FIVA: Facial Image and Video Anonymization and Anonymization Defense - Supplementary Materials

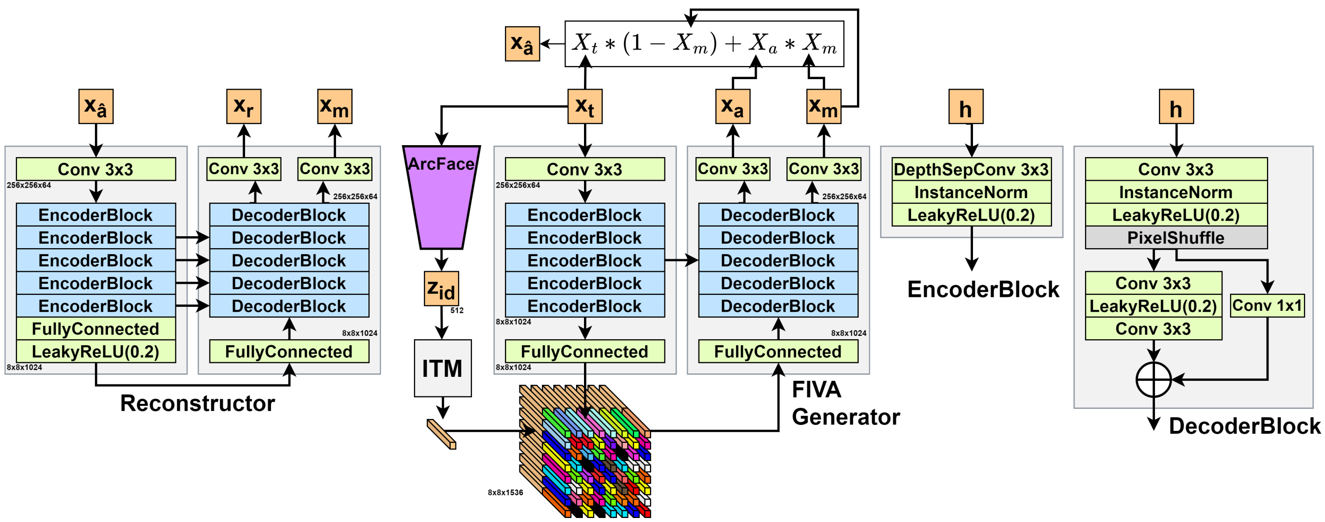

Figure 1: Overview of the reconstruction attack model and the *FIVA* generator. The skip connections are concatenated with the upsampled feature maps from the previous DecoderBlock before being passed into the next DecoderBlock.

A . Fake Identity Sampling Details

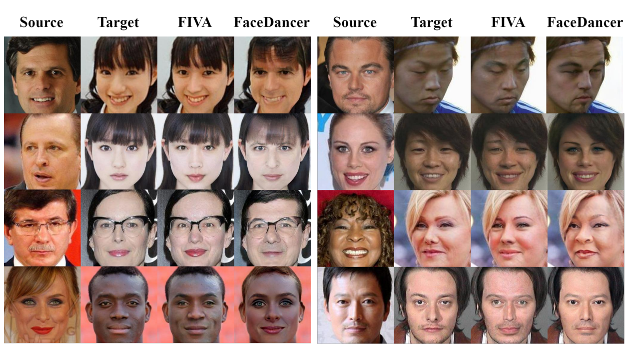

Figure 2: Qualitative comparison between *FIVA* and FaceDancer for gender and ethnicity retention.

In Figure 3 we visualize in a 3d projection how the anchor set of mixed identities is constructed. In reality, the dimensionality of these points are 512 (embedding size of the ArcFace output). In this example we only show three points for clarity. This is mentioned in the main manuscript as well, but using following equation:

$$z_a = \mathcal{S}_a[argmin(|cos(z_{id}, \mathcal{S}_a| + m))] \ , \qquad (1)$$

we can control a desired approximate cosine distance from the target identity by changing the margin m. In Figure 4 we illustrate anchor matches as the margin changes, where the green line represent the match that would occur for the margin m of 0.7. Because *FIVA* was trained counter-factually to drive the identity away we want to find an anchor close. When using target-oriented face swapping methods, which are trained to drive the identity towards the source, you have to sample identities far away.

B . Gender and Ethnicity Preservation

Because *FIVA* is trained counter-factually to drive the identity away it only uses target faces, compared to face swapping which generally uses pairs of target and source faces. It makes sense that the generator learns to preserve gender and ethnicity to be able to generate convincing samples for the discriminator. We find interesting evidence for this when looking into face swapping using *FIVA*.

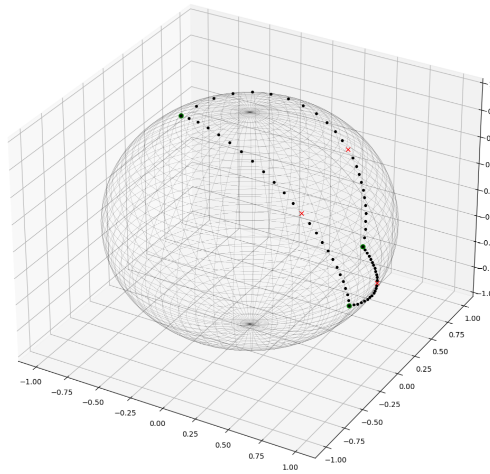

Figure 3: Illustration of how we create the anchor vectors that we sample fake identities from. The green points illustrates an identity extracted by ArcFace, the black dots shows the spherical interpolation path on the unit sphere and the red crosses represent the resulting vectors used in the anchor set which becomes a mix of actual identities.

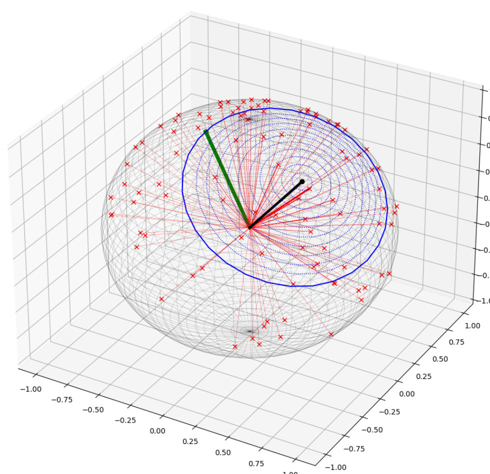

Figure 4: Illustration of matching a desired anchor. The red lines illustrates matches to a desired anchor based on desired approximate distance from the target vector (black line). The green line illustrates the match that would occur for when sampling for *FIVA*. Blue circle illustrates the desired distance.

FIVA manage to reach state-of-the-art for identity transfer for face-swapping. However, we notice qualitatively that the gender and ethnicity it preserved, even if you would use a male target and female source or Asian target with a Caucasian source. Shown in Figure 2, we can see that *FIVA* tends to preserve gender and ethnicity for face swaps compared to FaceDancer. Even if *FIVA* performs better for identity transfer quantitatively, it does not qualitative. However this behaviour is useful for other tasks, such as anonymization, allowing us to ignore the need to sample identities based of gender.