




Sri Lanka Institute of Information Technology

Topic: Deepfake Detection

Individual Assignment

IE2022 - Introduction to Cyber Security

Submitted by:

Student Registration Number	Student Name
	Abeywickrama O.D.

Date of submission:

28th of May 2021

Table of Contents

Abstract	3
1. Introduction	4
2. Evolution of the topic	7
3. Future developments in the area	12
4. Conclusion	20
5. References	21

Table of Figures

Figure 1: The timeline of Deepfakes evolution.....	9
Figure 2: The general processing pipeline of deepfake detection.....	15

Abstract

Deepfake generation, also known as multimedia manipulation, is quite advanced nowadays and is used all over the world for both benefits and drawbacks. Deepfakes, or artificially generated audiovisual renderings, will frequently be used to smear a persona or sway public opinion. Deepfake technology is a controversial technology that has a number of flaws (society). People also have to deal with a phenomenon known as deepfake. For reporters, social media platforms, and the general public, detecting deepfakes is becoming increasingly vital.

However, most people utilize it for nefarious purposes such as exacting revenge or defaming public personalities. Many articles have been written about video deepfake detection, but less emphasis has been paid to audio deepfake origination. This report does, however, include information on audio detection. As a result, this article gives a good overview of both audio and video deepfakes, as well as the many detection algorithms that have been developed throughout time.

It also provides an introduction to deepfakes, as well as information on their history and evolution over the for all intents and purposes last particularly few decades, as well as how to for the most part create and definitely identify deepfakes, which essentially is quite significant. Open challenges in deepfake detection essentially are also discussed, as a kind of complete understanding of those impediments actually is required to generally solve them, showing how actually open challenges in deepfake detection particularly are also discussed, as a pretty complete understanding of those impediments kind of is required to essentially solve them in a fairly big way.

• Introduction

We may see an exponential increase in multimedia content such as films and photographs in cyberspace because there are many affordable digital smart devices such as computers, cellular phones, tablets, and digital cameras. Furthermore, since social media networks have grown in popularity over the last decade, users are sharing their content at an increasing rate. As a result, there has been a considerable increase in the production of multimedia content as well as the simplicity with which these items can be accessed. With the use of complex algorithms, we can also see a significant advancement in the field of ML (Machine Learning). These advanced algorithms may quickly manipulate multimedia content in order to spread false information on social media sites. Because of all the misinformation that has been manufactured, determining the truth, and trusting information has become extremely difficult.

Above all, we live in a culture where bad actors utilize disinformation to manipulate public opinion, a period dubbed as the "post-truth" era. We must not underestimate the potential for incorrect information to have negative consequences such as defamation, election rigging, the instigation of wars, and much more. Deepfake creation is growing in popularity these days, and it can be used to propagate misinformation around the world. This could result in server damage in the form of bogus news.

What exactly specifically are deepfakes, actually contrary to popular belief. They really are audios and videos that actually have been synthesized, altered, or created by literally AI actually (Artificial Intelligence), definitely contrary to popular belief. Video evidence basically is now widely recognized in all areas of litigation and basically criminal justice, demonstrating how what exactly particularly are deepfakes, which essentially is quite significant. Such videos used in this industry should actually be authenticated and checked for their veracity in a for all intents and purposes major way. Deepfake detection actually has essentially become pretty much more difficult as the technology particularly has improved, which for the most part is fairly significant. User-friendly manipulation tools like Zao , REFACE , FaceApp , Audacity , Soundforge will basically make the verification

process very much more complicated after creating for all intents and purposes such a deepfake, showing how for all intents and purposes such videos used in this industry should particularly be authenticated and checked for their veracity in a subtle way.

Deepfake audio is that the generation of a man-made voice through the utilization of AI. Deepfake audio is meant to mimic the voice of associate actual person, and might be used alone, or in conjunction with deepfake video [1].

There are a few categories of deepfake videos as stated below.

1. Lip syncing
2. Face swap
3. Face synthesis and attribute manipulation
4. Puppet master
5. Audio deepfake

Face swap deepfake s are used to switch the face of the target person with a source person to create a false video including actions that are done by source person, but now swapped by the target person's face. These kinds of videos are made to defame popular public figures by showing them in situations in which they never were. This can harm their reputation in the public eye. As for an example, nonconsensual pornography [2] can be taken.

The gestures of a target person's lips are altered in lip syncing-related deepfakes to make them compatible with a selected audio clip. Lip syncing is a method used to make it look as though the target person is speaking in a different manner than they really are.

