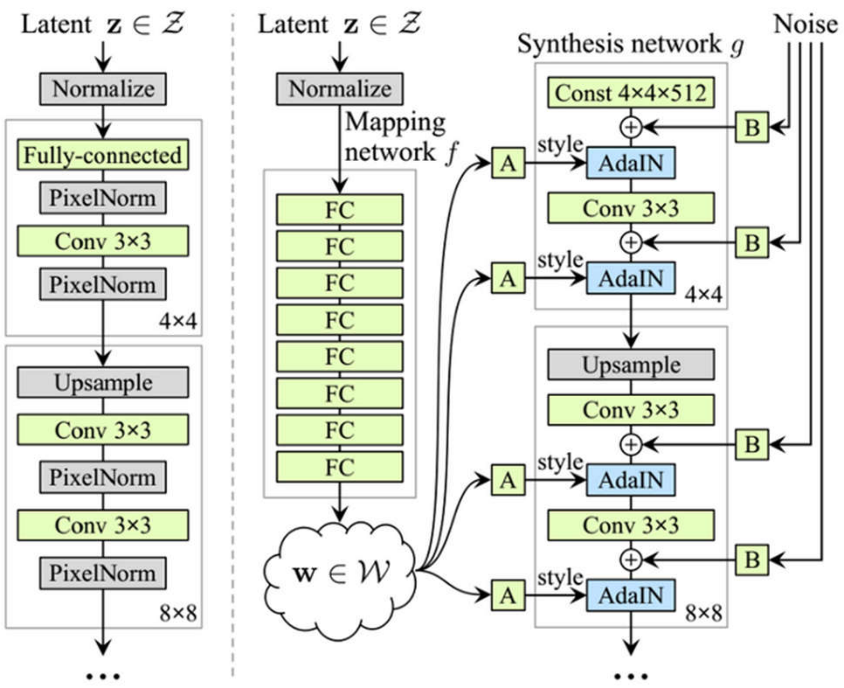
$$IN(x) = \gamma(\frac{x - \mu(x)}{\sigma(x)}) + \beta$$

Adaptive Instance Normalization

$$AdaIn(x, y) = \sigma(y)(\frac{x - \mu(x)}{\sigma(x)}) + \mu(y)$$

GAN for Face Manipulation

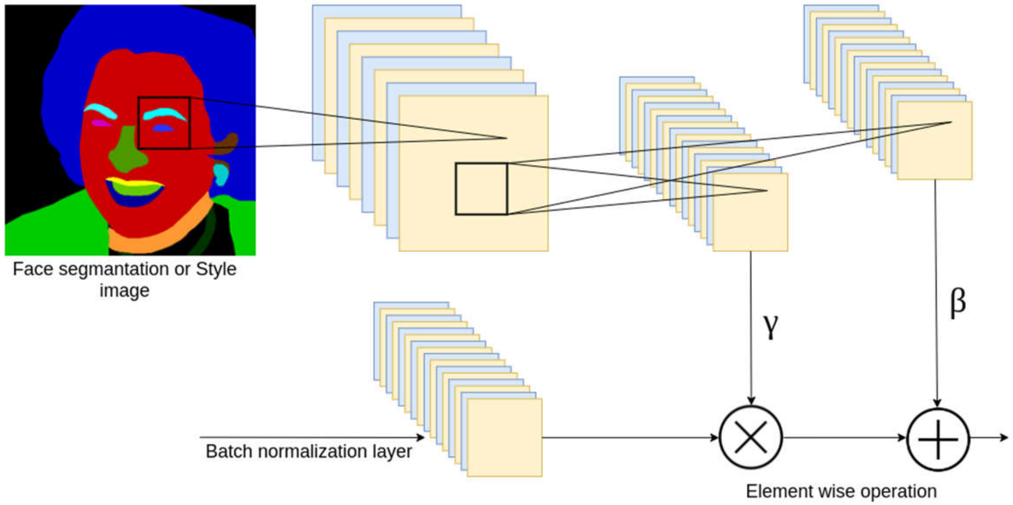
StyleGAN



(a) Traditional (b) Style-based generator

GAN for Face Manipulation

Spatial Adaptive Denormalization (SPADE)

