Advanced Techniques in Artificial Intelligence

Deepfakes

by Facundo Dartayete, Diego Dabezies and Florian Deublein

Abstract

Deepfakes have risen in popularity over the course of the last few years. Their quality rises with increasing computing power and as new techniques are being found frequently, anybody with access to a computer is able to create decent deepfakes. In this report we will explain the basic principle behind such deepfakes and discuss their possibilities and dangers. Subsequently we will also talk about deepfake detection and compare deepfakes to other domains of AI.

1 Introduction

The term “deepfake” is derived from the terms “deep learning” and “fake”. Deepfakes are synthetic media where an image, video or audio is modified to mimic someone else's appearance and/or voice.

The extreme advances in AI techniques have seen this tool gain a lot of popularity in the last few years, generating more convincing results every year. Due to the abundance of malignant uses for this technology and its widespread use, governments and organisations have created related AI tools to detect these deepfakes.

Deepfakes computing heavily relies on other well known AI structures, mainly Deep Neural Networks, Autoencoders and Generative Adversarial Networks (GANs). The latter - with a specific focus on generating new data from the training set - has very powerful use in photography and deepfakes as it has the ability to generate the target’s face on top of the original source.

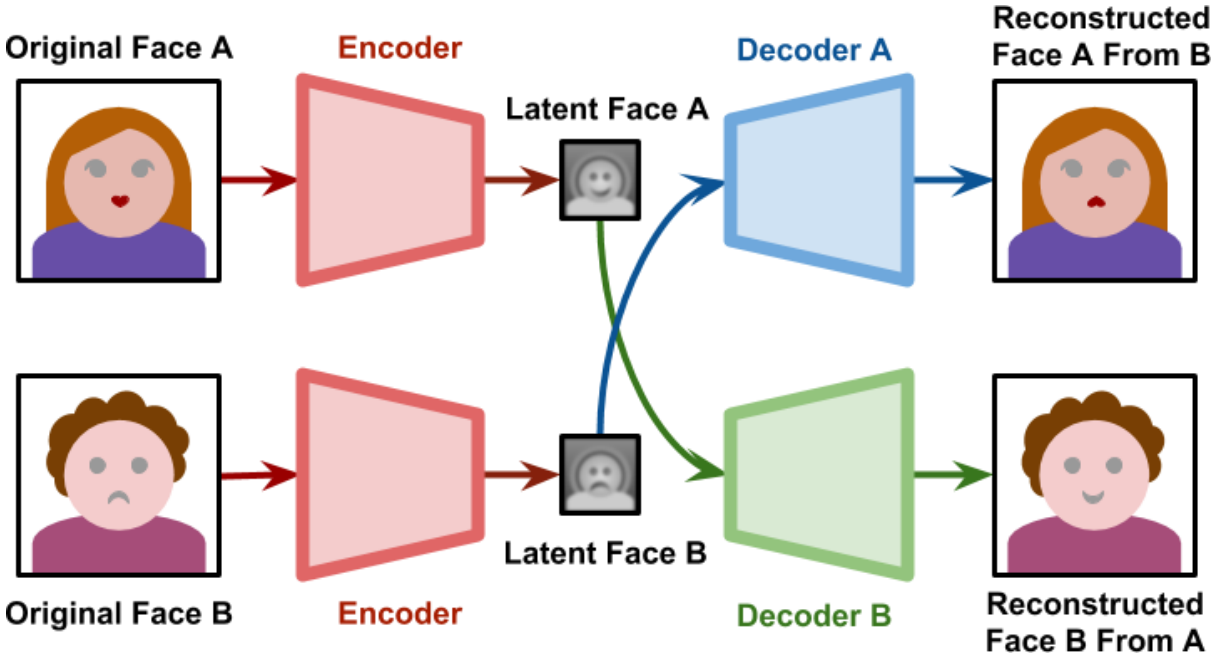
Nowadays, although very realistic deepfake videos still require a lot of computing from a powerful machine, quite good results may be obtained in a couple of seconds from a free app on any smartphone.

2 How do they work?

The first step for creating a deepfake video is to train an AI encoder with data from both the target and the original video. The algorithm will learn similarities between them and generate a compressed result with them.

In the next step a decoder is trained that is able to recreate the target’s face from the encoder result using deep neural networks. Once a good result is achieved, the original video can then be fed to the decoder which will recreate the target’s face in the face of the

original video, matching its facial expressions and looks.