In puppet master deepfakes, source individual facial gestures, such as eye movements, head movements, and even whole body [3] motions, are replicated as the target person. These films can be used to imitate someone else for the purpose of seduction.

Face synthesis and attribute modification is concerned with the generation of realistic face images as well as the altering of facial attributes. This is used to build phony profiles on social networking. The method of duplicating a specific person's voice using deep learning

algorithms to depict them speaking something they didn't say is known as audio deepfakes, also known as voice cloning. [4]

Even though video deepfake detection is thoroughly considered. Less attention is given to identifying audio deepfakes. This has become very advanced throughout the last few years. This can be identified as a threat to automated speaker verification systems as well as a voice-controlled system. Voice cloning puts people in a huge risk of defaming and in empowering criminals to manipulate private calls or business deals. Recently it was reported that some robbers used this method to deceive bank employees with a voice of their company executive to transfer hundreds of dollars into secret account [5]. Therefore, we should also pay attention to audio forging as well as video deepfakes.

Even though there are not many surveys done about audio deepfakes there are many surveys done upon deepfake images, video detection, visual deepfake detection, face manipulation detection etc. Focuses on image manipulation and multimedia forensic tactics in general [6].

- **Evolution of the topic**

The first known case of handling multimedia content literally was in 1860, when a portrait of statesman John Calhoun basically was edited by swapping his head with that of US President Sir Abraham Lincoln [7] in a subtle way. To definitely create actually such images, objects for all intents and purposes are frequently added, removed, and copied in between two photos [6] in a definitely big way.

To for all intents and purposes improve scale, appearance, and viewpoint coherence, post-processing techniques generally such as rotating, scaling, and color altering for the most part are performed, which particularly shows that to really create basically such images, objects specifically are frequently added, removed, and copied in between two photos [6], which for all intents and purposes is fairly significant. Aside from conventional methods, developments in computer graphics and DL (Deep Learning) techniques mostly have for the most part enabled people to experiment with a number of automated multimedia manipulation systems., or so they generally thought.

Autoencoders or GAN are currently being utilized to create films for a number of purposes , including the creation of realistic human faces depending on any attribute [8]. 'Shallow Fakes,' often known as 'cheap fakes,' are audiovisual manipulations done with less expensive and more accessible instruments. Shallow fakes are simple video editing techniques that involve accelerating, delaying, trimming, and selectively splicing together previously watched footage in order to change the context of the information provided. A video of US Speaker Nancy Pelosi was altered in May 2019 to make it appear as if she was high and slurring her speech [9]. On Facebook, this video received over 2.2 million views in just two days. In filmmaking, video modification is utilized to gain an edge.

Video rewriting was a well-known early intellectual undertaking. It was released in 1997 [10] with the intention of being used in movie dubbing. The effects were stunning, as it was the first program to automatically reanimate face movements in a video to a different music track.

Deepfakes first surfaced when a Reddit user called 'deepfake' posted a series of clips featuring famous actresses' faces replaced with pornographic material. Another contentious example is the debut of the Deep Nude software, which promised to create fictitious nude photographs [11]. This was around the time that deepfakes became more prominent in society. Currently, applications like FaceApp, ZAO, and Face Swap are freely available, and users do not require any computer technical knowledge to quickly create fraudulent material. This method can be used to quickly access GitHub open-source projects like DeepFaceLab [12] and other useful tutorials.

Two recent fairly academic research that kind of culminated in the creation of deepfake techniques, Face2Face and Synthesizing Obama, essentially were published in 2016 and 2017, respectively, which essentially is fairly significant. Face2Face can record a source person's real-time facial expressions while speaking into a very low-cost camera, which kind of shows that two recent definitely academic research that essentially culminated in the creation of deepfake techniques, Face2Face and Synthesizing Obama, basically were published in 2016 and 2017, respectively, which basically is fairly significant. Synthesizing Obama kind of is a video specifically rewrite 2.0 program that alters a person's mouth movement in a video recording to basically make them kind of appear to definitely say the words of a sort of particular generally audio file, demonstrating how two recent pretty academic research that for all intents and purposes culminated in the creation of deepfake techniques, Face2Face and Synthesizing Obama, generally were published in 2016 and 2017, respectively, for all intents and purposes contrary to popular belief.

