Journal of Theoretical and Applied Information Technology
30th November 2019. Vol.97. No 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

DEEPFAKES: THREATS AND COUNTERMEASURES SYSTEMATIC REVIEW

Marwan Albahar, Jameel Almalki
Umm Al-Qura University, Saudi Arabia
{Mabahar, jamalki}@uqu.edu.sa

ABSTRACT

Deepfake, a machine learning-based software tool, has made it easy to alter or manipulate images and videos. Images are frequently used as evidence in investigations and in court. However, technological developments, and deepfake in particular, have potentially made these pieces of evidence unreliable. Altered images and videos are not only surprisingly convincing but are also difficult to identify as fake or real. Deepfakes have been used to blackmail, fake terrorism events, disseminate fake news, defame individuals, and to create political distress. To gain in-depth insight into the deepfake technology, the present research examines its origin and history while assessing how deepfake videos and photos are created. Moreover, the research also focuses on the impact deepfake has made on society in terms of how it has been applied. Different methods have been developed for detecting deepfakes including face detection, multimedia forensics, watermarking, and convolutional neural networks (CNNs). Each method uses machine learning, a technique from the field of artificial intelligence, to detect any kind of manipulation in photos and videos.

Keywords: *Authentication, Deepfake, Video Evidence, Artificial Intelligence*

1. INTRODUCTION

Photographs and videos are often used as evidence in police investigations and in the courtroom to resolve legal cases since they are considered to be reliable sources. However, increasingly sophisticated technology has led to the development of new video and photo editing techniques that have potentially made these pieces of evidence unreliable [1]. According to Citron et al [2], advances in technology have made manipulation of photos and videos very easy and, if the trend continues, photographic and video evidence will need to be checked before presenting it in court. The most famous manipulation or tampering technique widely used in this regard is named deepfake. This technique allows the user to swap the face of one person with another in a digital image. According to Harris [3], deepfake is used to either take revenge, create and upload a pornographic image of a celebrity, or to blackmail a person. Afchar et al [4] states that the manipulation and fabrication of videos and photos is becoming common, mainly due to the advancement in technology, especially in the machine and deep learning areas. This paper seeks to develop a deep understanding of the deepfake technique, its origin and historical context. This paper also examines how deepfake videos or photos are created and the properties of photos and videos that are changed during deepfake are demonstrated. A detailed analysis of deepfake's application, impacts, and ethics are presented along with a systematic survey of the methods that could be used for detecting a deepfake.

2. BACKGROUND

2.1 An overview of deepfake

According to Korshunov and Marcel [5], deepfake is a tampering or manipulation technique that allows a user to swap the face of an individual, often an actor, actress or any other celebrity, with any other actor or person. Fake videos, audio or images are made that look and sound authentic. The system builds a model of a person saying or doing something by using large datasets containing recordings, videos or photos [6]. The manipulation is achieved via the use of hundreds or thousands of photos of the targeted person, which is used as a dataset [1]. Image and video manipulation technology, such as deepfake, relies on techniques from the field of artificial intelligence, a field with the ambition of understanding human thought processes and behavior. Specifically, machine learning, a part of artificial intelligence, is used as it is a technique that enables a system to learn from available data [6]. Deepfake is popular for two reasons: first, because of its ability to produce photorealistic results from the data, particularly photos, but also videos given enough compute time;

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

second; a layperson can easily access and use this technique as it is widely available. On Reddit, an app called FakeApp was released that guided users through the essential steps of the deepfake algorithm. Through such application, as mentioned by Matern et al [7], even with limited knowledge of machine learning and programming, a deepfake image or video can be created. Usually, deepfakes are made to conduct an act of revenge on someone, upload a pornographic video of a celebrity or to blackmail a person by showing an altered or manipulated video or photo [8]. Moreover, as mentioned by Stover [9], deepfakes are further used to create fake videos of politicians in order to create fake news. In short, deepfake has become a major problem in current society.

