Data driven challenge

Spotting fake videos



Fontys Hogeschool ICT

Data Science Minor – Max de Goede

Contents

[Topic 2](#_Toc23590900)

[Describe the innovation 2](#_Toc23590901)

[Describe approach using data analytics 6](#_Toc23590902)

[How to implement 6](#_Toc23590903)

[Process report 7](#_Toc23590904)

[Results 7](#_Toc23590905)

[Conclusion 7](#_Toc23590906)

[Reflection 8](#_Toc23590907)

[Digital visualization product 8](#_Toc23590908)

[Dataset used 8](#_Toc23590909)

[Code 8](#_Toc23590910)

[Figure list 8](#_Toc23590911)

## Topic

### Describe the innovation

#### Problem/Issue

Each day that passes by, it becomes clearer and clearer that there is an issue surrounding fake videos and impersonations in videos. These videos keep popping up more often in the media. The problem surrounding these videos is the fact that for humans it becomes harder and harder to distinguish real from fake. In the current day, this is something doesn’t happen often, but AI that creates these fake videos are becoming better. If used wrongly this could have catastrophic impact on society.

As an example, one could frame Donald Trump of saying something he did not say.

This is an issue that has to be addressed and countermeasures have to be built before damage can be done.

The scope does not only stop here, the idea that ai can recreate faces with near-perfect facial recognition indicates that this technology could lean towards the side of warfare. Everyone knows the (near sci-fi) drones that are designed to kamikaze on someone by facial recognition (a great example is the movie Angel has fallen). One can imagine a day where people can wear masks that change their facial appearance (It sounds sci-fi but so did mobile phones years ago)

I could keep going on nearly forever on cases where this can be a problem or even a solution. The message I am trying to send is that this technology is a technology where countermeasures need to be created as the AI evolves.

#### Innovation/Future prevention

So what exactly is an innovation that could deal with this? Before we can even think of countermeasures the following thought needs to be brought to mind. Can it be stopped? I think everyone can answer this for themselves, there is no way we are going to stop faked videos. This technology will come. So the next question arises, what countermeasures can we create to counter these videos? The simple thought of what could be is already a big project in itself. But the focus of this challenge is not on what the problem is, but what innovation could be a possible solution.

To narrow the scope of the challenge, let’s look at how we can spot these fake videos. And how to systematically bring them down/counter them. But what could counter these videos/imagery? The first thing one would come up with is the fact that humans will spot the fakes out. Many tests have already been conducted and it turns out that this is not the case, spotting out all fake videos is impossible. For this, an AI (or some kind of technology) is necessary to spot out the fakes.

One might think that there has been nothing that could set the groundwork on how this can be spotted, but I differ in this thought. There are many examples of how we could create an infrastructure. We just need to correctly “steal” the idea and implement it into this new topic. So what is this idea we have to steal? Bots in games, have existed for years and companies are trying to constantly find ways to find spot these bots! So why not take inspiration from the groundwork others have laid out?

#### Detection of manipulated images

How would we create a system that would be able to detect the difference between real and fake imagery/videos? The first step would be to create a dataset consisting of manipulated and non-manipulated images. This data would exist to find train a model, that would find patterns in faked imagery.

Where would the acquisition of such data come from? The best scenario would be a world in which we have access to enough aired televised data from news channels and create a set amount of fakes within the dataset.

But before something like this is even possible one has to think of the following, there will always be more than 1 network that creates fakes and each creates it in it’s own way. This forces the network to learn patterns in each of these networks. The biggest issue here is getting data from all networks. And what happens if there is a network that purposely creates bad fake data to mess around with the AI/system. These types of questions are really hard to answer and the answers can only be found by actually doing it.

This means that to get a data set for the network to train on, one would need the following;

* Big amounts of image data of each model
* Normal video data
* Purposely bad created data
* Faked videos

These 4 together will hopefully create a dataset that will be able to detect fakes, but a problem arises from this. The moment this system is trained it indirectly could train other models to find ways to improve itself so the system would not be able to detect the fakes. This that the project would be secretive and a constant flow of the newest images to keep the system up to date.

#### What current measures are available?

**Images in the following chapter have been taken from the paper listed below, I would love to refer the reader to the paper and take some time to read it. It is really interesting and worth the time! A summary has been written below, in case the reader doesn’t have the time necessary.**

Luckily there have been bright minds that have thought about this problem and decided to start on it. The (Technical University of Munich, University Federico II of Naples, University of Erlangen-Nuremberg, 2019) released the following [paper](https://arxiv.org/pdf/1901.08971.pdf).