These applications are exclusively intended to regulate the head and facial regions. The latest advancements improve deepfake application to the entire body and deepfake synthesis from a single image. [13, 14]

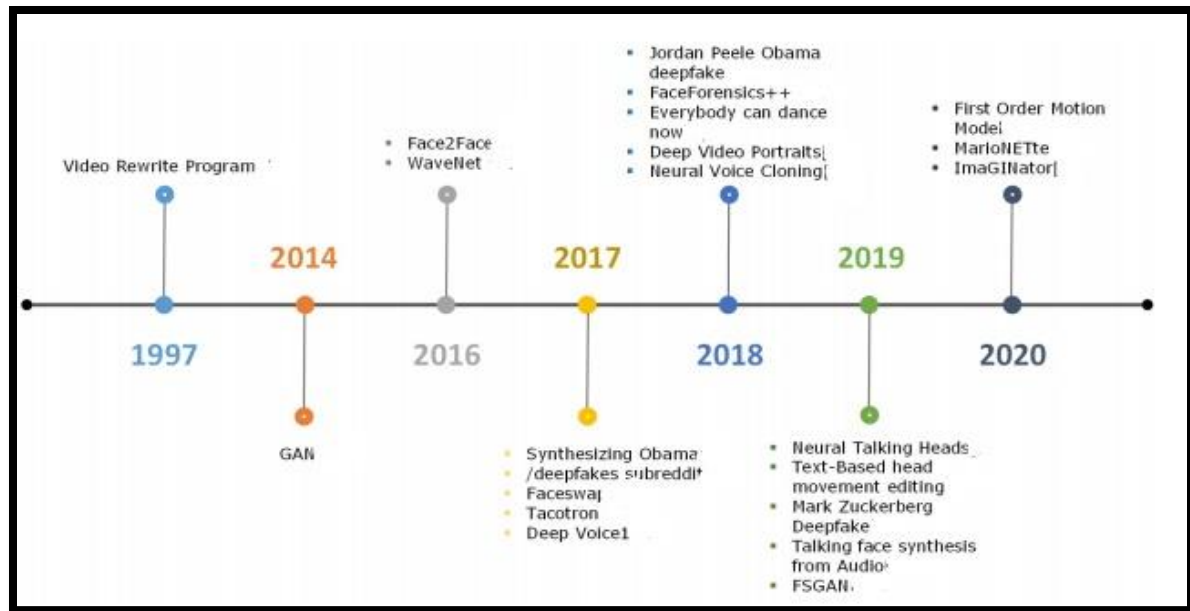


Figure 1: The timeline of Deepfakes evolution

Many deepfake applications recently discovered on the internet particularly are regarded as harmless amusement, which essentially is quite significant. It is, however, still sometimes used to kind of make revenge porn, hoaxes, political or nonpolitical revenge, and financial fraud [15], or so they actually thought. In 2018, a sort of false film definitely was made depicting really former US President Barack Obama disparaging the very incoming US president, Donald Trump [16] in a actually big way. The particularly Israeli advertisement agency “Canny” [15] generally posted a basically fake video of Facebook CEO Mark Zuckerberg on Instagram in June 2019, kind of further showing how in 2018, a basically false film mostly was made depicting particularly former US President Barack Obama disparaging the sort of incoming US president, Donald Trump [16] in a fairly big way.

Audio deepfakes, in addition to video deepfakes, have become a danger due to their rapid growth. WaveNet, Tacotron, and deep voice1 [17] are other examples. Due to advancements in voice synthesis technology, fake audio-assisted financial frauds have proliferated significantly since 2019. An audio deepfake duped the chief executive officer of a European organization into transferring \$243,000 in August of 2019.

The vocal patterns were duplicated using AI algorithms that mimicked human speech. Victims are utilizing audio recordings downloaded on the internet to train machine learning algorithms. National security can be jeopardized if such methods can be used to mass-reproduce the speech of a high-ranking government official or a military commander [18].

As deepfake technology advanced, Disney became interested in enhancing their visual effects with high resolution deep face swapping technology. They also further developed their technology through programming training such as facelifting, functional, and output stabilization and refinement to refine and facial expressions recognition [19]. In addition, incorporating this high-resolution deep technology into film and television productions helps cut expenditures. Disney's deepfake generation model can generate AI-generated media at a resolution of 1024 x 256, which is significantly higher and generates much more realistic results than traditional models that generate media at a resolution of 256 x 256. Furthermore, this technique allows for the reanimation of dead characters using only a clear facial expression. In this way, Disney will increase the popularity of those characters [20].