3 Possibilities

Deepfakes, like any other technology, depends on the use we give it. There is a broad range of possibilities with this technology, some of them very useful to society. We will mention the most common and popular use cases.

3.1 Education

Audiovisual content is very rich and entertaining in a classroom. But imagine how engaging it could be if you could bring back to life the very own characters of historical events to narrate their experience. This can all be done with the use of AI deepfakes.

3.2 Cinema

Deepfakes could produce excellent results where historically accurate movies could be played with the actual characters inserted through deepfakes into the film. Furthermore, deepfakes even the playing field in an industry where Hollywood's big corporations with seemingly infinite resources produce spectacular special effects that independent producers cannot keep up with.

3.3 Accessibility

Soon enough these tools will be able to synthesize voices well and fast enough to provide more independence for people with speech and hearing problems.

3.4 Public safety

In places where human rights activists are needed but may encounter a harsh response from the local government, deepfakes may help them spread their messages while keeping the anonymity they need.

4 Dangers

One of the main dangers of deep fakes is fake news. Due to the fact that fake but very realistic videos can be created at increasing speeds and are very difficult to differentiate from real ones, anyone with access to a simple computer can create convincing deepfakes of celebrities or politicians. These can then be used to put fake news in the world about certain topics that can threaten society by manipulating beliefs and can damage opinions about celebrities permanently.

4.1 Politics

There have been many times where deep fakes were used for political reasons to misinform the society in order to achieve certain goals. One example is in Belgium, where a Belgian political party created a deep fake of a Donald Trump speech where he was speaking about the Paris climate agreement. In the video he says: "As you know I had the balls to withdraw from the Paris climate agreement. And so should you." and only in the end of the video he said that the video was fake, but it wasn't translated into dutch subtitles, which generated a lot of debate in the belgian society about if belgium should withdraw from the Paris climate, showing the power of deepfakes to give rise to fake news.

4.2 Society

Many people think that deepfakes are a threat to society due to the power they can have to misinform people. The worst scenario they can generate is a society where people can't differentiate fake videos from real ones, making people highly distrustful with every video they watch. There is evidence that people are already using deepfakes to discredit genuine video evidence, saying that a real video is a deepfake to deny the evidence. One example is the case of one of the victims of Jeffrey Epstein who reported that had been forced to have sex with his powerful friends, including Prince Andrew. She presented a photo in which she appears with the prince and then in a BBC interview, Andrew came out to claim that the photograph was altered and false, casting doubt on the victim's evidence.

4.3 Personal harassment

Another unethical use deepfakes present is online harassment. In this case a humiliating video of someone can be synthesized and made viral in a couple of hours, presenting tremendous psychological repercussions to the victim, even if the content is proven fake afterwards.

Celebrities are often victims of these acts as there is a lot of video material for the encoder to practice with.

5 Deepfake detection

So won’t anybody ever again be able to trust any video they see online and chaos will arise from distrust in politicians and government systems?

Luckily, science is already fighting fake news by inventing and improving other deep learning-algorithms that are able to distinguish real from fake. This works because deepfakes - although hard to observe with the human eye - are still far from perfect. Hints for fake news are for example inter-frame inconsistencies or people blinking less frequently than average in videos because there are fewer photos with shut eyes to learn from.

6 Conclusion

As deep fakes seem potentially extremely dangerous compared to the still quite powerful achievements reachable, we believe there is a lot of investigation to be done in the following decades to win the race between deepfake detection and fake news and also to exhaust the possible improvements for society. The use of this technology should be further regulated by the law to prevent it’s misuse.

Stopping or even forbidding further research on the topic would have disastrous consequences because malicious uses of deepfakes will continue to improve nevertheless and the damage once done to the image of people and institutions is often irreversible.

Bibliography

Works Cited

Hu et al. “Dynamic Inconsistency-aware DeepFake Video Detection” in *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21,* International Joint Conferences on Artificial Intelligence Organization, Aug. 2021, https://doi.org/10.24963/ijcai.2021/102.

Alan Zucconi. “Understanding the Technology Behind DeepFakes.” *Alan Zucconi*, 18 Apr. 2018, www.alanzucconi.com/2018/03/14/understanding-the-technology-behind-deepfakes/.