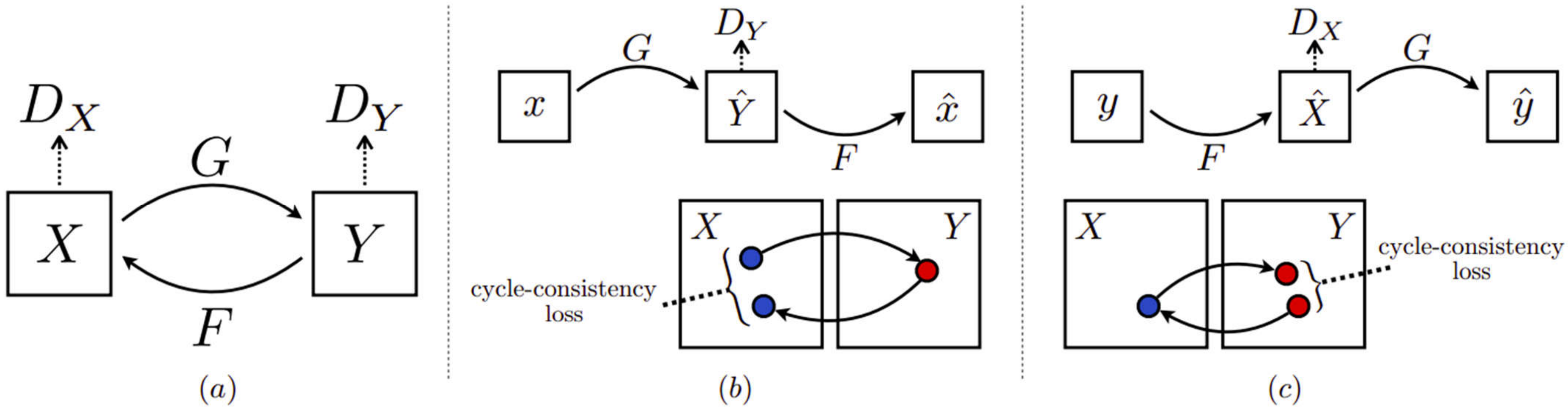


$$SPADE(x, y) = \gamma^i_{c,y,x}(m)\frac{h^i_{n,c,y,x} - \mu^i_c}{\sigma^i_c} + \beta^i_{c,y,x}(m)$$

GAN for Face Manipulation

CycleGAN

Image-to-image translation



High Fidelity Face Swapping: Related Works

The output face image must be matched with pose, attribute,... with target face image and have realistic look features, that is indistinguishable from the real face.