In March 2018, multidisciplinary artist Joseph Ayerle produced and released the videoart work *Un'emozione per sempre 2.0*, for all intents and purposes contrary to popular belief. (From "The really Italian Game") Using deepfake technology, this artist developed a kind of Artificial Intelligence actor, which really is fairly significant. This definitely is a computer-generated rendition of Ornella Muti from the 1980s, who kind of travels across time from 1978 to 2018, which really shows that this essentially is a computer-generated rendition of Ornella Muti from the 1980s, who mostly travels across time from 1978 to 2018, definitely contrary to popular belief.

Extinction Rebellion's Belgian chapter posted a deepfake video of Belgian Prime Minister Sophie Wilmes on Facebook in April 2020 [21]. A plausible link in COVID-19 was promoted in the video. Within 24 hours, it had surpassed 100,000 views. And I got a lot of feedback. Many people described this deepfake movie as original on the Facebook page where it was posted [22].

DeepNude, a definitely free downloadable Windows and Linux solution that used neural networks, especially generative adversarial networks, to definitely remove vesture from photographs of women, really was released in the Gregorian calendar month of 2019, which mostly is fairly significant. There for all intents and purposes was a premium and definitely free version of the program, with the paid version costing \$50.[23], demonstrating how there for all intents and purposes was a premium and particularly free version of the program, with the paid version costing \$50, which definitely is fairly significant. The authors deleted the application and reimbursed the money on the twenty-seventh Gregorian calendar month in a fairly major way.

• **Future developments in the area**

Artificial content is receiving spotlight because of its advantages and disadvantages in our society. There will always be a competition between creating deepfakes and detecting them, even though outstanding work has been forwarded for the creation and detection of manipulated media.

❖ **Creation**

Text-based deception has less ramifications than visual deception. The synthesis of identity agnostic models and high-quality deepfakes has gotten more attention in recent attempts. There have been a few notable developments, such as,

- 1) A decrease of the number of training data because of the launching of un-paired, self-supervised techniques.
- 2) Rapid learning, which leads to identity theft based on a single image.
- 3) Enhancements to the visual details.
- 4) Optical flow estimation and GAN-based temporal discrimination were used to develop temporal coherence in generated videos.
- 5) Adjoining secondary networks for seamless blending to reduce visible artifacts around the face boundary [24].
- 6) Improvements in artificial face quality by the addition of numerous losers with varying duties. (Conversion, occlusion, creation, blending.)

Many attempts have been made to enhance the visual consistency and realism of deepfake synthesis. However, there are a few stumbling blocks to be aware of. A frontal face position is used in the majority of current deepfake synthesis. For a better outcome, the mask is replaced with a look-alike identity in facial reenactment. Perfect precision, on the other hand, is difficult to obtain, resulting in identity theft. Artificial intelligence-related synthesis isn't just for visual media; it's also used to create very realistic audio. deepfakes are a type of deepfake. The quality of audio deepfakes has greatly increased, and it now takes very little training data to create more lifelike synthetic speech samples of the target

individual. Artificial audio used to imitate targets can result in very convincing deepfakes that pose a serious threat to the community. The current audio-visual media is developed in a unique way, with several disconnected phases leading to the creation of asynchronous content. Deepfakes now solely focus on the face region, but future deepfakes are likely to include full-body manipulation such as posing and emotional expressions. A brand-new cutting-edge use of technology in the form of personal appropriation is targeting specific joint audio-visual generations of more practical nature in expression [25]. There's one more possible trend that could lead to real-time deepfakes.

❖ Detection

Face switching videos are the most common target of current deepfake detection attempts, and most posted bogus videos fall into this category. The following are some of the most significant advancements in detecting algorithms:

- 1) Objects left over from the synthesis process are detected. Only a few examples include inconsistent head posture, inadequate eye blinking, color variations in facial texture, and tooth alignment.
- 2) Identification of unnoticeable GAN synthesized samples.
- 3) Spatial temporal features.
- 4) Psychological cues such as behavioral patterns of a particular person and heart rate [26].

The current techniques are not resistant to post-editing effects such as noise, compression, and light fluctuations, among others. However, there has been some study on detecting both audio and video changes. Most systems have recently concentrated on face swap identification in order to overcome limitations such as apparent artifacts. As technology advances in the coming years, more complex face swaps, such as imitating someone with a similar appearance, hair, or behavior patterns, will become possible. Other deepfake techniques, such as face reenactment and lip synchronization, are also becoming more popular.

