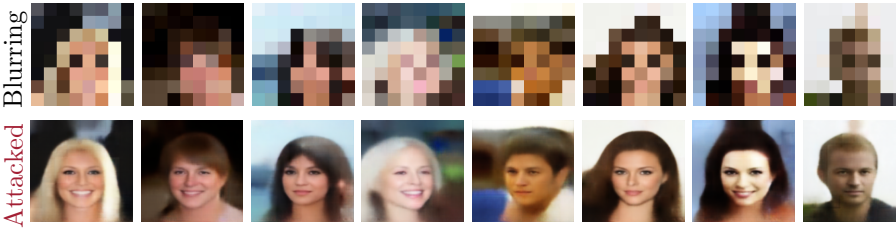
1 Additional Results of Face Reconstruction

In this section, we provide additional visualized results for defense and attack on the two datasets, CelebA and LFWA, to prove that our FaceMix consistently outperforms other defending methods, including adding noise and blurring.

(a) Raw images of human faces from the test set of the CelebA dataset.

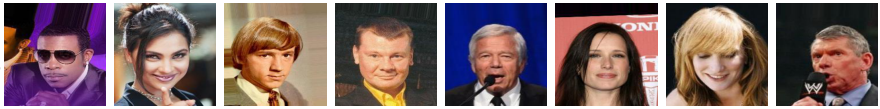
(b) Adding Noise $\mathcal{N}(0, 8)$. The attack model can recover the faces with many details and recognizable personal identification.



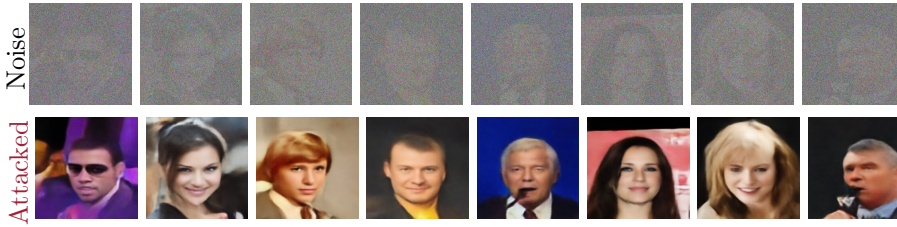
(c) Blurring (8×8). Artifacts appears in the reconstructed faces but many sensitive face attributes still retains, e.g. race and hair color.

(d) FaceMix (Ours). The attack model fails to separate the information from the mixed faces and the privacy is well preserved.

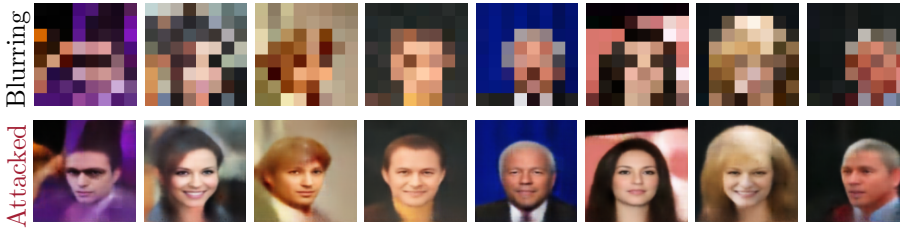
Fig. 1: Defense and attack on th CelebA dataset. It is much harder for the attack model to reconstruct the raw data from our mixed input sent to the cloud than adding noise and blurring. For each method, 1st row is the mixed images applied the defense method on the raw image; 2nd row is the image recovered by the GAN-based attack model.



(a) Raw images of human faces from the test set of the CelebA dataset.



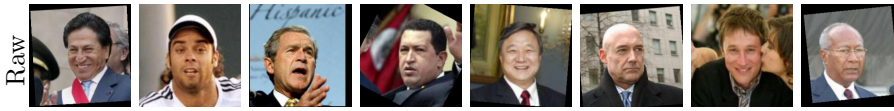
(b) Adding Noise $\mathcal{N}(0, 8)$. The attack model can recover the faces with many details and recognizable personal identification.



(c) Blurring (8×8). Artifacts appears in the reconstructed faces but many sensitive face attributes still retains, e.g. race and hair color.

(d) FaceMix (Ours). The attack model fails to separate the information from the mixed faces and the privacy is well preserved.

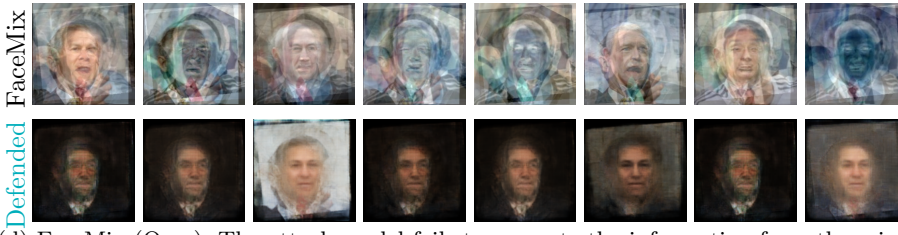
Fig. 2: Defense and attack on th CelebA dataset. It is much harder for the attack model to reconstruct the raw data from our mixed input sent to the cloud than adding noise and blurring. For each method, 1st row is the mixed images applied the defense method on the raw image; 2nd row is the image recovered by the GAN-based attack model.



(a) Raw images of human faces from the test set of the LFWA dataset.

(b) Adding Noise $\mathcal{N}(0, 8)$. The attack model can recover the faces with many details and recognizable personal identification.

(c) Blurring (16×16). Artifacts appears in the reconstructed faces but many sensitive face attributes still retains, e.g. race and baldness.

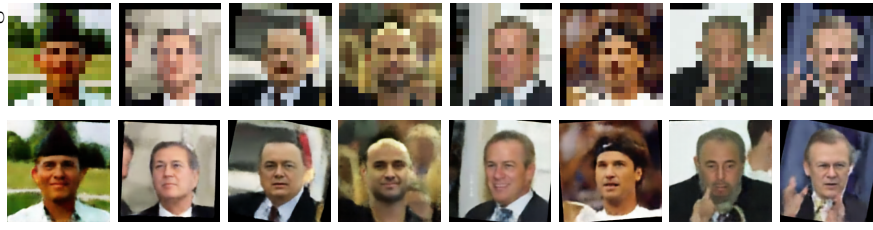


(d) FaceMix (Ours). The attack model fails to separate the information from the mixed faces and the privacy is well preserved.

Fig. 3: Defense and attack on the LFWA dataset. It is much harder for the attack model to reconstruct the raw data from our mixed input sent to the cloud than adding noise and blurring. For each method, 1st row is the mixed images applied the defense method on the raw image; 2nd row is the image recovered by the GAN-based attack model.

(a) Raw images of human faces from the test set of the LFWA dataset.

(b) Adding Noise $\mathcal{N}(0, 8)$. The attack model can recover the faces with many details and recognizable personal identification.



(c) Blurring (16×16). Artifacts appears in the reconstructed faces but many sensitive face attributes still retains, e.g. race and baldness.

(d) FaceMix (Ours). The attack model fails to separate the information from the mixed faces and the privacy is well preserved.

Fig. 4: Defense and attack on the LFWA dataset. It is much harder for the attack model to reconstruct the raw data from our encrypted input sent to the cloud than adding noise and blurring. For each method, 1$^{\text{st}}$ row is the encrypted images applied the defense method on the raw image; 2$^{\text{nd}}$ row is the image recovered by the GAN-based attack model.