ImSecu DeepFake challenge

Francesco Dente

October 26, 2024

1 TASK 1: How easy is it to create a deepfake ?

1.1 DeepFake generator and detector

This section is about both the first part about generating deepfakes and the third part about using deepware detector.

1.1.1 Face Swap deepfakes

I experimented with three different pictures of myself under various lighting conditions and with different accessories (see figure 1).

For the two deepfakes intended to look as realistic as possible, I used image 1a, which features good lighting and no accessories. I also selected two celebrities with similar face shapes to mine[1], in videos showing them from the front.

For the two deepfakes intended to look as fake as possible, I used images 1b and 1c for Millie B. Brown and Billie Eilish, respectively. Additionally, these are cross-gender celebrities with different face shapes from mine, in videos where they move their heads frequently.

All deepfakes were created using Roop.

Source	Avatarify	Deepware	Seferbekov	Ensemble	Detected
Elijah Wood	36%	4%	3%	1%	NOT DETECTED
Jeff Bezos	19%	0%	3%	1%	NOT DETECTED
Millie B. Brown	23%	69%	99%	96%	DETECTED
Billie Eilish	23%	81%	99%	96%	DETECTED

Table 1: Results of deepware on the four generated deepfakes

(a) Wood, Bezos deepfakes

(b) Millie B. Brown deepfake

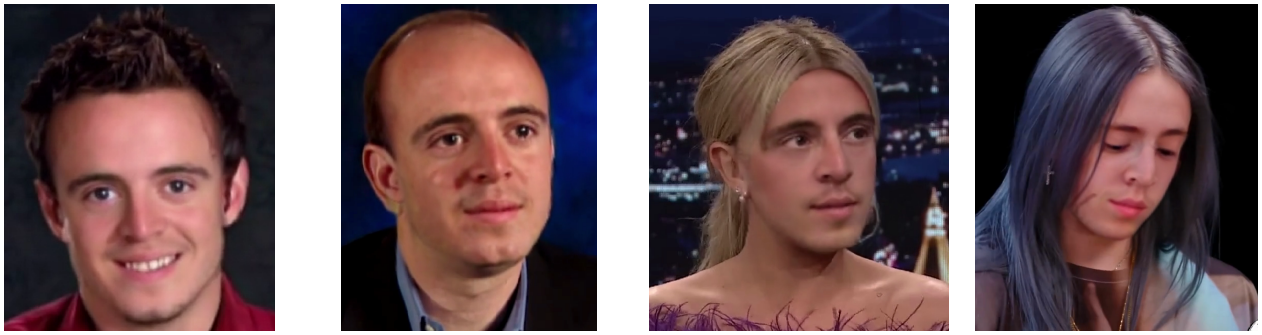
(c) Billie Eilish deepfake

Figure 1: Driving images for face swaps

[1] Elijah Wood is the actor I used for the morphing attack in Challenge 1

(a) Elijah Wood (b) Jeff Bezos (c) Millie B. Brown (d) Billie Eilish

Figure 2: Original frames



(a) Elijah Wood (b) Jeff Bezos (c) Millie B. Brown (d) Billie Eilish

Figure 3: Frames after face swap

1.1.2 Face reenactment deepfakes

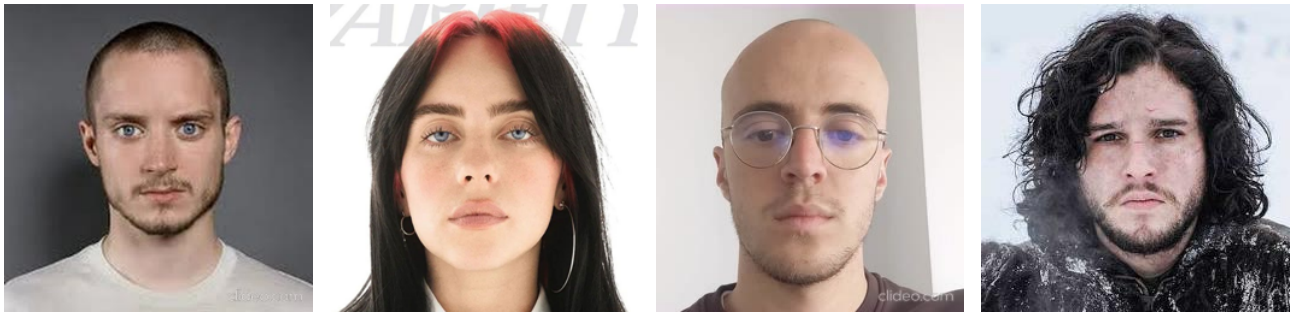
To get less detectable results, I used as driving video, a video of me without glasses, trying to move my head as little as possible. In this way I created the deepfakes of Elijah Wood and Billie Eilish, which were not detected by deepware. On the other hand, the John Snow deepfake, for which I used a video of myself with exaggerated facial expressions, was easily detected by deepware. Finally, I also tried to recreate a deepfake of myself using a stock video as a pilot.

