# A. Ablation Analyses

## A.1. General Ablation Analysis

An ablation analysis is shown in Fig. 12. In order to efficiently run multiple-models, it is done with the low-res architecture. The various options include: a no-mask option, a partial adversarial loss that applies only to the masked output and not to the raw output, training without the gradual increase of $\lambda$, and an attempt to incorporate an additional output with a lower resolution to be taken into account, as part of the compound loss. All of these ablation experiments were conducted on the lower-resolution model.

The following description of methods and the associated artifacts correspond to the columns of Fig. 12: (c) No mask. *Bad face edge, glasses occlusion handled poorly.* (d) Adversarial loss on masked output only. *Various artifacts, e.g., around the right eye, one can also observe green stripes near the mouth.* (e) No gradual increment of $\lambda$. *Collapse into unnatural blurred face*. (f) Lower resolution output added for the compound loss. *Weak de-id, checkerboard pattern near the center of the face, and when handling occlusions.* (g) Weak $\lambda$, adversarial loss on masked output only. *Weak de-identification, artifacts near the eyes and eyebrows*.

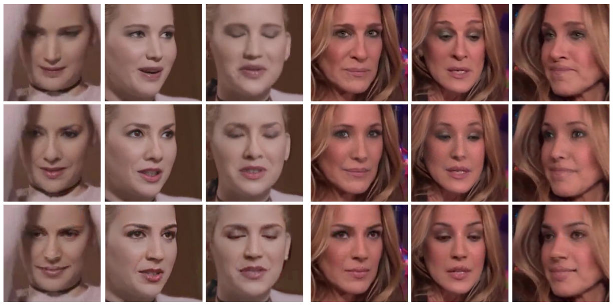

Figure 10. No face-descriptor ablation study. Source (row 1), our model (row 2), and no face-descriptor (row 3), resulting in lower quality results, with noticeable artifacts in the rendered identity.

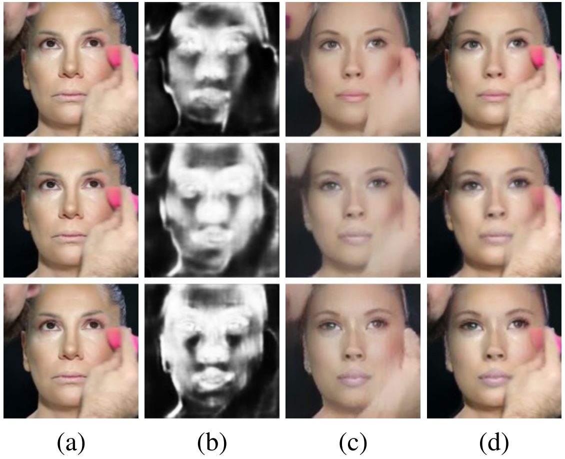

Figure 11. Mask regularization ablation study and mask outputs. (a) Source, (b) mask, (c) raw output, (d) masked output. Compared to our model (row 1), the effects of not minimizing the mask norm ($\alpha_4 = 0$, row 2) can be observed as occlusions (hand, pink element, upper-left face) not handled well, and excessive face regions taken from the rendered image, resulting in distortions. No mask derivative regularization ($\alpha_5 = 0$, row 3) effects can be seen as high-frequency patterns generated by the mask and output frame.

A numerical analysis plot of the ablation study with the same options is provided in Fig. 13. Each method is evaluated along two axes of comparison between the input image and the output image: on the x-axis we show the difference in appearance as measured by the L1 norm between the images; the y-axis shows the difference in ID, as computed by the L1 norm between the VGGFace2 representation of the two images. The plot shows mean results obtained for our method (marked (b) to match the columns of Fig. 12) and the various ablation methods (marked (c)–(g)). As can be seen, our method maintains image similarity and also has a difference in ID that is similar or larger than any other method, with the exception of the method marked as (c). This is expected, since this variant is the mask-less one, which does not blend-in the original image. Variant (f) is considerably more similar to the original image on both axes, since the de-ID performed is very weak with this variant.

## A.2. Face-Descriptor Ablation Analysis

A face-descriptor-specific ablation analysis is provided to emphasize its necessity in Fig. 10. The face-descriptor is highly motivating for the decoder to use, otherwise, minimizing the high-level perceptual loss ($l_{1\times1}$) would be more challenging, as can be seen in Fig. 10. For each source image (row 1), our model result (row 2) can be seen to produce higher-quality results with less artifacts, compared to the model that lacks a face-descriptor concatenated to the latent space (row 3). In the results of the third row, the face descriptor is not concatenated to the z embedding, but still used in the perceptual loss.

