FSGAN: Subject Agnostic Face Swapping and Reenactment

Authors: Yuval Nirkin, Yosi Keller, Tal Hassner

**Paper Review: SUBHANI SHAIK(MT18117)**

**I. SUMMARY**

Face swapping is a task of transferring a face from source to target image. Face Re-enactment means involving facial movements and expressions based on appearance of source image to target image using transformations. FSGAN (Face Swapping GAN) is an end-to-end trainable architecture for swapping and for re-enactment of the face from source to target image. There are many other approaches which worked on the same problem like 3D face representations for face swapping, DeepFake architecture, latent feature based architectures, disentanglement and GANs which generates face with respect to subject. The main advantage of this architecture is- it is subject agnostic i.e., we can transfer to face along with re-enactment without training that subject.

The FSGAN has three main components- Face Reenactment and Segmentation, Face Inpainting and Face Blending.

To get better results they used Perceptual Loss Lperc, Pixelwise loss (L1 loss) Lpixel, Reconstruction loss Lrec which is combination of Lperc and Lpixel, and Adversarial loss Ladv which is GAN loss.

**A. Face Reenactment and Segmentation:**

In this, Gr (re-enactment generator) which takes an image *I* and a heatmap *H(p)* which has facial landmarks. Gr generates the re-enacted image Ir and segmented image of hair and face Sr with the help of GS (Segmentation CNN- UNet), which further used this intermediate generated image to generate enhanced image with face which might also have missing pixels. GS is used to analyse and generate the segmented hair and face image St of target.

Here for training authors have used Stepwise consistency loss for Gr and cross-entropy loss for GS. This network was trained Gr and GS alternatively. To get better interpolation of face for given subject images {Is1..sn}, Euler angle {e1..n}and Faces{Fs1..sn}we find the closest triangular plane position of target image and calculate barycentric coordinates and generates Sr.

By end of this component we will have Ir, Sr, St which are further send to next component.

**B. Face Inpainting:**

Generally re-enacting the occluded faces is difficult and more chance to generate artifacts. To deal with that problem authors have proposed this Inpainting generator Gc. It takes Ir, Sr and It and render the better context with respect to target image by randomly removing elliptical shape parts like hair etc.

It used generator loss *L*(*Gc*) = *λrecLrec*(*Ic; It*) + *λadvLadv.*

We will get Fc image as output which is further send to next component.

**C. Face Blending:**

This component take Fc and blends with respect to Ft. To take skin-tone and lighting conditions into account authors have used Poisson blending loss. For generating better re-enacted face authors have combined this poisson blending loss with perceptual Loss Lperc.

*L*(*Gb*) = *λrecLrec*(*Gb*(*It*; *Irt*; *St*)*; P*(*It*; *Irt*; *St*))+*λadvLadv.*

**II. LIMITATIONS OF FSGAN**

The FSGAN works better than earlier approaches like DeepFake etc. and have advantage like subject agnostic architecture, but it has some limitations:

* For large images with occlusions, at the time of inpainting it may generate artifacts which can result in blurriness and degrade the quality of re-enacted image.
* Because of iterative process it may generate degraded texture of the image.
* It wraps the texture directly from the target images i.e., more dependent on attributes of training data like resolution etc.
* The authors have used facial landmarks for re-enactment with can be sparse, because of that architecture may not capture facial expressions correctly.

**III. SUGGESTIONS FOR THE PROPOSED TECHNIQUE**

* For segmentation authors have used UNet with bilinear interpolation for upsampling. We can use segmentation architectures like FastFCN, Gated-SCNN which give better results. So that it will improve quality of re-enactment.
* We can use occlusion removal architecture like 3DMM which will perform better images by wrapping texture from the images itself.
* For better facial land marking we can use RNN with attention architecture in re-enactment component. So that it can generate better image Ir depicting that features of It.

IV. REFERENCES

1. FSGAN: Subject Agnostic Face Swapping and Reenactment, Yuval Nirkin et.al.
2. FastFCN: Rethinking Dilated Convolution in the Backbone for Semantic Segmentation, Huikai Wu et.al.
3. Gated-SCNN: Gated Shape CNNs for Semantic Segmentation, Towaki Takikawa et.al.
4. Face De-occlusion using 3D Morphable Model and Generative Adversarial Network, Xiaowei Yuan et.al.