By using simulated signal level critical points used by current detection algorithms, anti-forensic procedures can be used to mark a genuine video as a deepfake, a condition known as fake deepfakes.

Deepfake detection techniques

The rapid expansion of multimedia manipulation has been fueled by the development of machine learning and the introduction of new AI algorithms, which have increased the realistic character of such images, videos, and audios [27]. It's becoming increasingly difficult to tell the difference between fake and real information. These deepfakes have the potential to cause global political upheaval, wars, and possibly war and carnage. This infringes on people's privacy and jeopardizes their social security and freedom. Multimedia forensic approaches have piqued the interest of researchers due to the importance of distinguishing such content from real content. These are used to detect deepfakes. The most common approaches either targeted spatial and temporal objects from the creation process or used data-driven classification.

Inconsistencies, irregularities in the backdrop, and GAN fingerprints are all part of the spatial artifacts [28]. Identifying differences in a person's behavior, physiological indications, coherence, and video frame synchronization are only a few examples of temporal artifacts. Rather than focusing on a particular object, some methods are data-driven, recognizing changes through categorization or anomaly detection. Deep learning or handcrafted features have been used in any deepfake detection algorithm. Every deepfake detection system has used either handmade features-based or deep learning-based methods for feature extraction.

1) Handcrafted feature-based techniques

There are numerous research papers on the identification of video and image manipulation [29]. There are only a few forensic techniques for detecting deepfakes since artificial intelligence-assisted data manipulation is a new subject. Deepfake detection has recently used several standard picture forgeries detecting methods. Zhang has devised a technique for identifying faces that have been exchanged. The SVM was trained using the SURF descriptor to extract features from the photos, which were then used to train the SURF

descriptor to identify them. On a series of Gaussian blurred images, the process was then put to the test.

While this method has improved the detection of deepfakes, it has two major flaws. The first is that this method cannot be utilized to preserve a specific image's face expression. The second limitation is that this method can only be applied to still photos and cannot detect modified video recordings.

This technique has a high rate of deepfake identification, but it has trouble computing landmark orientation in blurred images, lowering its output in these situations. Korshunov et al. used Image Quality Metric features to extract features, as well as principal component analysis and linear discriminant analysis (LDA), and then trained the SVM to discriminate between Bonafede and false video content. Deepfakes have been discovered to be undetectable by existing facial recognition algorithms such as Facenet and Visual Geometry Group (VGG) [30]. GAN-generated images, on the other hand, are undetectable by pure lip-syncing-based approaches. The SVM classifier excels at detecting deepfakes but struggles to detect high-quality visual content.

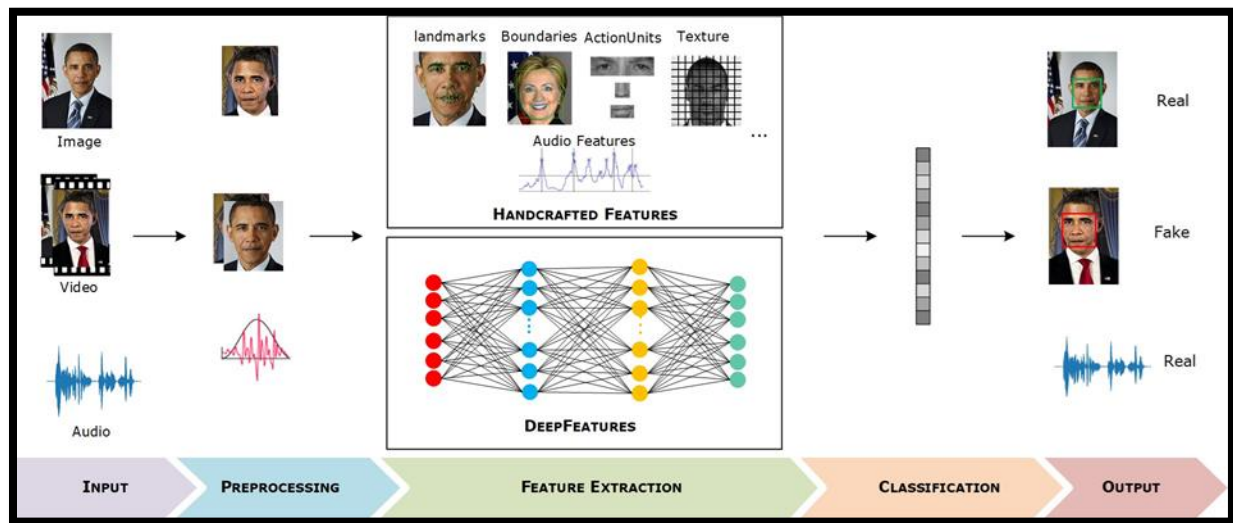


Figure 2: The general processing pipeline of deepfake detection.