## A.3. Mask Regularization Ablation Analysis

The mask regularization parameters $\alpha_{4,5}$ importance can be observed in Fig. 11. They assist in dealing with occlusions, and handling irrelevant regions, that can be taken from the source image, rather than generated (e.g. regions that are not related to the generated face, teeth, etc.). $\alpha_4$ keeps the mask minimal, i.e. blending maximal regions from the source image, rather than the generated one. By avoiding excessive blending of generated regions, less artifacts are apparent on the final output (as observed in row 2). $\alpha_4$ keeps the mask smooth, by penalizing mask derivatives. This can be seen to reduce high-frequency patterns, (as observed in row 3).

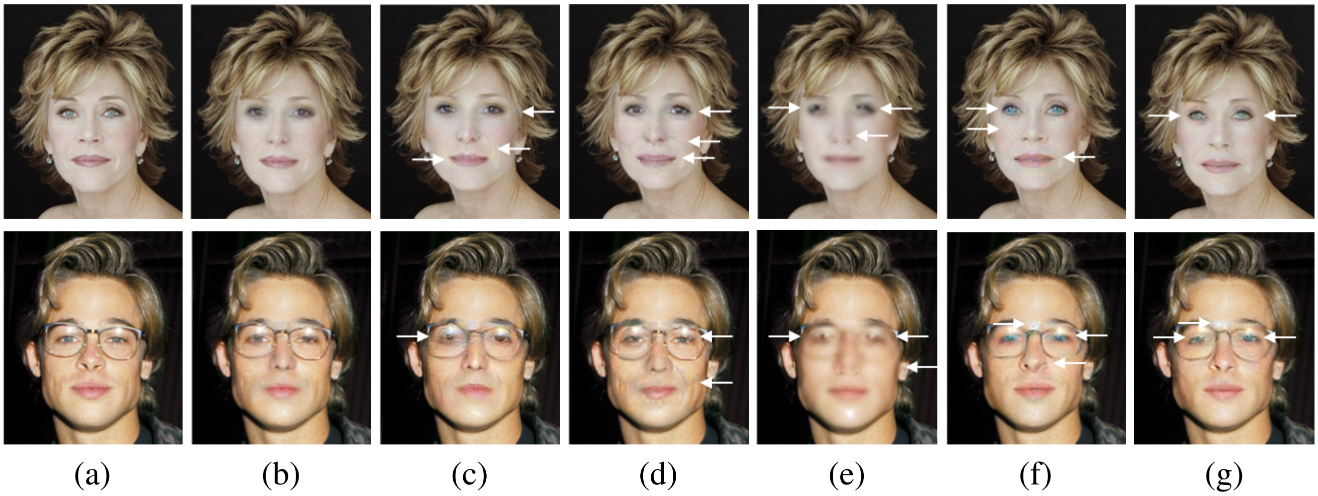

Figure 12. An ablation study. (a) Source image. (b) Our result. (c)–(g) variants, see text for details.

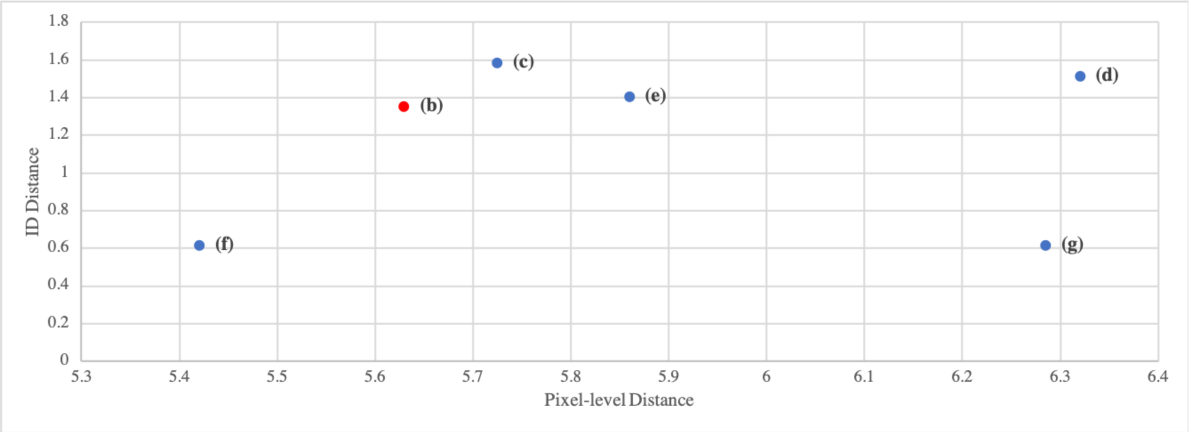

Figure 13. The mean pixel-level distance vs. the mean ID distance. The first should be low, while the second should be high. Shown are the methods of each column in Fig.12(b)–(g).

