FSGAN: Subject Agnostic Face Swapping & Reenactment

—

Team members:
Neppalli Balakrishna Prasad, Srilekha Kadambala

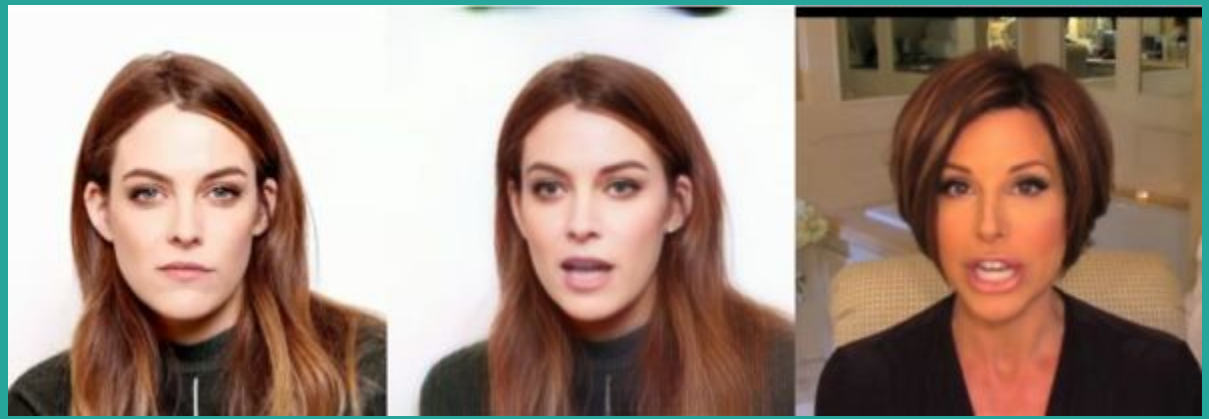
Introduction

- **Face swapping** is the task of transferring a face from source to target image, so that it seamlessly replaces a face appearing in the target and produces a realistic result.

Source **Face Swapping** **Target**

Introduction

- **Face reenactment** uses the facial movements and expression deformations of a control face in one video to guide the motions and deformations of a face appearing in a video or image



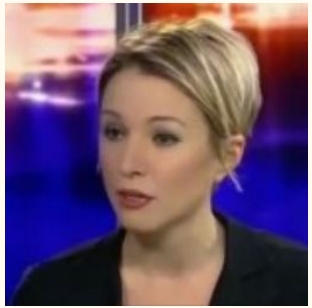
Source **Face Reenactment** **Target**

Objective

We derive a GAN based approach for face reenactment which adjusts for both pose and expression variations and can be applied to a single image or a video sequence. For video sequences, we introduce continuous interpolation of the face views. Occluded face regions are handled by a face completion network. Finally, we use a face blending network for seamless blending of the two faces while preserving target skin color and lighting conditions.