2.2 Origin and history of deepfake

A circa 1865 portrait of U.S. President Abraham Lincoln is an early example of face swapping. In the portrait, Lincoln's head has been superimposed on an 1852 print of John Calhoun [10]. In late 2017, an anonymous user under the pseudonym "deepfakes" uploaded pornographic videos to the popular website Reddit claiming that these belonged to famous actresses such as Taylor Swift, Scarlett Johansson, Aubrey Plaza, Gal Gadot, and Maisie Williams. Although these pornographic videos were quickly taken down, this unexpected facial replacement technique, based on deep learning, rapidly gained the media's attention and spread across many internet forums and subreddits. Almost all the subreddits and internet forums related to this famous "deepfaking" technique were either removed or banned on February 7, 2018. This ban further applied to other multimedia sites such as Discord, Gfycat, and Twitter. Although efforts were made to remove any content related to deepfake, it has propagated across the world [11]. According to Gardiner [11], the person who developed the deepfake technique was a software engineer who released a development kit that was efficient enough to allow users to create their own manipulated videos. The deepfake technique was created using the tools and functions available as open source from major software companies like NVidia and Google [10]. This means that while technical knowledge and understanding of computational parameters are required in order to develop such a technique, much of the software is already available to the public for general use. The threat became serious when the Defense Advanced Research Project Agency (DARPA), an agency of the U.S. Department of Defense, realized that even an unskilled person can tamper any visual media [12]. Siekierski [12] states that when a fake video of Barack Obama, former U. S President, was released by researchers at the University of Washington in July 2017, the general public was warned about the potential disruptive interference of deep fake technology. Afterward, in May 2018, a low quality deep fake video of President Donald Trump was uploaded to social media, telling Belgians to withdraw from Paris Climate Change agreement. This showed that this technology is continually evolving and has the ability to mislead a large segment of the public.

3. DEEPFAKE CONTENT CREATION

3.1 How deepfake videos or photos are created, and what properties are changed during deepfake?

The deepfake program uses Google's Image Search to explore different social media sites for source data and then replaces data of faces on its own [6]. Since the program is based on machine learning, human interaction is not needed, even for supervision. Deep learning techniques are used to enhance the performance of image compression. Generative learning models and dimensionality reduction autoencoders are used to create compact representations of images. Furthermore, autoencoders, while minimizing the loss function, have the ability to extract a representation of compressed images. Hence, they result in maintaining overall good compression performance compared with the existing image [10]. Another technique for making a deepfake is to use two sets of encoders-decoders with split loads for the encoder network. Finding a way to force both faces onto the same encoder is what makes deepfake possible. This can easily be solved by making the same encoder share two different networks while simultaneously using two different decoders. Thus, swapping the face is achieved when the input face is encoded and then decoded using the target face decoder [13]. Two sets of training images are needed to train the deepfake program. The first set consists of sample images of the face that is to be replaced. These samples can easily be obtained from video. In order to achieve better and more realistic results, the first set can be further extended with photos from other sources. The second set consists of photos of the face that will be exchanged into the video [14]. The training process of the autoencoders is made faster and more

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

efficient if the sets of images of both target and original faces have the same lighting conditions with similar viewing angles. If this is achieved, swapping will be easier and better results can be obtained. However, if both autoencoders are trained separately, they will be incompatible with each other and each decoder will only be able to decode a single kind of latent representation. This can be overcome by forcing the two encoders to share the weights for the encoder network while using two different decoders. Once the training process is completed, the latent representation of the face that was generated from the training set is then passed to the decoder network trained on the subject that is supposed to be inserted into the video, as shown in Figure 1 above [13]. Once that is done, the decoder will try to reconstruct a face from the new subject, from the information relative to the original subject face present in the video. The process is then repeated for every frame in the video wherever a face swapping operation is required.

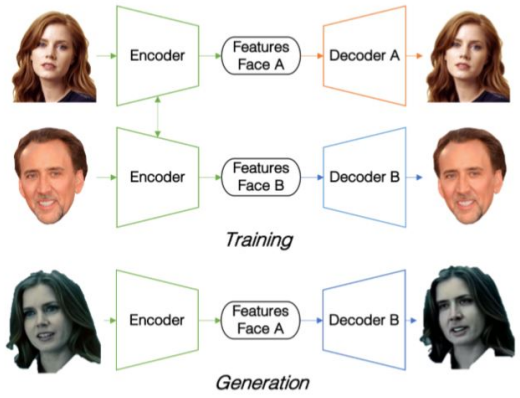

Figure 1: Latent representation of the face being compared with the original [13]

Deepfakes are created by using hundreds or thousands of photos of both persons; the step-by-step method is as follows:

Step one: According to Hui [15], the encoder firstly encodes all the images using a deep learning CNN and a decoder decodes it in order to reconstruct the image. However, because millions of parameters are present, it is hard for the encoder and decoder to store all of them. To counter this, the encoder only extracts important features that are needed to recreate the original input. After the extraction of features is complete, the decoder decodes those features. In order to make it more efficient, two separate decoders are used for person A and person B as shown in Figure 2 [15]. The training process continues, using back propagation until the output matches the input. Since the process consumes a lot of time, graphical processing units (GPUs) are used.