Ben Dickson is a software engineer and tech blogger. He writes about disruptive tech trends including artificial intelligence. “What Is a Deepfake?” *PCMag UK*, 4 Mar. 2020, https://uk.pcmag.com/news-analysis/125140/what-is-a-deepfake.

Catanzaro, Michele. “¿Qué Peligros Entrañan Realmente Los 'Deepfakes'?” *Elperiodico*, El Periódico, 15 Apr. 2021, www.elperiodico.com/es/entre-todos/20210414/peligros-entranan-realmente-deepfakes-11652389.

“Deepfake.” *Wikipedia*, Wikimedia Foundation, 2 Sept. 2021, https://en.wikipedia.org/wiki/Deepfake.

Jaiman, Ashish. “Positive Use Cases of Deepfakes.” *Medium*, Towards Data Science, 15 Aug. 2020, https://towardsdatascience.com/positive-use-cases-of-deepfakes-49f510056387.

“¿Qué Son Los Deepfakes y Cómo Protegerte De Ellos?” *Grupo Atico34*, 6 Oct. 2020, https://protecciondatos-lopd.com/empresas/deepfakes/.

slide 1

Have you ever been in the situation to be bored of university from home and you'd rather go on holiday but the university demands your presence in lectures?

If you could just let somebody else attend to the lectures ... well, your time is about to come!

Why? That's what we're going to explain to you in today's presentation about deepfakes

---slide 2

made by Facundo Dartayete, Diego Dabezies and me, Florian Deublein

---slide 3

So ... what are deepfakes?

The term deepfake is derived from Deep Learning and fake,

because synthetic media is modified to mimic someone's appearance and/or voice (=fake part)

by techniques like Deep Neural Networks, Autoencoders and Generative Adversarial Networks (=Deep Learning part)

To show you an example of a deepfake, we prepared a video for you (!video)

---slide 4

It should be mentioned that the program only had a single photo from each of us

and the result - although obviously detectable as fake even with the human eye - is already quite good

---slide 5

But how do deepfakes work?

One of three steps consists of an AI encoder learning similarities between the target and original video.

In another step another encoder uses DNNs to learn an internal representation of the expressions and movements of the target's face.

Finally a decoder will recreate the target's face in the face from the original video.

---slide 6

Let's now talk about the possibilities achievable with deepfakes:

In education, deepfakes could be used to let historical characters narrate their experience during a historic event,
leaving deeper impressions for students which makes it easier to learn.

Also in cinema historical characters could actually play their own role or one could even "revive" a dead actor for an ongoing series.

Deepfakes could also allow people with speech problems to speak with authentic synthetic voices, probably improving their quality of life.

And last but not least, human right activists in places where there is to be expected a harsh response from the government
could be anonymized to keep their identity safe.

So far, deepfakes seem like a very positive and useful technology, but let's also have a look at the dangers:

---slide 7

In politics, parties or people with malintentions could misuse deepfakes to misinform the society about a politician's opinion.

Also with rising popularity of deepfakes nobody might be able to distinguish real videos from fake ones such that real videos could be doubted, too.

Another very bad misuse is to create a humiliating video of sb and make it viral within a couple of hours. Victims and their images can suffer tremendous damage
even after the video's content is proven fake afterwards.

So ... won't anybody be able to trust any video ever again?

---slide 8

As facundo said, deepfakes can be very dangerous for society, so luckily, science is already fighting fake news by inventing and improving other deep learning-algorithms that are able to distinguish real from fake videos and images.

This works because deepfakes are still far from perfect but they are improving day by day.

Hints for distinguish fake videos are for example less frequent blinking (less photos with eyes shut, image 1) or inter-frame inconsistencies (show in graphic, image 2).

---slide 9

So ... what's our conclusion about deepfakes?

in our opinion, the extreme dangers of deepfakes **outweigh** the possible improvements.
That means, deepfakes should be regulated and misuse has to be heavily punished by law.

At the same time, stopping or even forbidding the research on deepfakes would have disastrous consequences,
because malicious uses of deepfakes will continue becoming better and better and without further research on the "good side", the race between creation and detection of deepfakes will be lost.

As damage once done to the image of people and institutions is often irreversible, further research is obligatory!

---slide 10

Here are the sources we used for the presentation - feel free to dig further if the presentation sparked your interest.
Thank you a lot for your attention. Are there any questions?