Source

Target



FSGAN

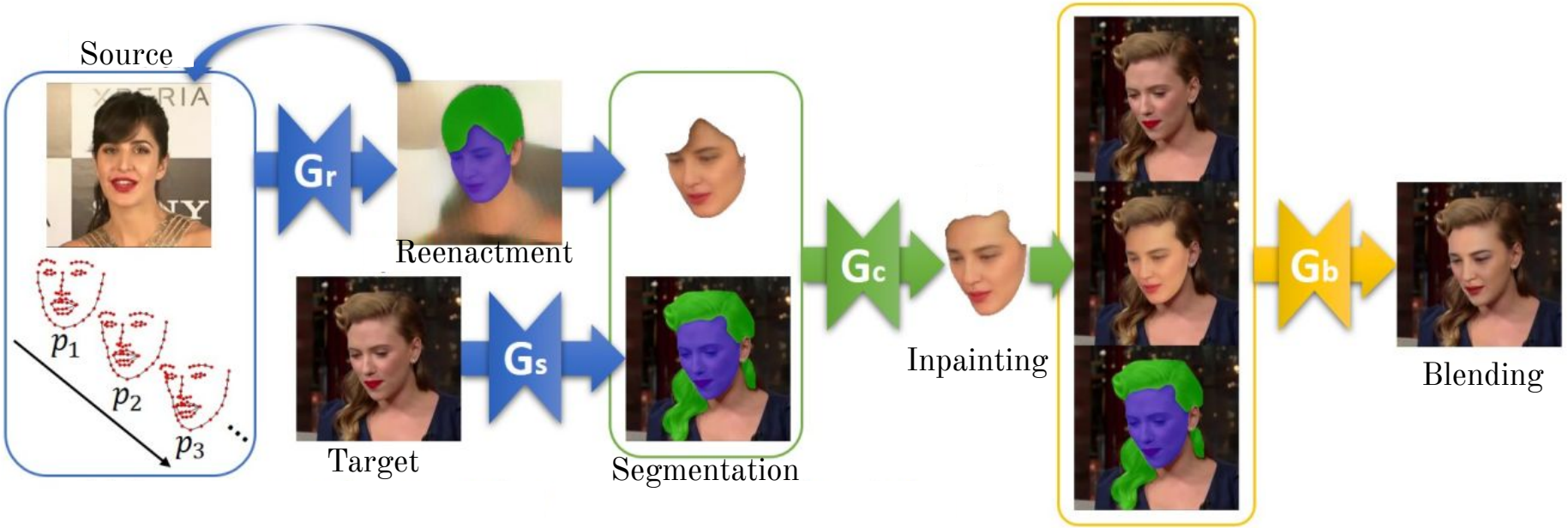
Dataset and Training details

- IJB-C dataset (~11k face videos, of which 5,500 were used which were in HD)
- VGGFace2 dataset (3.3M images depicting 9,131 identities)
- CelebA (202,599 images, annotated with 40 binary attributes)
- LFW Parts Labels set (~3k images) and Figaro dataset (1k images) for face and hair segmentations
- FaceForensics++ (provides 1000 videos, from which they generated 1000 synthetic videos on random pairs using DeepFakes and Face2Face)
- Proposed generators were trained from scratch, where the weights were initialized randomly using a normal distribution.
- Adam optimization was used.