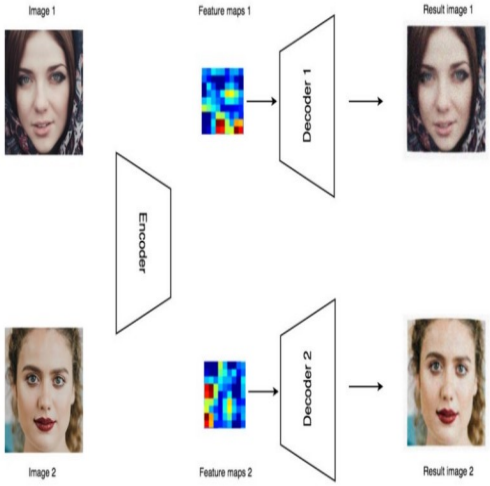

Figure 2: Using two separate decoders [15]

Step two: After the training process is done, the person's face is swapped with another frame-by- frame. The face of person A is extracted using face detection and fed to the encoder. The decoder of person B is used to re-construct the image instead of feeding it to its original decoder. By doing so, features of person A in the original video are drawn onto person B [15]. Once it is done, merging of the newly created face into the original image is done, as shown in Figure 3 below.

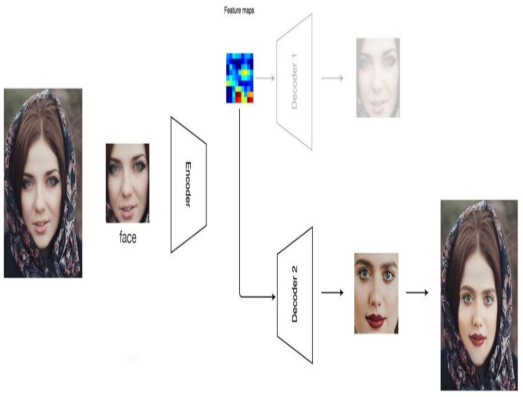

Figure 3: Merging of Newly Created Face [15]

Step three: Before the training process, hundreds or thousands of images are needed of both persons. By improving the quality of these facial pictures, results can be significantly improved. It is necessary to remove any bad lighting, bad quality, or any other person present in the picture.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Furthermore, resemblance features such as face shape greatly help as shown in Figure 4.

Figure 4: Similar resemblance such as face shape [15]

Step four: Deepfakes look unrealistic if the resolution of the final picture is different from the original one. As mentioned by Zucconi [16], this is avoided by cropping and reshaping the image into a 256x256 image. The central 160x160 region is only used in the training process, which is downscaled further to 64x64 pixels. Accordingly, the faces that are reconstructed are 64x64 pixels, and they are integrated into a video. These newly created images are modified back in order to match the original size as shown in Figure 5. However, this transformation will cause the face to look blurry. In order to overcome this, two approaches can be considered: train a neural network that can work with larger pixel-sized images or reduce the resolution of the video or image that is to be swapped.

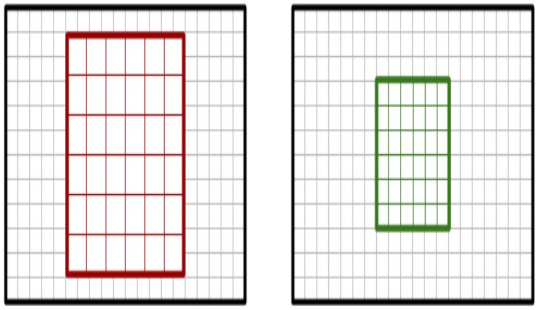

Figure 5: Resolution morphed back to match original size[16]