*\*Code,* [*GIT*](https://github.com/ondyari/FaceForensics)*\**

*What were the results from the paper?*

The paper worked with the dataset FaceForensics+++ that mainly consists of 4 networks (Face2Face, FaceSwap, DeepFakes and NeuralTextures). In the paper faked vidoes/imagery is referred as synthetic image generation, the paper mainly focusses on creating a standard for detection of the networks mentioned above.

The paper splits synthetic image generation into two categories, facial expression manipulation and facial identity manipulation. (As seen in figure 1).

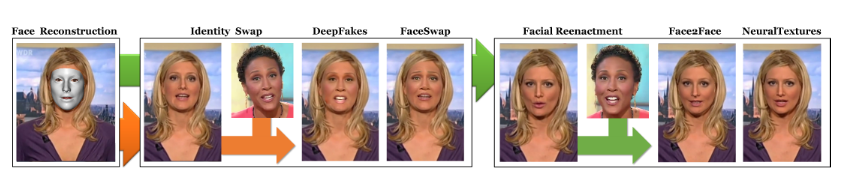


Figure 1, categories of synthetic image generation. Image comes from The paper of TMU.

The paper has a nice image that show’s how their AI works. (Figure 2)

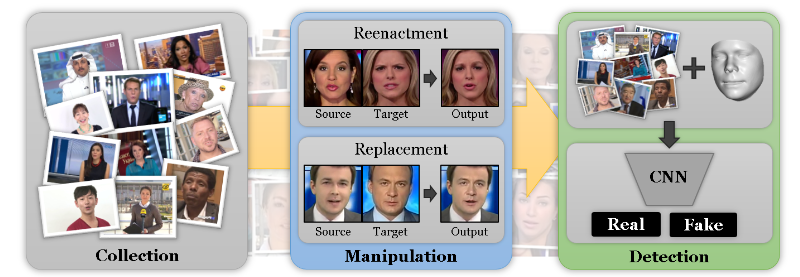
But what was the actual result from this AI? This can be seen in figure 3, depending on image quality the score of the AI can fluctuate.

Figure 2, How AI is trained and how it works, Image comes from the paper of TMU

*\*note if the reader is interested in how the tests are conducted, you are referred to chapter 4 in the paper\**

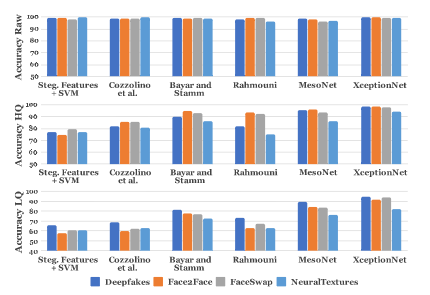


Figure 3, results from the paper, Image comes from the paper of TMU

### Describe approach using data analytics

So there is an AI that has results that could deliver something that can be useful and could improve current services and usages. The problem with this is not an AI that does not work, it has been proven that it can work with the correct data.

The problem now shifts direction, the question is changed from creating technology that can identify. To create an infrastructure and data flow that would be able to cope with the constant training of such a network. Let’s get some requirements straight.

* Data flow needs to be constant
* New networks must be developed and implementable
* Usage of the most current version of the AI must be useable

The first thought that comes to mind when reading the requirements are the requirements even plausible with the current technology. But there is something we are cannot oversee here, 5G. If 5G is created in such a way where it is implementable for this solution, that would solve quite some data availability issues. For the sake of thinking outside the box, let’s imagine we live in a perfect world where the 5G infrastructure is working.

The most important thing to make this work is to have enough computing power to be able to at least run 2 networks. Why 2? Well, firstly you would need the network that detects the faked imagery and then you would need one that can create faked imagery. Why would this be the case? This would mean that you could have the best-trained network for both, and yes the data might get overfit with the particular network. But this does not mean that you could have multiple different parallel networks running. This is the most important part, the ai needs to be trained up to a level that nearly any small faked imagery can be detected. The data needed for this is nearly unimaginable. But that’s the fun part theorizing about this.