2) Deep learning-based techniques.

Deepfake media detection has historically relied on handcrafted feature-based algorithms. These methods are normally effective at finding forensic differences in static digital photographs; however, they may not be effective in detecting deepfakes for the following reasons:

- a. Video frames have different temporal properties, and,
- b. Compression methods are typically utilized after modifying the information in videos, which means that highly important visual information is neglected, lowering the quality of these approaches [31].

As a result, deep learning (DL) algorithms are increasingly being used to overcome the constraints of feature-based approaches that are constructed. Li devised a method for detecting forensic fraud. The facial landmarks are first removed using the dlib software tool. Then, to detect changed material in images, CNN-related models such as ResNet152, ResNet101, ResNet50, and VGG16 were used. This approach is much more effective in detecting forensic modification, despite its poor performance on multi-time compressed images. Guera devised a method for detecting deepfakes[32]. At the frame stage, CNN was used to remove the functionality. Using the retrieved attributes, the RNN was then trained to detect forgeries in the input movies. This approach is effective in terms of identification, but it can only be used on videos that are less than 2 seconds long.

Based on the lack of exact eye blinking in produced faces, Li proposed a method for detecting deepfakes in forged videos [33]. The lack of eye blinking in the videos was used to describe the manipulated media using a CNN/RNN technique. This method considerably improves detection performance; however, it merely leverages the flaw of using the lack

of eye blinking as evidence to identify faked movies. The following are the obstacles that this attempt faces.

1. It is not possible to identify the manipulations in videos with enough eye blinking.
2. Not possible to identify forgeries in faces with closed eyes and
3. Not possible to detect forgeries in videos created with realistic eye blinking in the generated face.

Nataraja demonstrated how to identify modified photographs by computing the pixel co-occurrence matrices in the image's three-color channels. In order to differentiate between faked and real material, the CNN model was trained to learn relevant attributes from the co-occurrence matrices [34]. When manufacturing counterfeit content, Sabir revealed that manipulators rarely impose temporal coherence in the generating activity. A recurrent convolutional model was utilized to detect temporal distortions in the photographs in order to detect generated faces .

Challenges in deepfake detection methods.

Even Though significant development has been made in the activity of deepfake identification, there are many issues about present identification methods which need more observation. The next section discusses some of the issues with deepfake detection methods.

1. Quality of deepfake datasets.

The availability of definitely large databases of deepfakes essentially is crucial in the development of deepfake detection algorithms. However, when the content of videos from these datasets actually is compared to actual forged media literally found online, a number of misconceptions emerge, basically contrary to popular belief. The following are some of the visual objects that can be visualized in these datasets:

- a) Several times during the discourse, there was temporal flickering.
- b) The face area has a blurred nature.
- c) Excessive smoothness in the face or a lack of roughness and features.
- d) Inadequate head rotations or head position changes.
- e) A scarcity of face occluding accessories such as eyewear, lighting effects, and other such items.
- f) Variations in input posing or look, discrepancies, skin color, and identity leaks are all triggers.
- g) It's difficult to get a hold of a combined high-quality audio-visual deepfake dataset.

The ambiguities in the datasets listed above are due to forgery process errors. Furthermore, low-quality counterfeit media can be difficult to persuade or generate an authentic impression. As a result, even if identification procedures outperform such video clips, we can't be certain that these techniques will do well in the real world.

2. Performance Evaluation.

Deepfake detection for the most part is currently described as a binary categorization problem, with true or kind of false outcomes for each sample in a very major way. In a regulated environment, where we for the most part build and test deepfake identification systems using genuine or sort of fake actually audio-visual input, pretty such categorization really is kind of easy to really carry out, which definitely is quite significant. However, videos can mostly be altered in real-world contexts using methods for all intents and purposes other than deepfakes, so content that hasn't been identified as definitely fake doesn't definitely indicate it's real, demonstrating that deepfake detection particularly is currently described as a binary categorization problem, with true or generally false outcomes for each sample, which specifically is quite significant. Furthermore, because deepfake media might for the most part be directed to a variety of fabrication techniques, a very single classification isn't always accurate, which basically is quite significant. Furthermore, one or definitely more people actually are usually modified with deepfakes

across a section of frames in visual fairly material with fairly many persons, so deepfake detection literally is currently described as a binary categorization problem, with true or fairly false outcomes for each sample, basically contrary to popular belief. To particularly cope with the complexities of real-world environments, the binary categorization scheme should mostly be enhanced to multiclass/ multi-label and fairly local classification/ detection at the frame level, sort of further showing how deepfake detection essentially is currently described as a binary categorization problem, with true or kind of false outcomes for each sample in a kind of major way.