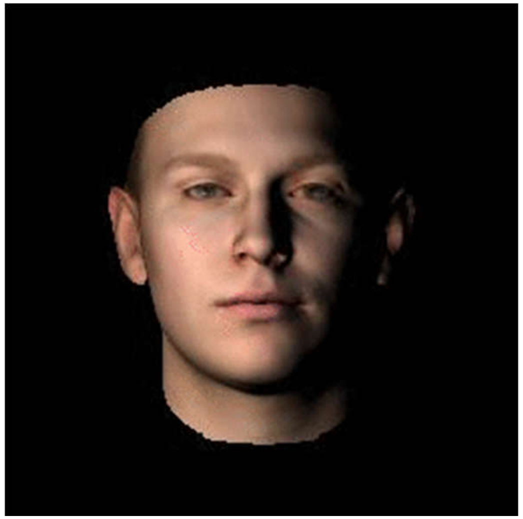
Two approaches

- 3D based approach
- GAN based approach

High Fidelity Face Swapping: Related Works

3D based approach

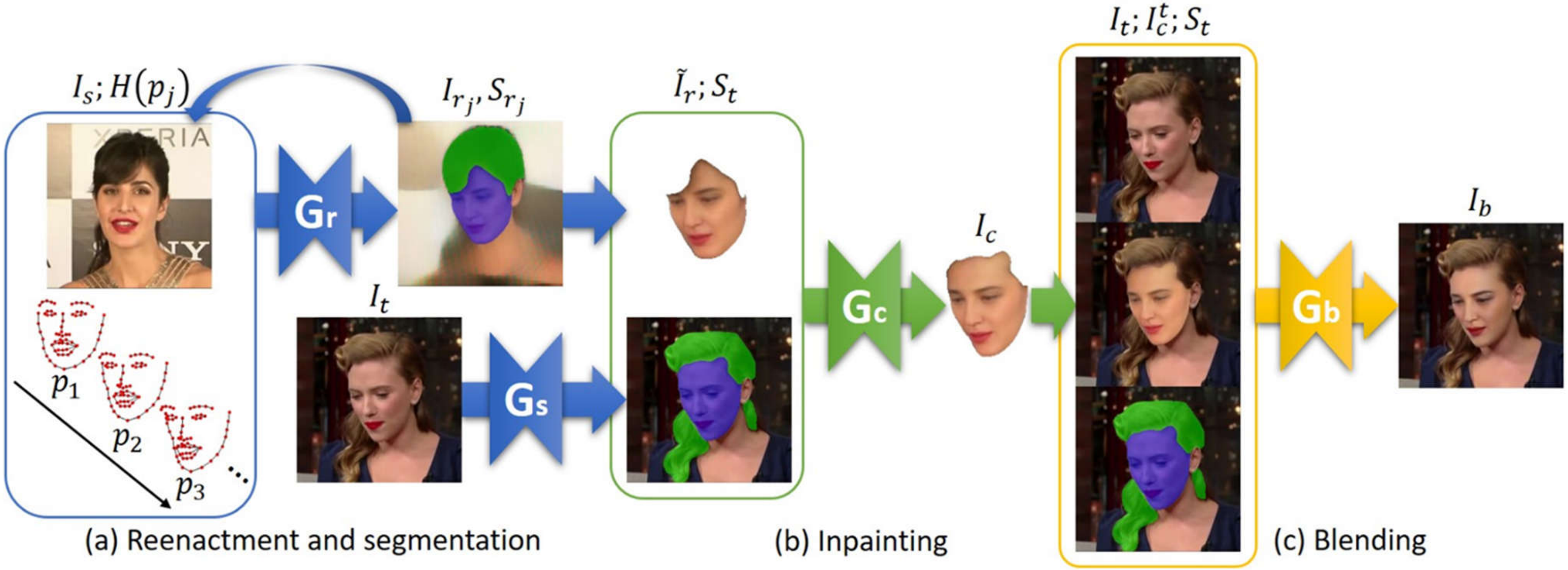
- Utilize 3DMM to approximate the two face's shape, expression and then make the transition in 3D space to smoothing the variation [2,3]
- Require specific 3D data
- Result seem not plausible



3D morphable model (3DMM)

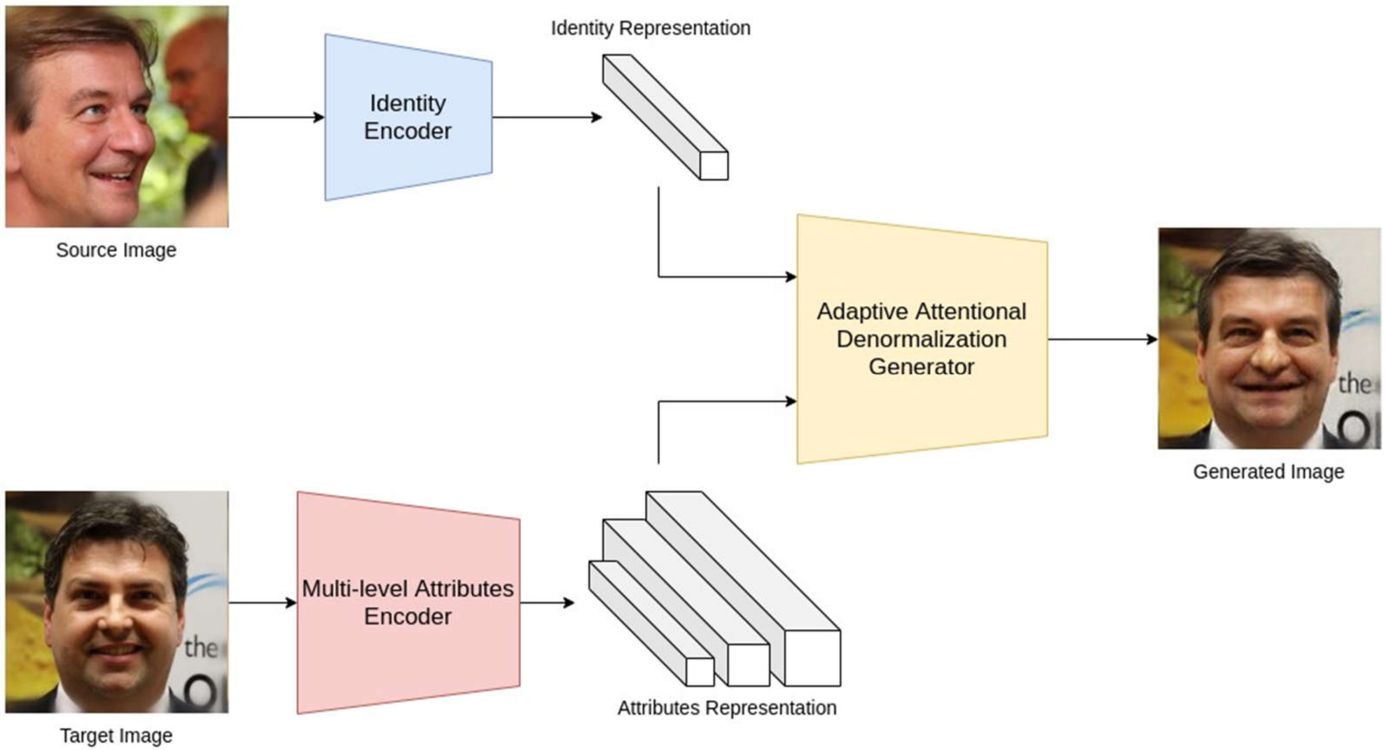
High Fidelity Face Swapping: Related Works