All deepfakes were generated using LIA[2].

Source	Avatarify	Deepware	Seferbekov	Ensemble	Detected
Elijah Wood	49%	6%	2%	4%	NOT DETECTED
Billie Eilish	54%	0%	0%	0%	SUSPICIOUS
Myself	98%	0%	1%	1%	DETECTED
John Snow	93%	0%	1%	1%	DETECTED

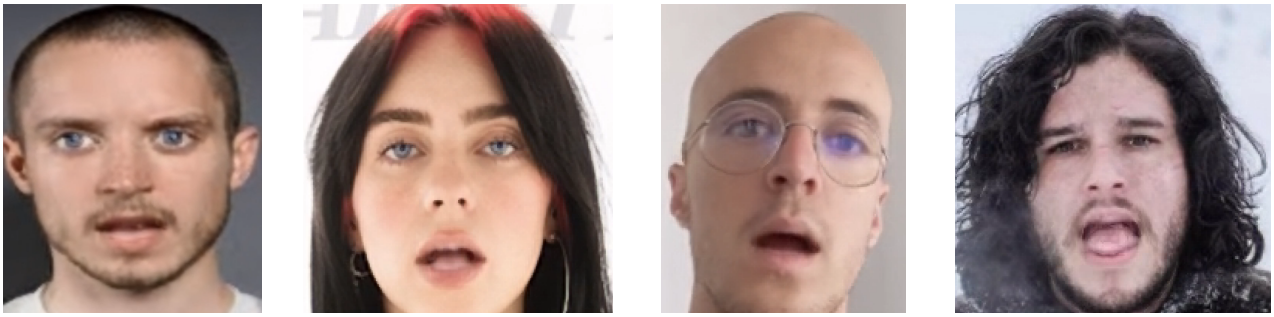
Table 2: Results of deepware on the four generated deepfakes

[2]In order to make it work, I had to resize both images and videos to 256x256



(a) Elijah Wood (b) Billie Eilish (c) Myself (d) John Snow

Figure 4: Images used in face reenactment



(a) Elijah Wood (b) Billie Eilish (c) Myself (d) John Snow

Figure 5: Frames from resulting face reenactment videos

1.2 Questions

1.2.1 Explain what a face-reenactment and a face-swap mean:

They are both methods for generating **deepfakes**:

- **Face-swap**: it consists in replacing one person's face with another person's face in a video or image, trying to preserve the original facial expressions and movements of the target person, creating the effect that the second person is present in the original scene. The traditional approach had some limitations such as the lost of expressions of the input face and the result that looked unnatural. Nowadays, thanks to DeepLearning, the results that we can obtain look very real.
- **Face reenactment**: an image/video is used to drive the expression, gaze, pose, or body of another person's image. It turns an identity into a puppet, that can say or do thing that he never actually did.

1.2.2 What is the main difference between these two categories?

Face reenactment focuses on modifying the facial expressions and movements of a person in a video based on another person's expressions and movements (Use an identity as a puppet), while **face-swap** involves replacing one person's face with another's in a video or image, creating the effect that the second person is present in the original scene.

Furthermore, **face-swap** usually needs to detect and align faces in target/source, and then blend the face seamleassly onto the original replaced face. **Face-reenactment**, instead, focuses more on detecting and tracking key facial landmark points, in order to apply these movements and expressions to the face we want to *use as a puppet*.

1.2.3 What kind of input data do you need for creating a face-reenactment or face-swap deepfake? Explain the source and pilot/driving video

Since I did not find a precise definition of the distinction between source and driving data, in different papers they give different meanings to the words. I will quote the definitions from the paper *Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward*, which is referenced in the course slides about deepfakes.