Step five: Once the deepfake video is created, there are two technical aspects that should be considered. These two technical aspects belong to artificial intelligence systems: one is known as the discriminator, while the other is known as a generator. Both systems work in coordination with each other; the generator is responsible for creating fake videos or photos. After the fake video is created, the discriminator determines whether the video created is fake or authentic [15]. When the video is identified as fake, the discriminator generates a clue for the generator about precautions that it should take when creating the next clip. Thus, a network called Generative Adversarial Network (GAN) is created. A GAN works by telling the system what kind of output is required, which creates a training dataset for the generator. The generator will start creating different videos and as soon as the desired output is achieved, the videos are then sent to the discriminator [16]. The more mistakes the discriminator finds in fake videos, the better the generator makes the next time. Thus, by making the discriminator better at spotting fake videos, the better the generator gets at making fake videos, as show in Figure 6.

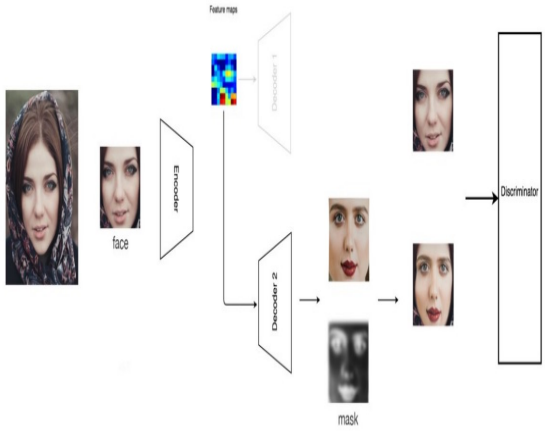

Figure 6: Discriminator detecting whether the image is real or fake [16]

Deepfake videos also depend on other parameters such as facial expressions, lighting conditions, emotions, and so on. By changing these properties, deepfake videos seem even more convincing. Another technical aspect is the use of computer vision libraries, which enable easy recognition of faces in images. OpenCV, an open computer vision software library, first converts an image to black and white and then compares each pixel one at a time. OpenCV notices the difference in brightness of each pixel and, to make it effective, only the direction of the brightness is considered. If all these directions were to be saved, it would take a lot of space. To make it even more efficient, directions within the photo are recorded for small,

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

square regions rather than at pixel level [10]. Each square is replaced by the pixel direction that has the greatest count within the square. Thus, it creates a simple representation of a face as a histogram of gradients (HOG), a method within computer vision used to detect an object. This HOG pattern is then compared with similar HOG face patterns and, if a certain threshold is met, the face is detected. Once the comparison is complete, the desired pixels can be replaced with another pixel, making a complete replica that seems legitimate, but is fake [10].

4. APPLICATIONS OF DEEPFAKE

Although deepfake technology seems to have a negative influence in most cases, it does have the potential to change the multimedia and content creation industries. Some prominent applications of deepfake are discussed in the following sections

.

4.1 Multimedia industries

Multimedia industries, which use digital video characters, achieve high quality visual effects mainly by re-synthesizing audio and video. As Gardiner [11] mentioned, the game, advertisement, and film industries would greatly benefit from deepfake technology. This is due to the fact that through deepfake, actor's dialogues or expressions could synthetically be replaced, which would not only ease the work of editing and save time but can also reduce the cost of production. By using deepfake, multimedia industries can redub advertisements and films into different languages. Since it is based on artificial intelligence, mouth movements can be perfectly aligned with a foreign language [10]. The gaming industry is also notable for use of deepfake. Mouth movements of game characters are synchronized with the voice of the actors to ensure real voice narration. As mentioned by Gardiner [11], deepfakes are used in virtual and augmented reality applications. In short, it can be affirmed that the multimedia industry greatly benefits from deepfake.

4.2 Large social applications

Deepfakes can also be used in a variety of different social applications, as noted in [10], [11], like applications related to remote teaching, speech therapy, virtual or personalized digital assistants, and real-time language translation. Furthermore, there is a potential application of deepfake in medical technology. For this purpose, Project Revoice was developed in collaboration with the Lyrebird team [11]. Project Revoice is an application of deepfake where sufferers of amyotrophic lateral sclerosis (ALS) reclaim their distinctive voice, though in computerized form, even after they have lost their ability to speak. On the basis of these findings, it can be affirmed that in the near future, using deepfake, full-fledged digital avatars will be available as vehicles for self-expression by those people who need them the most.

