Deep Face Swap with GAN

(https://youtu.be/bvopA4qkH-w)

Chi Wang, Jinil Jang

{chiwang, jinil}@stanford.edu

Introduction

Human image synthetic technology plays an important role in movie editing. In nowadays, AI based approach, such as "DeepFake", is getting attention, but it has constrained with factors such as light condition or skin tones. We overcome the limitation with GAN.

Dataset

- Collected 500 celebrity – "George Clooney" – images from Google
- 200 our own images (200/Chi or 200/Jinil) for model training
- Randomly split 500 celebrity's pictures to 400 / 50 / 50 as train, validation and test sets.
- In pre-train, detected faces are extracted from images, warped, and scaled to 64 x 64 x 3.

Methods

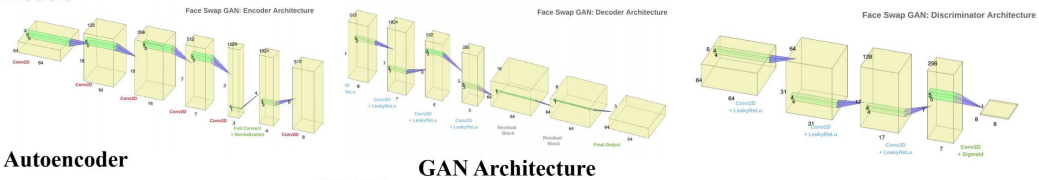
- MSE loss for the Discriminator
- SSIM loss for the Autoencoder

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

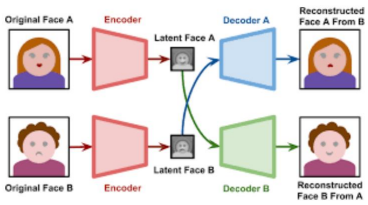
- ADAM with $learning\ rate = 5e^{-5}, \beta_1 = 0.5, \beta_2 = 0.999$

Approach

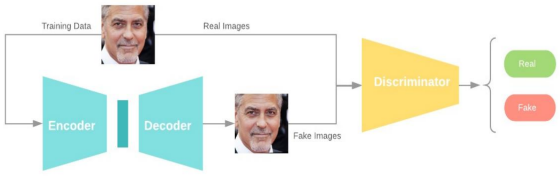
Models



Autoencoder



GAN Architecture

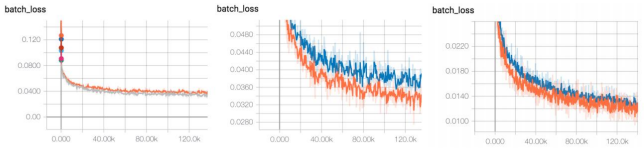


Training

Performed on a GPU-enabled EC2 instance per each algorithm (Baseline, More Intermediate layers, and GAN) in parallel over 72 hours.

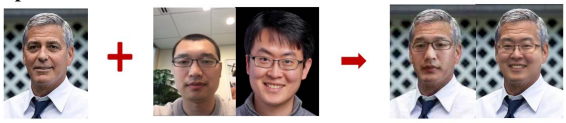
Results

Training Loss



Baseline | More Intermediate layers | GAN

Output



Quality Comparison

Baseline

More Intermediate layers

GAN

Discussion

- To get reasonable output, it's required at least 500 dataset, and over 50 hour training process on GPU-enabled machine. Long training process made us hard to tune hyper-parameters.
- If there is multiple faces in an image, all faces were swapped with source face. Face recognition was needed to swap only the target face.
- After 40K epoch, learning was slow and gave blur output. Early stopping and applying GAN on output result can reduce train process.

Future Directions

Bad output example:

- Face edge is not smooth with bear.
- Black/white image gets blur result.

Future Work:

- Data Augmentation: Generate Black/White images to improve model performance on dark environment.
- GAN on the swapped image to make edge smooth.

References

[1] Bitouk et all, Face swapping: automatically replacing faces I photographs, 2008
[2] Ganin et all, Deepwarp: Photorealistic image resynthesis for gaze manipulation, 2016
[3] Kim et all, Deep video portrait, 2018
[4] Lin et all, Face replacement with large-pose differences, 2012
[5] Radford et all, Unsupervised representation learning with deep convolutional generative adversarial networks, 2015
[6] Smith et all, Joint face alignment with non-parametric shape models, 2012