- In **face-swap**, or **face replacement**, the face of the person in the source video is automatically replaced by the face in the target video

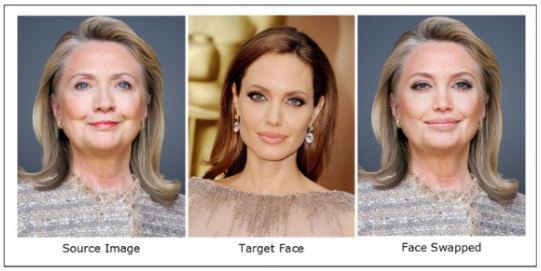


Figure 3: A visual representation of Face-Swap based deepfake

- In **face-reenacmtent** the source data is the video of the person whose facial movements or expressions we want to transfer to the target/driving data. The target/driving data is the image of the person whose facial movements or expressions will be altered.

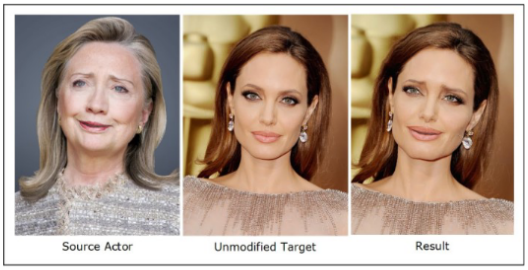


Figure 6: A visual representation of face-reenactment based deepfake

1.2.4 Opinions about deepfake generators

In my opinion they have both helpful and dangerous use, but the threats are worse than the benefits we could have.

- **Helpful uses**
  - Entertainment, for movies, dubbing, visual effects
  - Virtual reality, for creating more realistic experiences
- **Dangerous uses**
  - Create misleading content that can spread misinformation.
  - Manipulate people's faces without their consent (privacy violations)
  - Harm someone's reputation portraying them in particular scenarios
  - Impersonate someone for identity theft

2 TASK 2: Did you win an award at the Cannes Film Festival?

2.1 Chosen algorithm

I use Roop (**Face-swap**) because I want to get the effect that I am present in the original awards scene, by superimposing my face on the real person receiving the award (Cillian Murphy). This would be very hard with either of the other two face-reenactment algorithms.

This section answers to question 2

2.2 Choice of inputs

In order to make the deepfake as real as possible, I had to choose wisely both the target image and the driving video. I noticed different important criteria:

- Camera angles, light and pose of both image and video must match
- Driving video with people with a huge beard creates difficulty in generating a realistic lipsync, and can create strange effect around the mouth.
- Face occlusion, caused by things like hands covering the face, hair covering part of the face, or objects blocking the face, leads to low quality deepfakes. The main problem is that this can cause some frames to have the original face rather than the target face [3].

This section answers to question 1

2.3 Credibility of the video

To generate the final video, I tried with two different approaches:

- **Using only face swap**: The swapped face lacked clarity and sharpness.
- **Using both face swap and face enhancement**: this approach resulted in a clearer swapped face, but it appeared less natural in relation to the surrounding environment, making the fake effect more noticeable[4].

The **Deepware** deepfake Detector flags both of my videos as deepfakes. In table 3, you can see how various detectors respond differently to each video.

The detection of the deepfake is expected since there are several frames in which Cillian Murphy's original face is still visible. This poses a significant problem, not only because it allows Deepware to identify the video as a deepfake, but also because it makes the video not credible enough to fool real people. As previously mentioned, this issue is mostly due to face occlusion in the video and can be addressed in the following ways:

- Acting directly on the Deepfake generator
- Using pre-processing techniques to make life easier to the deepfake generator
- Using post-processing techniques, in order to manually patch the target image in the few frames where we can still see the original one. For example, in my video I managed to remove the original face popping up for few frames by simply replacing them with the previous frame with the correct face [5].
- Using videos with as less face occlusion as possible, or do some tricks in order to replace the short intervals in which we have face-occlusion, with other frames in which the face of the *victim* is not shown.

Source	Enhanced	Avatarify	Deepware	Seferbekov	Ensemble	Detected
Cillian Murphy	No	19%	37%	99%	86%	DETECTED
Cillian Murphy	Yes	80%	0%	42%	8%	DETECTED

Table 3: Results of deepware on award-winning deepfake

[3] This is the biggest problem I encountered when creating my deepfake

[4] This is very clear when watching the whole video rather than just a frame of it

[5] This was possible only because the original face appeared for very few frames, otherwise the cut would be clearly noticeable, that's why there are still problems in my final video.

(a) Not enhanced

(b) Enhanced

Figure 6: Me winning an oscar award

In conclusion, I opted for the version without enhancement, since, given the overall low quality of the video, the enhancement effect appeared particularly unnatural.

This section answers to both question 3 and 4