GAN based approach: RSGAN, FSGAN



FSGAN architecture

High Fidelity Face Swapping: FaceShifter



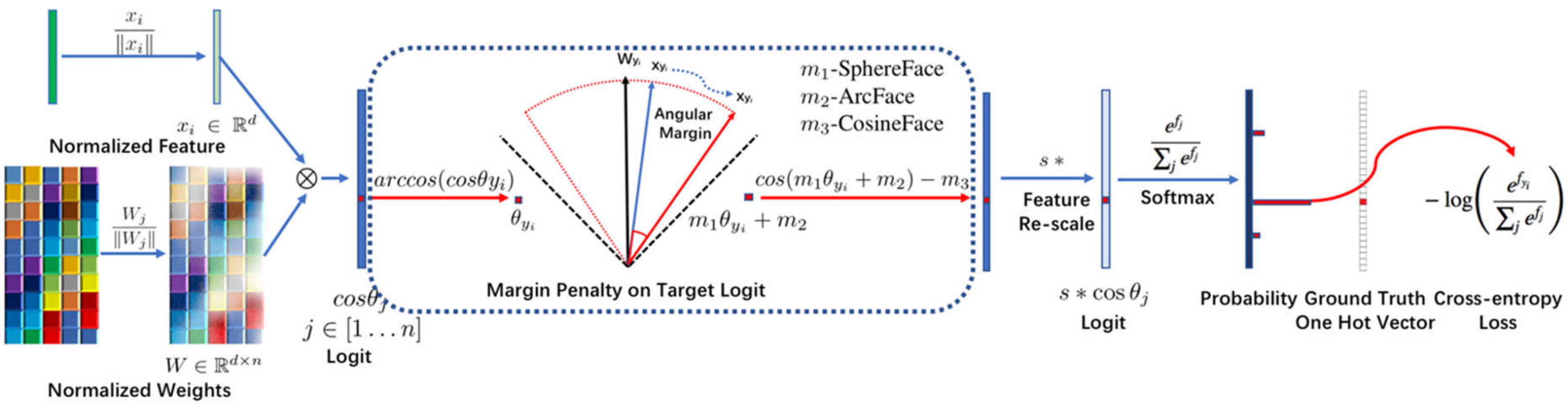
Overall architecture of the framework

High Fidelity Face Swapping: Face Shifter

Identity Encoder

Additive Angular Margin Loss

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{\exp(s\cos(\theta_{y_i}+m))}{\exp(s\cos(\theta_{y_i}+m))+\sum_{j=1,j\neq y_i}^{n}\exp(s\cos\theta_j)}$$