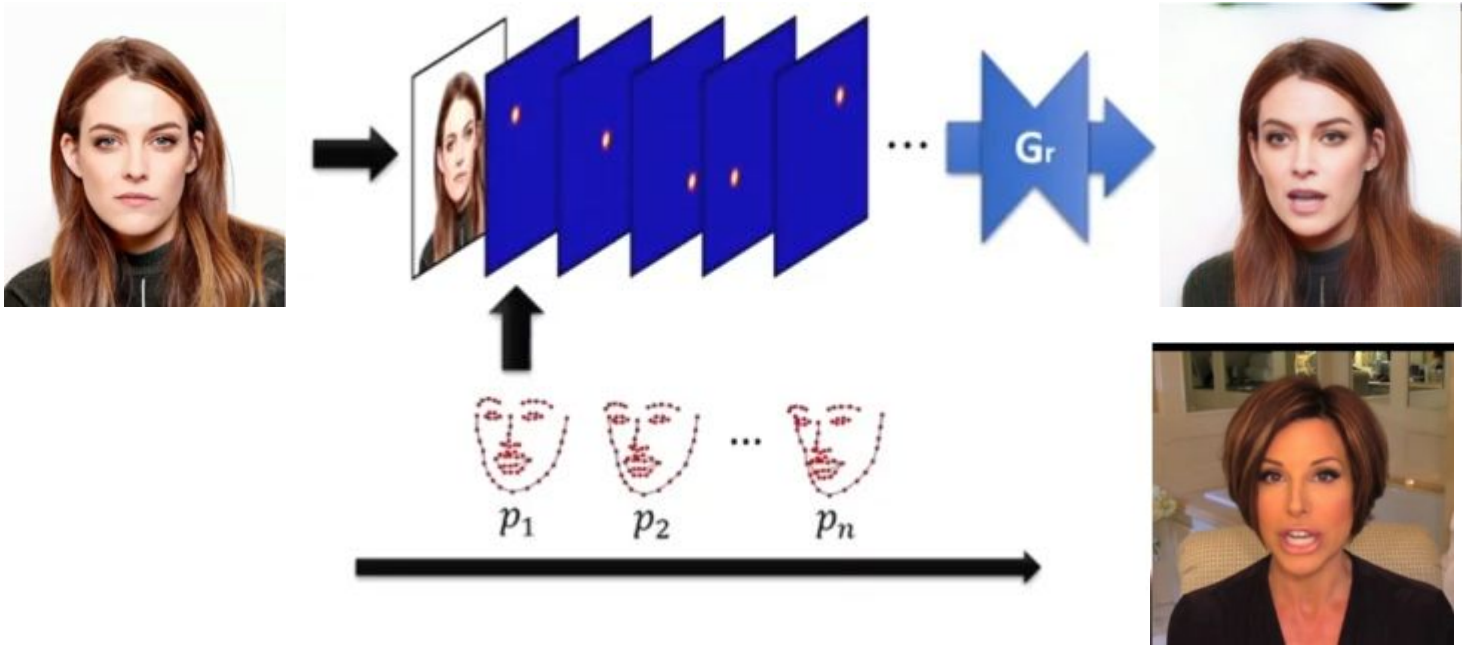
Methods Available

1. Face2Face
2. Nirkin et al.
3. DeepFakes
4. FSGAN (Face Swapping GAN)

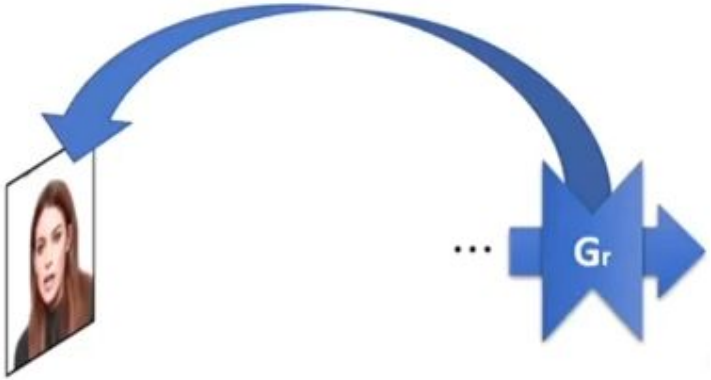
Model Overview

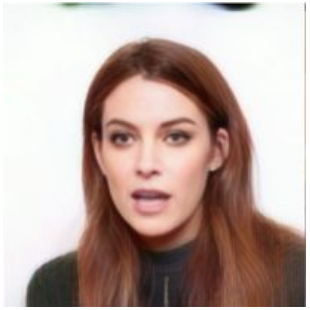


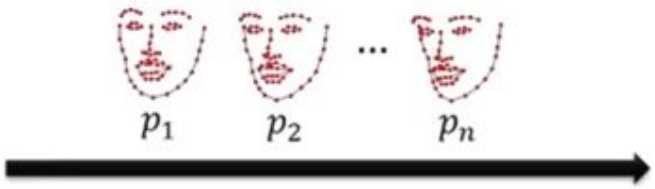
Face Reenactment



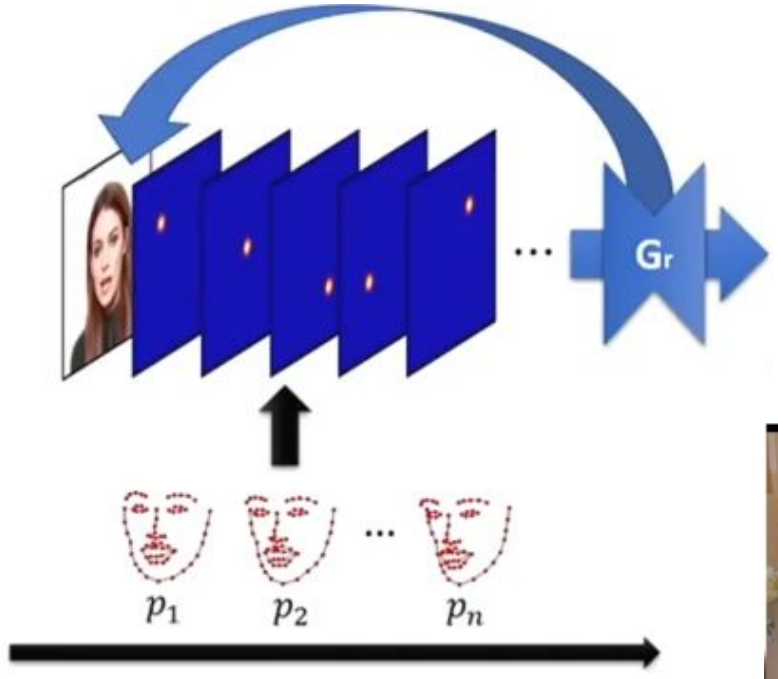
Face Reenactment

