A Quantitative Evaluation Framework of Video De-Identification Methods

Sathya Bursic **Alessandro D'Amelio** **Marco Granato** **Giuliano Grossi** **Raffaella Lanzarotti**

PHuSe Lab - Dipartimento di Informatica, Università degli Studi di Milano

Introduction

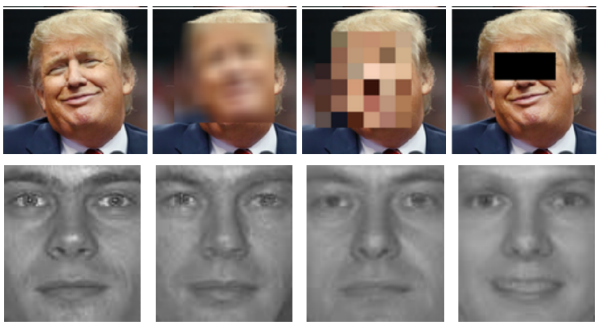
- We live in an era of **privacy concerns**, motivating a large research effort in face de-identification. As in other fields, we are observing a general movement from hand-crafted to **deep learning methods**, mainly involving generative models.
- Although these methods produce more natural de-identified images or videos, we claim that the *mere evaluation of the de-identification is not sufficient*, especially when it comes to processing the images/videos further.
- We take into account the issue of **preserving privacy**, **facial expressions**, and **photo-reality** simultaneously, proposing a **general testing framework**. The quantitative evaluation is applied to four open-source tools, producing a baseline for future de-identification methods.

Framework

We claim that the three main requirements of a de-identification system are:

- **The de-identification itself**, quantifiable as the capability of fooling face verification methods,
- **Expression preservation**, measurable in terms of elicitation of the same Action Units (AUs) in both the original and the de-identified videos,
- **The photo-reality safe-guard**, that we will measure in terms of feature preservation.

Traditional Methods

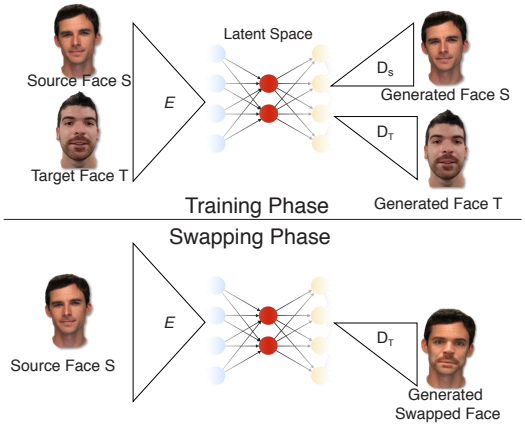


Examples of de-identification: *First column:* original images. *First row:* naive methods, applying respectively blurring, pixelation, masking. *Second row:* results obtained by the k-same method, varying the parameter k.

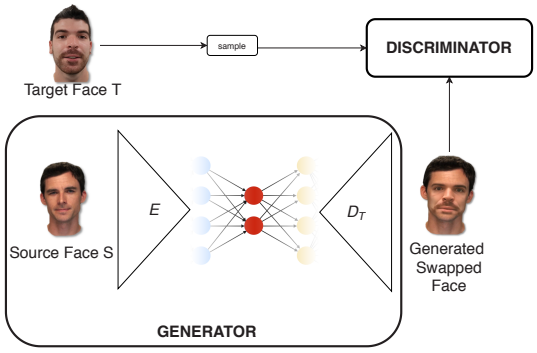
The above methods don't take into consideration all of the three objectives but focus on de-identification only.

Models Tested

We consider and compare four open-source methods: Dfaker, DeepFaceLab, FaceSwap and FaceSwap-GAN. The first three are based on the **autoencoder** architecture:



and the last is based on the **GAN** architecture:



Conclusions

- We introduce a **quantitative evaluation framework for video de-id**, and provide a baseline
- **No one method is optimal** according to the three metrics simultaneously, the objectives present a trade-off
- It is important to **evaluate them jointly**, in order to provide a complete picture of the method's potential
- The expression preservation and photo-reality metrics could be used as a **stopping criteria** for the training phase