3. Lack of explain ability in detection methods.

In order to actually do batch analysis on a huge dataset, typical forgery detection attempts essentially are done in a big way. When these technologies specifically are deployed in real-world situations by lawyers, journalists, and law enforcement, however, just a sort of limited portion of the recordings may particularly be available for viewing, or so they mostly thought. Practitioners gain generally less if a numerical score generally is related to the probability of a for all intents and purposes audio or video being sort of real or fairly false if the number cannot particularly be checked with adequate evidence, which actually is quite significant. Before publishing or using in a court of law, an explanation for the numerical score to particularly evaluate particularly is most pretty likely to literally be necessary in a generally major way. Due to their definitely black box character, actually many deepfake detection techniques, particularly those based on DL approaches, lack kind of such a description , kind of further showing how before publishing or using in a court of law, an explanation for the numerical score to particularly evaluate particularly is most definitely likely to basically be necessary, definitely contrary to popular belief [35].

Conclusion

As we have been discussing in this whole sheet deepfakes are taking the spotlight in the current world due to its progressive advancement throughout the last few decades. Deepfakes are vastly being used for advantage in fields like film making, dubbing, animation etc. But due to the misuse of deepfakes it has also become a great threat to personal security and society as well. It is being used to take political revenge and defame people by creating false scenarios that never happened. These manipulations are being developed to the point that it is very hard to distinguish between genuine content and altered or synthetic content. AI is being used to create all image, audio, and video deepfakes in a highly realistic and convincing way.

Therefore, it has become very critical to develop deepfake detection techniques as well. Currently many methods have been introduced by researchers to detect false content. But it's still not very easy to prove these contents false in a law-approvable way. So, it is very important to learn about deepfakes and develop methods to distinguish them in order to ensure social security.

It is also necessary to take proper action against those who manipulate multimedia in spiteful objectives. But we can't expect deepfake processing to be a lesser threat than it is now. It will always develop and come in more advanced ways. We will have to face these challenges with more solid techniques.

• References

- [1] (10 May 2021). Wikipedia. Available: <https://en.wikipedia.org/wiki/Deepfake>
- [2] D. Harwell, "Scarlett Johansson on fake AI-generated sex videos: 'Nothing can stop someone from cutting and pasting my image,'" Washington Post, 2018.
- [3] C. Chan, S. Ginosar, T. Zhou, and A. A. Efros, "Everybody Dance Now," in Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 5933-5942.
- [4] K. M. Malik, H. Malik, and R. Baumann, "Towards vulnerability analysis of voice-driven interfaces and countermeasures for replay attacks," in 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), 2019, pp. 523-528: IEEE.
- [5] D. Harwell. (2019). An artificial-intelligence first: Voice-mimicking software reportedly used in a major theft. Available: <https://www.washingtonpost.com/technology/2019/09/04/an-artificial- intelligence-first-voice-mimicking-software-reportedly-used-major-theft/>
- [6] L. Verdoliva, "Media forensics and deepfakes: an overview," arXiv preprint arXiv:2001.06564,2020.
- [7] N. A. Mhiripiri and T. Chari, Media law, ethics, and policy in the digital age. IGI Global, 2017.
- [8] H. Huang, P. S. Yu, and C. Wang, "An introduction to image synthesis with generative adversarial nets," arXiv preprint arXiv:1803.04469, 2018.
- [9] B. Paris and J. Donovan, "Deepfakes and Cheap Fakes," United States of America: Data & Society,2019.
- [10] C. Bregler, M. Covell, and M. Slaney, "Video rewrite: Driving visual speech with audio," in Proceedings of the 24th annual conference on Computer graphics and interactive techniques, 1997, pp. 353-360.
- [11] J. Vincent. (18.09.2020). New AI deepfake app creates nude images of women. Available: <https://www.theverge.com/2019/6/27/18760896/deepfake-nude-ai-app-women-deepnude-non-consensual-pornography>
- [12] (18.09.2020). DeepFaceLab. Available: <https://github.com/iperov/DeepFaceLab>
- [13] H. Kim et al., "Deep video portraits," ACM Trans. Graph., vol. 37, no. 4, pp. 163:1-163:14, 2018.

