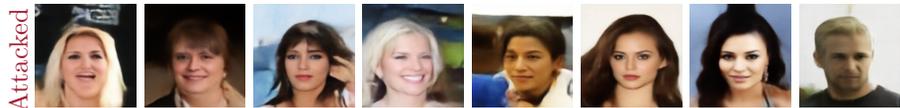
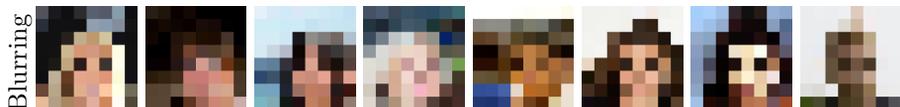


# 1 Additional Results of Face Reconstruction

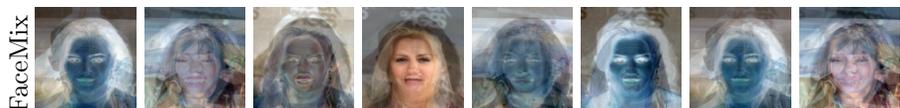
In this section, we provide additional visualized results for defense and attack on the two datasets, CelebA and LFWA, to prove that our FaceMix consistently outperforms other defending methods, including adding noise and blurring.



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



(c) Blurring ( $8 \times 8$ ). 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, 1<sup>st</sup> row is the mixed images applied the defense method on the raw image; 2<sup>nd</sup> 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 ( $8 \times 8$ ). 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.

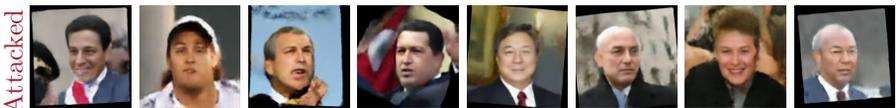
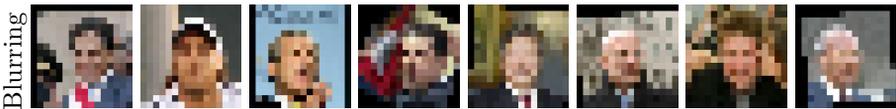
Fig. 2: Defense and attack on th CelebA dataset. It is much harder for the attack model to reconstruct the raw data from our mixed input sent to the cloud than adding noise and blurring. For each method, 1<sup>st</sup> row is the mixed images applied the defense method on the raw image; 2<sup>nd</sup>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



148  
149  
150  
151  
152  
153  
154  
155  
156  
157

Attacked

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



198  
199  
200  
201  
202  
203  
204  
205  
206  
207

Defended

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



248  
249  
250  
251  
252  
253  
254  
255  
256  
257

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

258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314

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, 1<sup>st</sup> row is the mixed images applied the defense method on the raw image; 2<sup>nd</sup> 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 ( $16 \times 16$ ). 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<sup>st</sup> row is the encrypted images applied the defense method on the raw image; 2<sup>nd</sup> row is the image recovered by the GAN-based attack model.