I think that there has to be near perfect infrastructure of data would one want this to go without any issues since the amount of data that needs to be fed, is immense. But we should not think about boundaries at all, this would be giving data a new meaning. In this case, data could be the networks that deliver faked imagery. I could go on all day about this, but I think I have speculated enough on how this could work and how this is delivered.

### How to implement

How could we implement this? I will go with the first idea I had which was looking at live television and how this could work.

I will paint a small scenario, in the future it is possible that hackers can get access to live news and forge the data live and can keep this under the radar, from the news channel. \*note In this scenario all the data coverage and ai is up to date and can works perfect\*

In this scenario the given channel has the AI running, the AI is not only present through with the actual recordings. But is also present on some external servers and reviews the broadcast externally. Here the AI immediately spots issues with the broadcast. It then immediately sends a message towards the broadcasting team and they can shut down the broadcast or even better the AI, is trained in such a way that it can correct what the is broadcasted.

## Process report

### Results

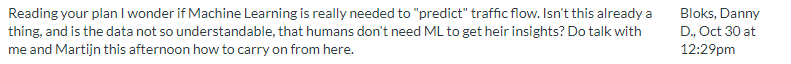
### I am very pleased with the result that came forth from my research. But what I liked the most came when the ideas started flowing into my head of what could be possible. I’m the type of person that loves to think about stuff and could even be possible. I tell people, ”there is no need to think outside the box. Just don’t think about a box, let your mind be free and see what comes forth.” I do like to just let my head explore possibilities and I will admit that for the challenge I did limit myself on what came from my ideas. Personally, it would not have been a good report if this went towards the side of sci-fi. I tried to keep this as realistic as possible, which is one of the reasons that I took my time to read and understand what the current technology that has been developed could achieve.

### Conclusion

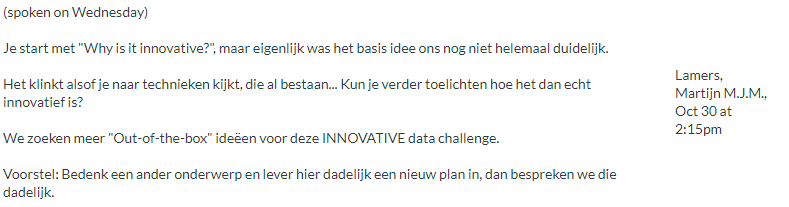
The first idea I had about detecting traffic abnormalities was not accepted and I will admit I am happy it was not, I was now able to take time and look into another topic and brainstorm on it and came out with some ideas I thought were pretty nice.

The report looks good and has both sides of the coin, it has the part where I explore what is currently available, what could be available in the near future (5G) and then my imagination work a bit on what could be. The distribution on how I did this felt natural. I would have loved to let my mind completely explode with ideas and just write these down, but I felt that my document personally was already leaning towards the longer side of things and it would just make it a longer story with no extra context.

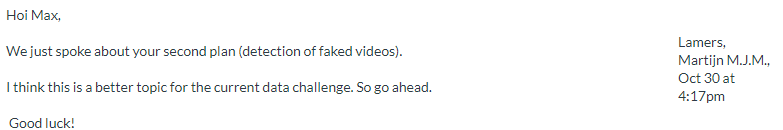
### Reflection



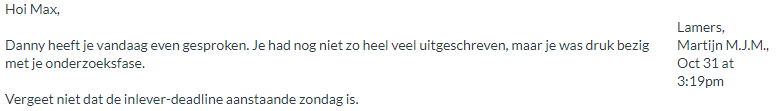
My first idea what not accepted, I had some issues understanding the first feedback, as my idea was not about improving the flow of traffic but detecting traffic anomalies.



After siting together with both teachers, I understood what the problem was and agreed with them. Even if the issue was not that my idea was misunderstood. A system is already in place and can start working. The conclusion we got from this idea was for me to create a new idea and present it.



I was glad to see that my new idea was accepted, though we did not have much time to discuss it, it felt as if the teachers agreed with what I had made.



My first day working on the idea (Thursday) I spent it on doing some proper research and finding even more background information and properly reading the paper. This is all I could show when Danny came by 😊.

## Digital visualization product

No visualization used, a reference however is made to the XceptionNet visuals and results.

## Dataset used

No dataset used, a reference however is made to the FaceForensics+++ dataset. The dataset used to train XceptionNet.

## Code

No code used, a reference however is made to the XceptionNet code, created by TMU.