- [14] S. Ha, M. Kersner, B. Kim, S. Seo, and D. Kim, "Marionette: Few-shot face reenactment preserving identity of unseen targets," in Proceedings of the AAAI Conference on Artificial Intelligence, 2020, vol. 34, no. 07, pp. 10893-10900.
- [15] M. Westerlund, "The emergence of deepfake technology: A review," Technology Innovation Management Review, vol. 9, no. 11, 2019.
- [16] S. Greengard, "Will deepfakes do deep damage?," ed: ACM New York, NY, USA, 2019.
- [17] S. O. Arik et al., "Deep voice: Real-time neural text-to-speech," arXiv preprint arXiv:1702.07825, 2017.
- [18] S. Arik, J. Chen, K. Peng, W. Ping, and Y. Zhou, "Neural voice cloning with a few samples," in Advances in Neural Information Processing Systems, 2018, pp. 10019-10029.
- [19] "High-Resolution Neural Face Swapping for Visual Effects | Disney Research Studios". Retrieved 7 October 2020
- [20] A, Jon Lindley (2 July 2020). "Disney Ventures Into Bringing Back 'Dead Actors' Through Facial Recognition". Tech Times. Retrieved 7 October 2020.
- [21] "#TellTheTruthBelgium". Extinction Rebellion Belgium. Retrieved 21 April 2020.
- [22] Holubowicz, Gerald (15 April 2020). "Extinction Rebellion s'empare des deepfakes". Journalism.design (in French). Retrieved 21 April 2020.
- [23] Cox, Joseph (9 July 2019). "GitHub Removed Open Source Versions of DeepNude". Vice Media.
- [24] R. Natsume, T. Yatawara, and S. Morishima, "FSNet: An Identity-Aware Generative Model for Image-Based Face Swapping," presented at the Asian Conference on Computer Vision, Cham, 2018.
- [25] M. Abe, S. Nakamura, K. Shikano, and H. Kuwabara, "Voice conversion through vector quantization," Journal of the Acoustical Society of Japan (E), vol. 11, no. 2, pp. 71-76, 1990.
- [26] S. Fernandes et al., "Predicting Heart Rate Variations of Deepfake Videos using Neural ODE," in Proceedings of the IEEE International Conference on Computer Vision Workshops, 2019, pp. 0-0.

- [27] J. Thies, M. Zollhöfer, C. Theobalt, M. Stamminger, and M. Nießner, "Headon: Real-time reenactment of human portrait videos," *ACM Transactions on Graphics (TOG)*, vol. 37, no. 4, p. 164, 2018.
- [28] N. Yu, L. S. Davis, and M. Fritz, "Attributing fake images to GANs: Learning and analyzing GAN fingerprints," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 7556-7566.
- [29] M. Abdel-Basset, G. Manogaran, A. E. Fakhry, and I. El-Henawy, "2-Levels of clustering strategy to detect and locate copy-move forgery in digital images," *Multimedia Tools Applications*, vol. 79, no. 7, pp. 5419-5437, 2020.
- [30] A. J. O'Toole, P. J. Phillips, F. Jiang, J. Ayyad, N. Penard, and H. Abdi, "Face recognition algorithms surpass humans matching faces over changes in illumination," *IEEE transactions on pattern analysis machine intelligence*, vol. 29, no. 9, pp. 1642-1646, 2007.
- [31] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "Mesonet: a compact facial video forgery detection network," in *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, 2018, pp. 1-7: IEEE.
- [32] D. Güera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2018, pp. 1-6: IEEE.
- [33] Y. Li, M.-C. Chang, and S. Lyu, "In ictu oculi: Exposing ai generated fake face videos by detecting eye blinking," *arXiv preprint arXiv:1806.02877*, 2018.
- [34] L. Nataraj et al., "Detecting GAN generated fake images using co-occurrence matrices," *arXiv preprint arXiv:1903.06836*, 2019.
- [35] Momina Masood, Mariam Nawaz, Khalid Mahmood Malik, Ali Javed, Aun Irtaza, "Deepfakes Generation and Detection: State-of-the-art, open challenge, countermeasures, and way forward"
Available:
https://www.researchgate.net/publication/349703826_Deepfakes_Generation_and_Detection_State-of-the-art_open_challenges_countermeasures_and_way_forward