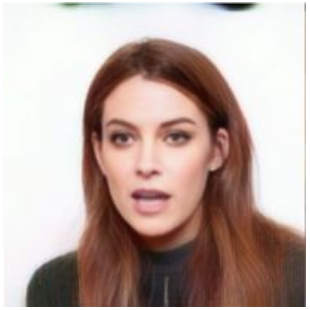






Face Reenactment





Face Inpainting

Reenactment

Completion

Target Face

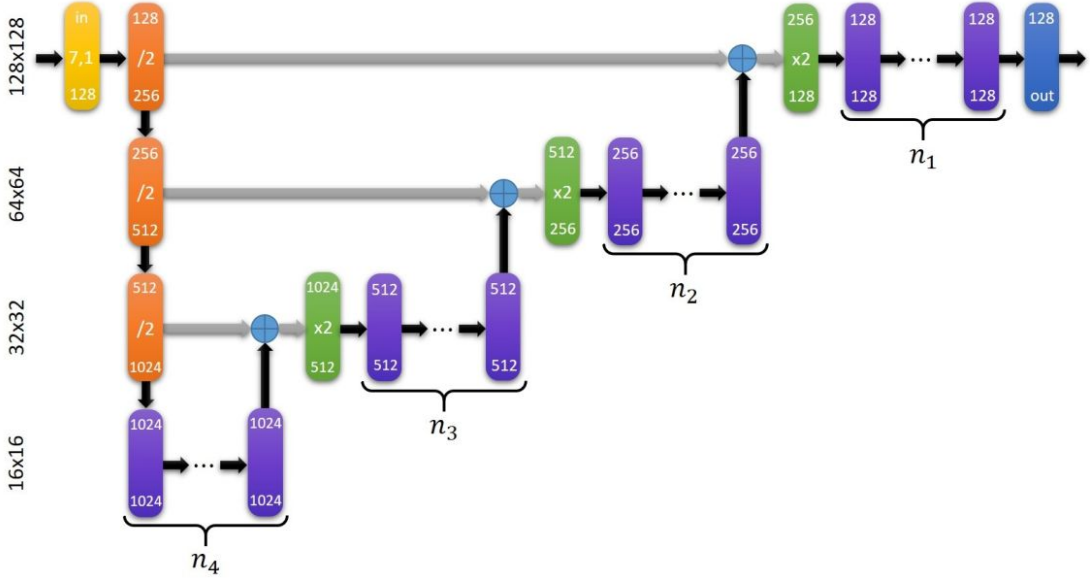
Architecture

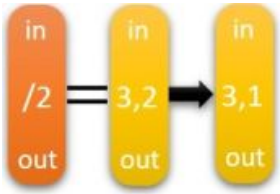
- U-Net
- pix2pixHD

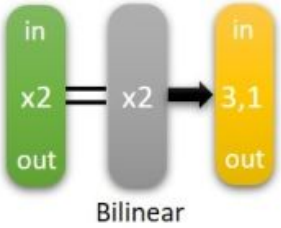
$G_r = G_c = \text{Enhancer}(\text{Global}(2,2,3),2)$