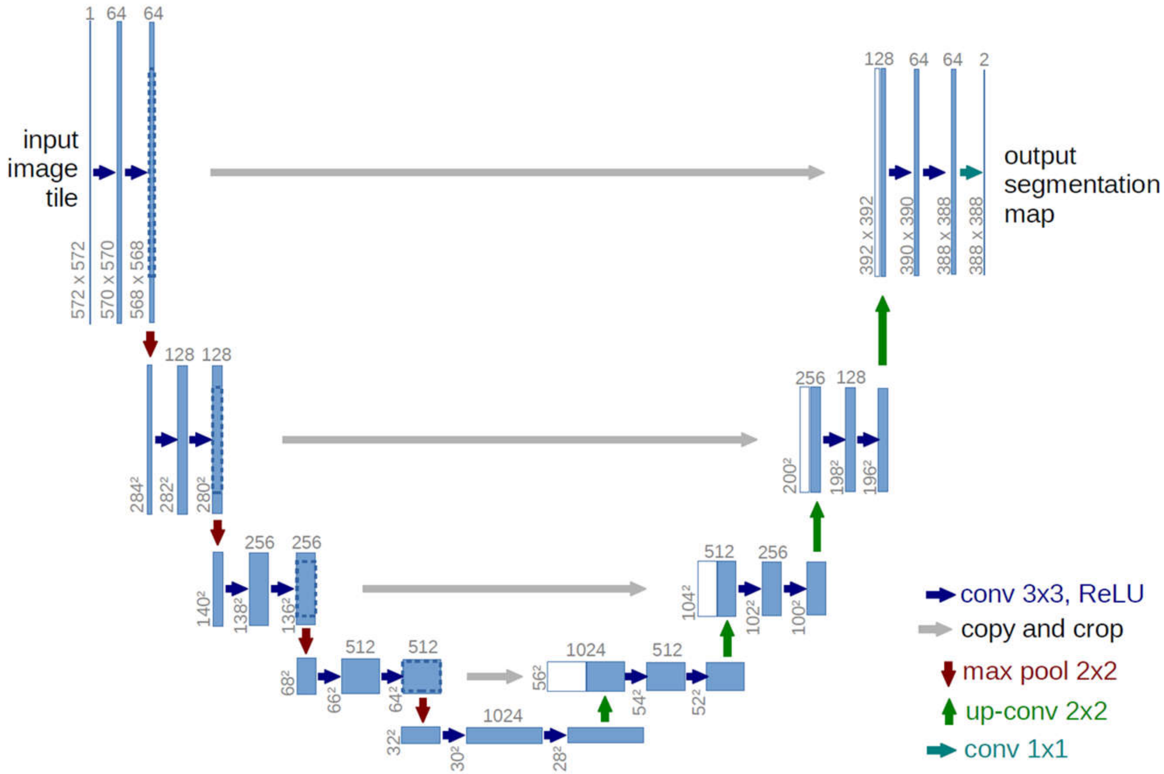
High Fidelity Face Swapping: Face Shifter

Multi-level attributes encoder

Specifically, we design a U-Net like structure that take input a image and output a list of n feature embeddings, where $z_{att}(X_t)$ corresponding to the k-level attribute feature map from the UNet decoder:

$$z_{att}(X_t) = \{z^1_{att}(X_t), z^2_{att}(X_t), ..., z^n_{att}(X_t)\}$$

High Fidelity Face Swapping: Face Shifter



Multi-level attributes encoder

High Fidelity Face Swapping: Face Shifter

Follow SPADE, the input first was normalized:

$$\bar{h^k} = \frac{h_{in}^k - \mu^k}{\sigma^k}$$

$$\mu_c^k(h_{in}) = \frac{1}{NHW} \sum_{n=1}^{N} \sum_{h=1}^{H} \sum_{w=1}^{W} h_{nchw}$$

$$\sigma_c^k(h_{in}) = \sqrt{\frac{1}{NHW} \sum_{n=1}^{N} \sum_{h=1}^{H} \sum_{w=1}^{W} (h_{nchw} - \mu_c^k(h_{in}))^2}$$

High Fidelity Face Swapping: Face Shifter

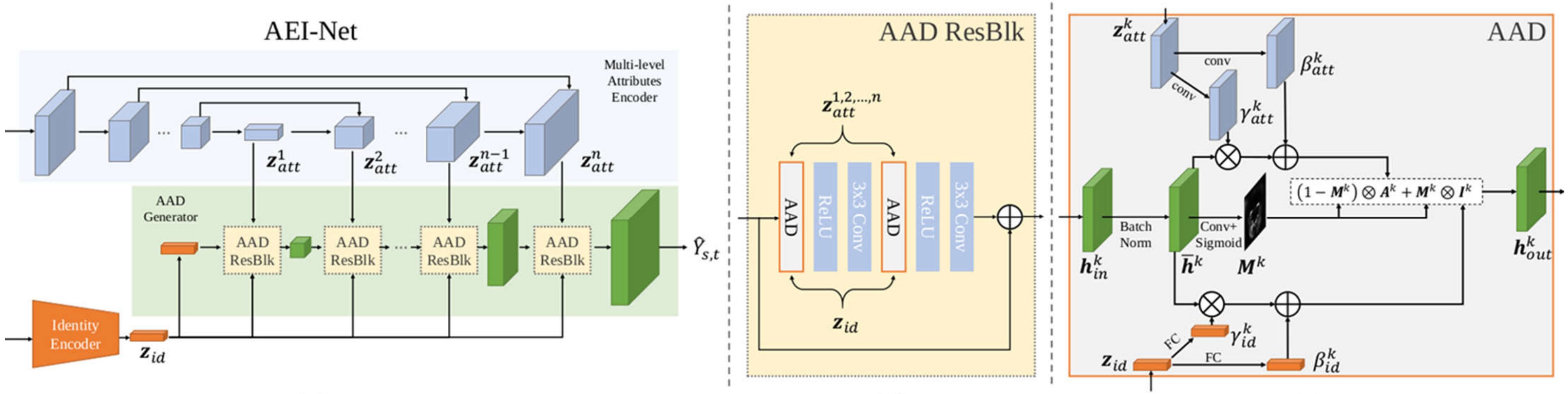
The input then go through the Adaptive Attentional Denormalization Layer, which compute the modulation parameters by utilizing a neural network

