Inverting Facial Embeddings with GANs

Edward Vendrow
evendrow@stanford.edu
Akash Singhal
akash13@stanford.edu

Introduction

- We seek to recreate an image of a face from only its “identity”, a 128-dimensional vector generated using FaceNet
- We take the identity of a person’s face as an input and find a latent vector generating a face with a similar identity.
- **Given just an identity, our algorithm generates images that fool FaceNet, a state-of-the art facial recognition system.**

Data

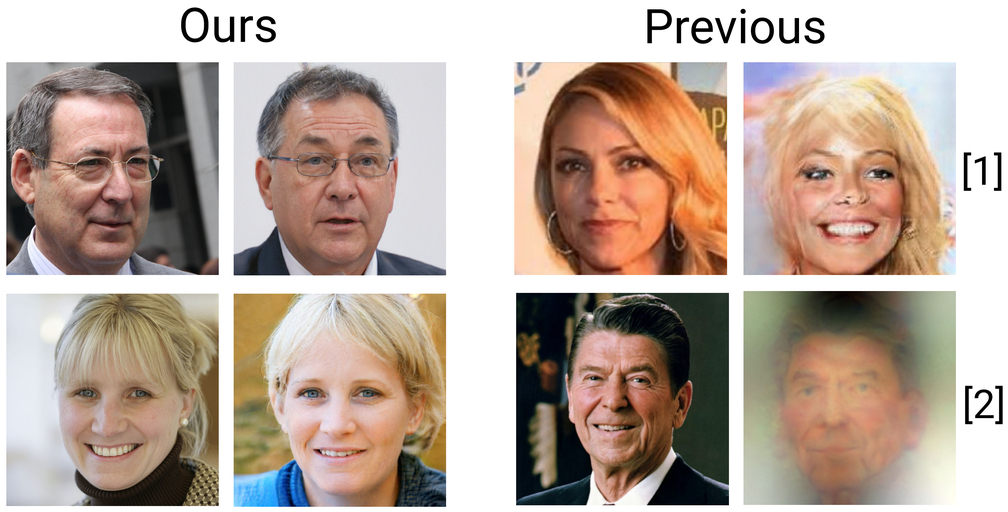
Examples from FFHQ (training StyleGAN):

We preprocessed data by running face detection to crop and resize images.

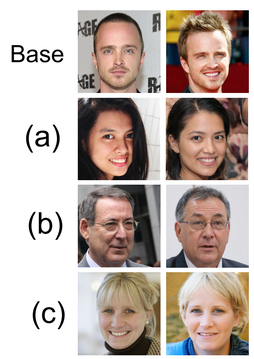
Method

- We use greedy search algorithm boosted by a large, pre-generated dataset
- We iteratively narrow down our search by repeatedly adding noise to the previous best latent to find a better one, judged by the FaceNet loss on its image.
- Decrease noise variance over time

Results



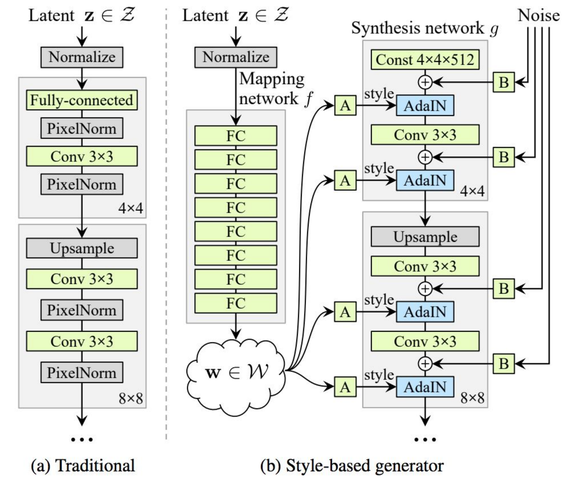
Our method presents a significant improvement over previous efforts to recreate faces from embeddings. By basing our face generation on NVIDIA’s state-of-the-art StyleGAN, we are able to create images that are both convincingly real and similar in identity to the original face.



Trial	FaceNet Loss
Baseline	6.112
(a)	6.736
(b)	6.929
(c)	**6.153**

FaceNet loss is defined as the norm of the difference in identity vectors. Loss below 6.5 generally signifies identical people. Different people usually have losses above 14.

Architecture



Architecture used for traditional models vs architecture used by the StyleGAN model [3].

Future Work

- Look for better search algorithms that can perform the search for the desired latent vector more efficiently.
- Apply re-sampling and then perform gradient descent to explore other approaches to recovering the desired latent vector.
- Plenty of other potential avenues as this is a relatively unexplored research area.

References

[1] Li, Zhigang et. al. “Generate Identity-Preserving Faces by Generative Adversarial Networks.” ArXiv abs/1706.03227 (2017): 1-9

[2] Zhmoginov, Andrey and Mark Sandler. “Inverting face embeddings with convolutional neural networks.” ArXiv abs/1606.04189 (2016): 4

[3] Karras, Tero et al. “A Style-Based Generator Architecture for Generative Adversarial Networks.” CVPR (2018).

[4] NVlabs. “NVlabs/Stylegan.” GitHub, 3 Dec. 2019, https://github.com/NVlabs/stylegan.