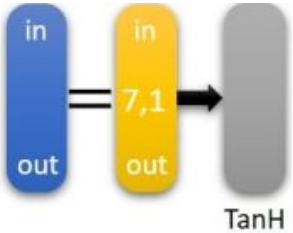
$G_b = \text{Enhancer}(\text{Global}(1,1,1),1)$

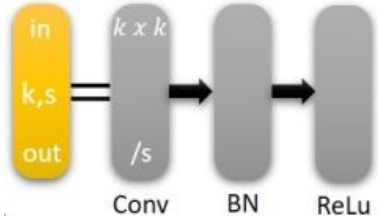
Figure showing $\text{Global}(n_1,n_2,n_3,n_4)$

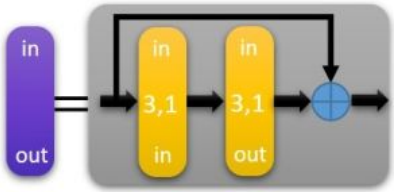






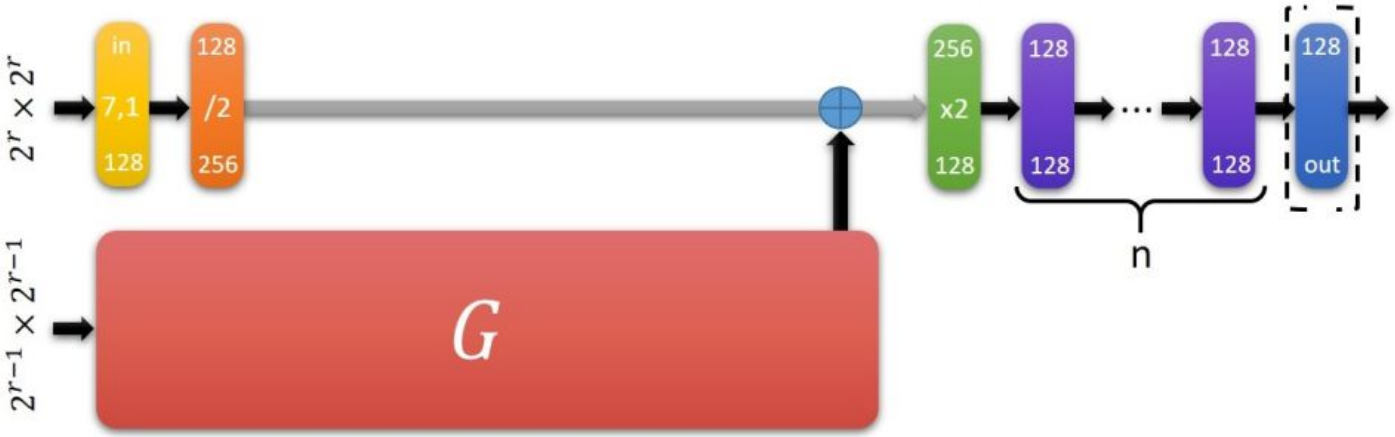




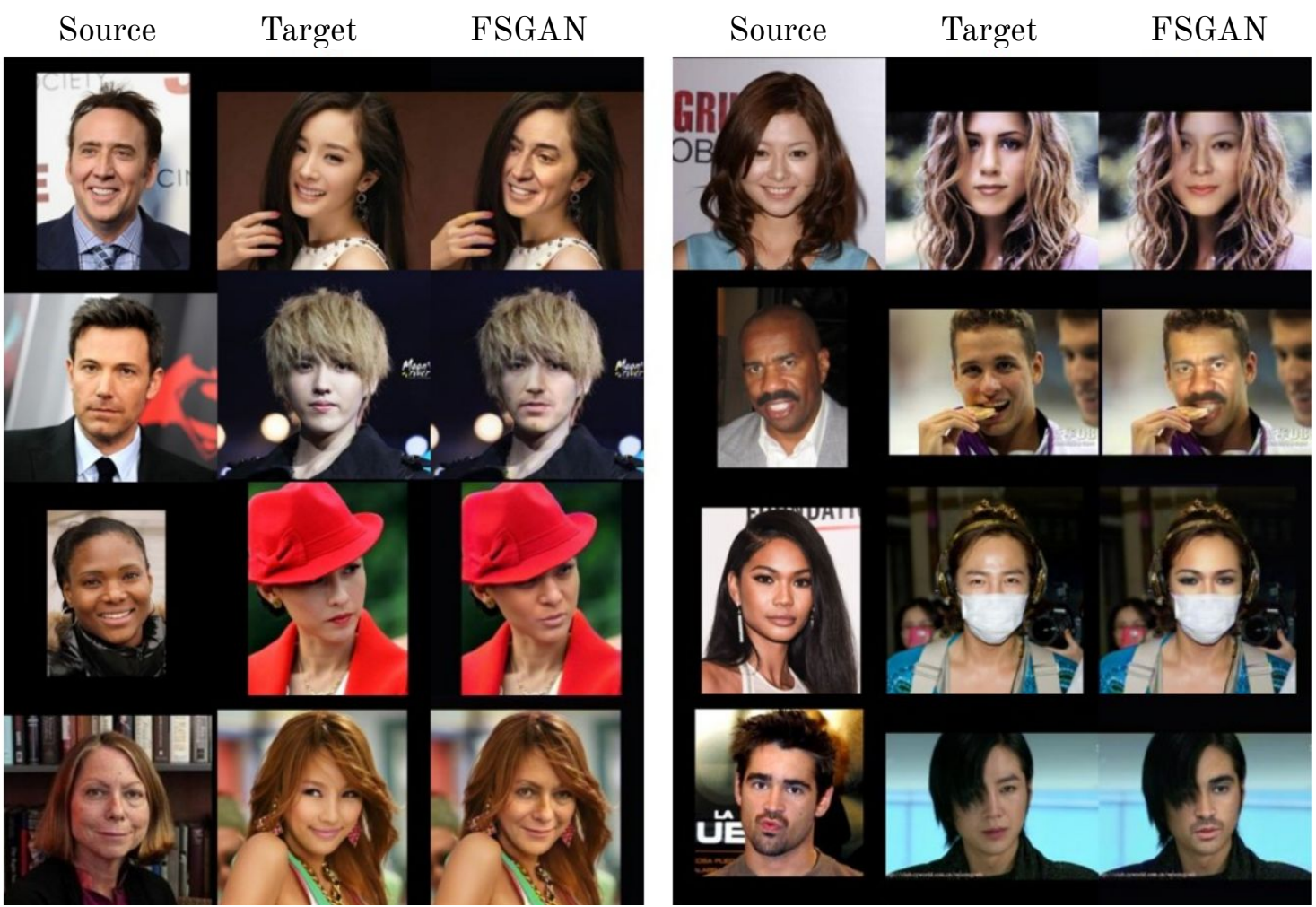


Architecture

Figure showing Enhancer(G,n)



Results



Limitations

- Larger the angular differences, the more texture and identity quality degrades.
- Too many iterations may lead to blurring the texture of the image.
- Limited to the resolution of the training data.

Conclusion

We can validate how well the methods preserve the source subject identity, while retaining the same pose and expression of the target subject.

The FSGAN model has outperformed many other models without even training subject-specific images. GANs are successful in generating fake faces from realistic images.

This method has various applications in various domains such as in entertainment for visual media production, graphics etc.

THANKYOU