## B. Additional Comparison with Previous Methods

We provide an extensive comparison with the work of [44]. In the paper we only included one of [44] generated outputs: the sample shown was the first output that gains <50% of recognition rates by an automatic face recognition algorithm, according to [44]. The work of [44] provides several models for different levels of de-identification. In Fig. 14, we present all faces from [44]. The reported recognition rate itself is given in Tab. 6.

As can be seen in the results of [44], the less recognizable the identity is, the less natural the face is. Note that: (1) our model provides for much stronger de-identification results, with the rank typically in the thousands, out of a dataset of 54,000 persons, as reported in the experiments section. (2) all models of the baseline method produce low resolution outputs (64 × 64) compared to our model's much higher resolution (256 × 256).

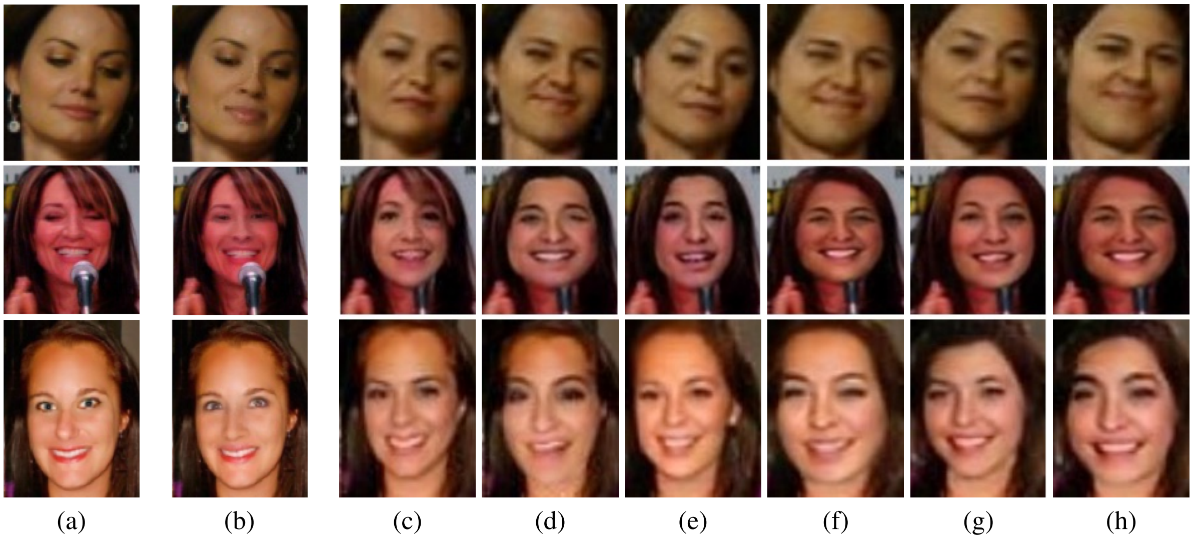

Figure 14. A full set of results for the comparison with the work of [44]. (a) Source image, (b) our generated output, (c-h) generated outputs for [44] of different models. The work of [44] provides several models that provide for different levels of de-identification. As can be seen, models that gain a rate of $< 50\%$ of head obfuscation effectiveness by machine recognizers, provide less natural faces.

| Row \| Column | (c) | (d) | (e) | (f) | (g) | (h) |
|---|---|---|---|---|---|---|
| Row 1 | 70.8% | 47.6% | 36.6% | 18.0% | 22.5% | 7.1% |
| Row 2 | 59.9% | 26.3% | 25.8% | 12.7% | 15.7% | 7.2% |
| Row 3 | 59.9% | 26.3% | 25.8% | 12.7% | 15.7% | 7.2% |

Table 6. Head obfuscation effectiveness for [44]: recognition rates of machine recognizers (lower is better), as provided by [44]