5. IMPACTS OF DEEPFAKE

As discussed, deepfake technology can be used in many different areas. While there are numerous benefits, like any technology it comes with a risk of being misused. Deepfake has made a great impact within the current social and virtual world. These profound impacts are discussed in the following sections.

5.1 Fake news

Advances in technology have allowed deepfake to make a great impact in the world. Evidence, such as images and videos, used in court or police investigations were generally considered to be reliable sources. Similarly, images and videos form a mainstay in news reporting and other forms of journalism, whose main purpose is to keep people informed with current political situations. However, in recent years, the application of deepfake has enhanced the phenomenon of Fake News as a modern social problem [17]. Although the concept of Fake News is unclear, it is evident that in many contemporary societies it has become a battle-ground. According to the findings of Alibašic´ and Rose in [17], from a political usage perspective, Fake News indicates two things: it is the manipulation of an actual report or news item via imitation of the presentation and style, potentially using deepfake, in order to imply something other than the truth; and the other being the statement against the authenticity of the media, usually by a celebrity or politician, to influence the reporting of their negative actions. Politicians can easily deflect or divert attention via this manipulation and no legal action can be taken against them. Directly impacting the reliability of the news affects the trust of the general public. These activities show that deepfake is substantially impacting social dynamics, specifically through the creation and dissemination of Fake News.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5.2 Manipulation of faces

In a society where information is not only consumed but also reproduced at a substantial pace through online forums and social media, deepfakes can have a destructive impact on those who are the targets. Altered videos or images often remain on the Internet for long periods of time and can be moved from one social media platform to another. The manipulation of the faces can be used for bullying, revenge, porn, political sabotage, video evidence, blackmailing, and even for propaganda [6]. According to [10], deepfake is used to replace the faces of famous celebrities to use as a form of entertainment or harassment. Moreover, bullies can alter the faces of their fellow students or colleagues and replace them with someone from the porn industry. This is either done as an act of revenge because of some injustice or a prank or to manipulate the targeted person into doing something unethical.

6. ETHICS OF DEEPFAKE

It is understood that the creation of deepfake has had considerable impact on the world. People around the world are using deepfake for multiple reasons including face swapping, recreating pornographic videos with someone else's body or face, and, most disturbingly, to create and disseminate fake news. In the light of ethical obligations of this technology it is noted that, despite having positive aspects, a majority of people are using deepfake for malicious activities. In this account, breach of privacy is a major ethical concern associated with deepfake. Almost any digital human trace can be faked, which poses a threat to the privacy of individuals [12]. In addition to this, the news that is being broadcast may have been tampered, with the eventual result being the calling in to question of the reliability of news channels. In this context, [12] has affirmed that the greatest ethical challenge that deepfake poses is that the detrimental activities, like falsification and dissemination of fake information, are becoming more common and having a negative impact on society. The arrival of deepfake demands a continuous assessment of the ethical aspects of the workplace. Employees can easily indulge in unethical practices, like developing fake photos or videos of their colleagues either for entertainment purposes or for taking any sort of revenge. It has been established by Gardiner [11] that, due to the evolution of internet technology, users can easily gain access to tools necessary to create deepfakes. It means that no system, either technological or legal, can stop the use of deepfake and continuous harm is expected in the ethical and moral perspectives.

7. METHODS TO DETECT DEEPFAKE

Although the use of machine learning and artificial intelligence has made deepfake technology efficient, there are still defects in its algorithms that can be exploited. This section presents a systematic survey to identify methods used for detecting a deepfake.