$$I^k = \gamma_{id}^k \otimes \bar{h^k} + \beta_{id}^k,$$
$$A^k = \gamma_{att}^k \otimes \bar{h^k} + \beta_{att}^k$$

$$h_{out}^k = (1 - M^k) \otimes A^k + M^k \otimes I^k$$

High Fidelity Face Swapping: Face Shifter

Adaptive Embedding Denormalization Network



(a) (b) (c)

High Fidelity Face Swapping: Face Shifter

Training Objective

$$\mathcal{L}_{adv}(\hat{Y}_{s,t}) = \max(0, (1 - Y_t) \cdot \hat{Y}_{s,t})$$

$$\mathcal{L}_{id} = 1 - \cos\left(z_{id}(\hat{Y}_{s,t}), z_{id}(X_s)\right)$$

$$\mathcal{L}_{att} = \frac{1}{2} \sum_{k=1}^{n} \left\| z_{att}^{k}(\hat{Y}_{s,t}) - z_{att}^{k}(X_t) \right\|_2^2$$

$$\mathcal{L}_{rec} = \begin{cases} \frac{1}{2} \left\| \hat{Y}_{s,t} - X_t \right\|_2^2 & \text{if } X_s = X_t \\ 0 & \text{otherwise} \end{cases}$$

$$\mathcal{L} = \mathcal{L}_{adv} + \lambda_{id}\mathcal{L}_{id} + \lambda_{att}\mathcal{L}_{att} + \lambda_{rec}\mathcal{L}_{rec}$$

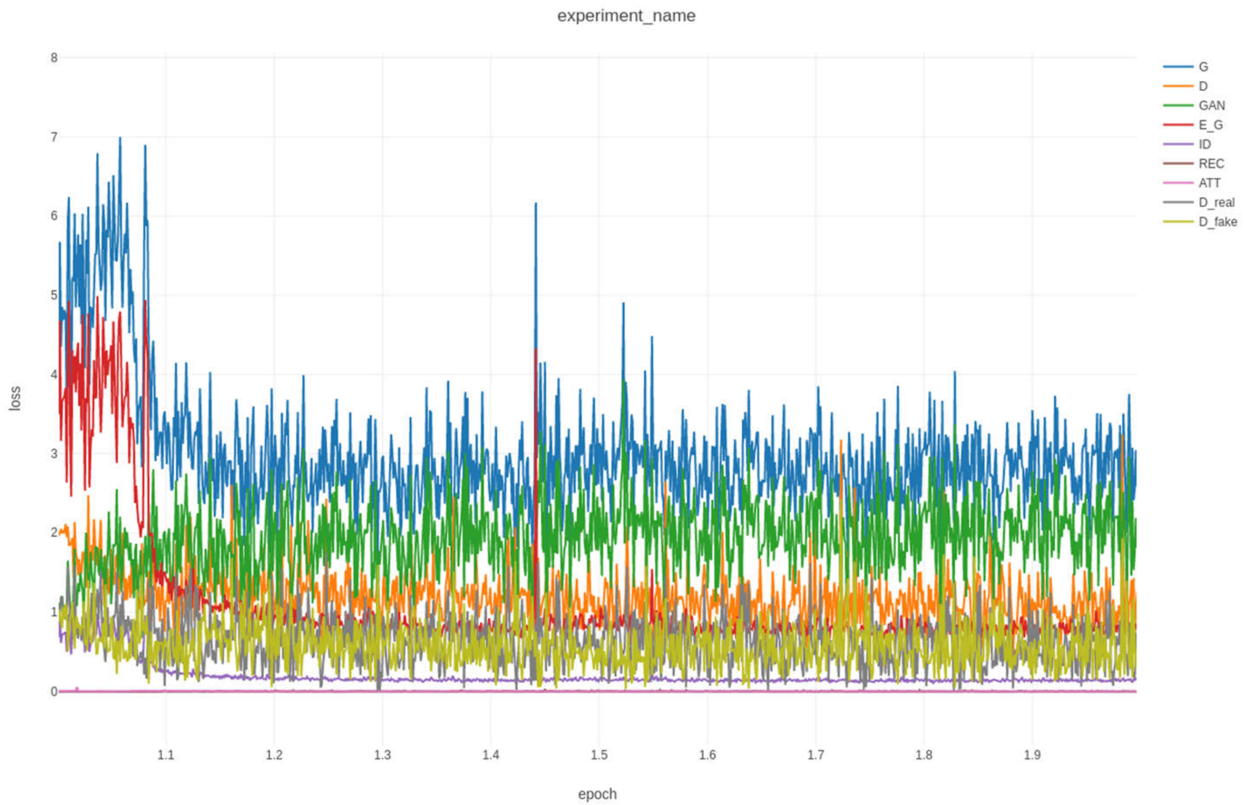
Experimental Result

Dataset

- VGGFace
- FFHQ
- CelebA-HQ

FFHQ Dataset

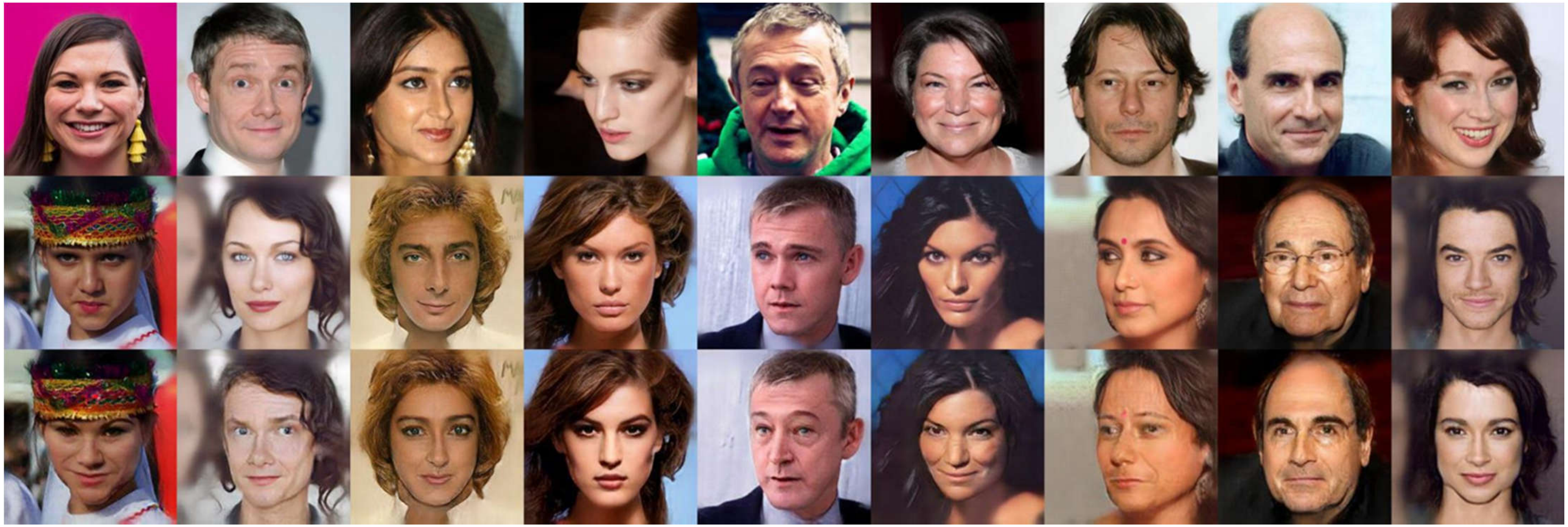
Experimental Result



Training loss in 500K steps

Experimental Result

Result on multiple datasets



Experimental Result

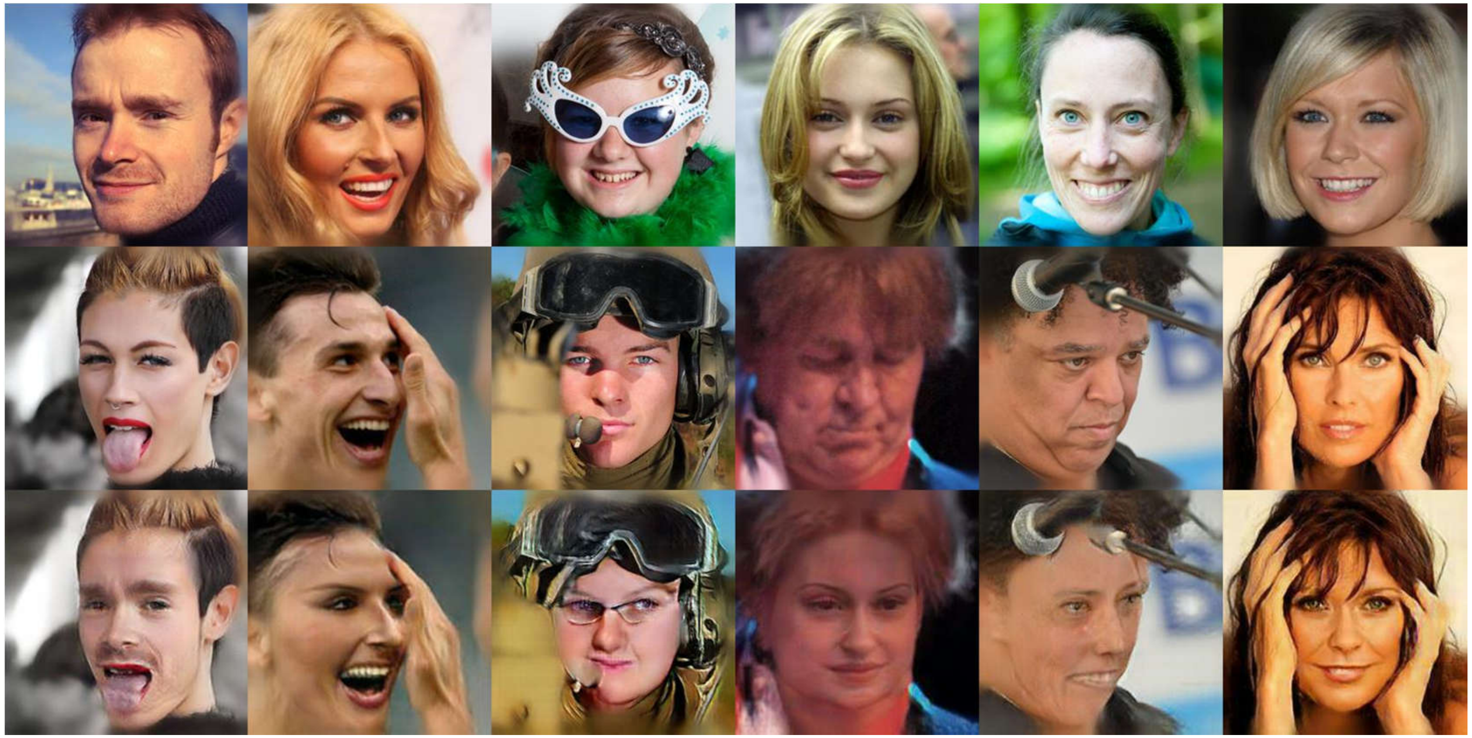
Compare with other state-of-the-art framework

Conclusion

- We studied a novel framework for high fidelity face swapping task by using generative adversarial network.
- Without subject-specific annotations, our works is able to surpass other approaches in producing accurate high fidelity facial images by providing just two face images.
- Extensive experiments show that our framework significantly outperforms current state-of-the-art face swapping methods
- We also publicize the code used for training and testing the model, which we hope will considerably contribute to further research works and related open-source projects.

Future work

Limitations when using our frameworks: Occlusion, extreme pose, GAN artifacts



Q&A session

Thank you for listening!

Reference

[1] Miles Brundage, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe, Paul Scharre, Thomas Zeitzoff, Bobby Filar, Hyrum S. Anderson, Heather Roff, Gregory C. Allen, Jacob Steinhardt, Carrick Flynn, Seán Ó hÉigeartaigh, Simon Beard, Haydn Belfield, Sebastian Farquhar, Clare Lyle, Rebecca Crootof, Owain Evans, Michael Page, Joanna Bryson, Roman Yampolskiy, and Dario Amodei. The malicious use of artificial intelligence: Forecasting, prevention, and mitigation

[2] Hyeongwoo Kim, Pablo Garrido, Ayush Tewari, Weipeng Xu, Justus Thies, Matthias Niessner, Patrick Pérez, Christian Richardt, Michael Zollhöfer, and Christian Theobalt. Deep video portraits. ACM Trans. Graph., 37(4), July 2018.

[3] Justus Thies, Michael ZollhÃ¶fer, Marc Stamminger, Christian Theobalt, and Matthias NieÃner. Face2face: Real-time face capture and reenactment of rgb videos, 2020

Yuval Nirkin, Yosi Keller, and Tal Hassner. FSGAN: subject agnostic face swapping and reenactment. CoRR, abs/1908.05932, 2019.

Lingzhi Li, Jianmin Bao, Hao Yang, Dong Chen, and Fang Wen. Faceshifter: Towards high fidelity and occlusion aware face swapping. CoRR, abs/1912.13457, 2019