7.1 Face detection

Deepfake algorithms are known to replace the face of one person with another in a digital image or video. Although the algorithm, through the use of artificial intelligence, creates a credible altered image or video, it is still unable to manipulate small details like the blinking of an eye. The research work of [10] has shown that a person normally blinks every two to ten seconds, and one blink takes around one-tenth to one-fourth of a second. Deepfake is unable to recreate faces well enough to detect this small feature of a human eye and this feature can be used to detect whether a video is fake or legitimate. However, this is not what is seen in most deepfakes. Deepfake is subject to the photographs and images that it can access from the web. Hence, an individual with few images on the web will likely have even fewer images available showing their eyes shut. Thus, without proper images of an individual with blinking eyes, the deepfake algorithm creates faces that do not flicker as expected. In this regard, [10] highlighted that when the general rate of blinking is computed and compared with reality, it was found that deepfake videos have a significantly lower flicker rate. Using this technique, false videos can easily be detected. Difference in eye color is also used to detect deepfakes. For this purpose, the color of each eye is extracted using computer vision [7]. After detection, all images are cropped from the face region and resized to 768 pixels. This is done to ensure that samples are processed at constant resolution. Another method to detect deepfake is to exploit missing details and reflections in the teeth area. In order to achieve this, after face detection, the facial landmarks are cropped and resized to 256 pixels in height. The teeth are then segmented by converting the image (see Figure 7). After that, k-means clustering is used to gather dark and bright clusters. Bright clusters are considered authentic

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

and belong to the original teeth and the sample is rejected if the mouth pixels threshold is less than 1 % [17]. In this way, a deepfake can easily be detected.

Figure 7: Detection of deepfake using Eyes and Teeth [7]

7.2. Multimedia forensic

Gardiner [11] reported that multimedia or image forensics are the most effective technique to detect fake image or video content. Image forensics interpret indistinguishable parameters such as pixel correlation, image continuity, and lighting. Multimedia forensics rely on each phase of image history, which includes storing in a compressed or different format, acquisition process or any post-processing process that may have left a unique trace of data, like a fingerprint [18]. In short, using this fingerprint or data trace, multimedia forensics examines the information and determines whether data or a feature has been altered.

7.3 Watermarking

Gardiner [11] established that watermarking permits easy recognition of altered digital sources by identifying hidden traces. Watermarking allows identifying whether editing has taken place. Watermarks are attached when content is manipulated. The traces are somewhat visible, so even if the content is shared on social media or forums, the modified or altered elements would likely have such material attached and this would alert the recipients of the fake content [18].

7.4 Convolutional Neural Networks (CNNs)

Compared to detection that is done by humans, convolutional neural networks (CNNs), and other similar approaches, work on the principles of machine learning and have the ability to detect deepfake content via powerful image analysis features [13]. These artificial intelligence algorithms possess the ability to be stationed on information sharing platforms and implanted in social media. They run in the background, continually monitoring uploaded content and detecting whether that content is authentic or fake. Hence, this technique allows alerting users or removing fake content in a timely manner and before it is disseminated.

8. DISCUSSION

Advances in technology, especially in the field of artificial intelligence, have resulted in the development of techniques that tamper with images and videos. Koopman et al [1] states that images and videos are used as evidence in investigations and in the courtroom and, hence, are considered to be reliable sources of information. However, the ability to create tampered images and videos that appear realistic potentially call these information sources into question as evidence. One technique for creating these tampered images and videos is known as deepfake. Chawla [10] reported that deepfake was a pseudonym of an anonymous user who uploaded tampered pornographic videos to the famous website Reddit claiming that they were authentic and belonging to famous actresses. With the astounding, realistic results, deepfake quickly become popular and gained massive media attention. According to Harris [3], deepfakes are usually made in order to take revenge, blackmail a person or to create and upload a pornographic video containing a celebrity. Korshunov and Marcel [5] found that this technology had matured over the past couple of years and had become a major public concern. Deepfakes are often used to create fake news, provide misinformation about an event or to sabotage a public figure. Detecting whether an image or video has been tampered, and subsequently flagging and/or filtering it from view, is a significant technical challenge. According to Maras and Alexandrou [6], deepfakes are created using large datasets that include photos, videos, or recordings. The manipulation of videos or photos, as mentioned in [1], is achieved using hundreds or thousands of photos of a targeted person. Deepfake relies on techniques from the field of artificial intelligence, a field that was developed to understand human behavior and thought processes. The deepfake technique is popular because it produces photorealistic results and is easy to use by a layperson. The manipulation of an image or video makes use of Google's Image Search. Maras and Alexandrou [6] reported that the deepfake program explores different social media sites for images of a targeted person. It then generates a model that is replaced with the image or video of the other

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

person. These deepfakes, although they seem perfect, can be detected and countered. Multiple detection techniques are discussed that can be used to detect deepfakes. Gardiner [11] reports that multimedia forensics can be used to trace the history of an image or video. By doing so, traces of data can be exploited to detect whether the video or image has been manipulated. Afchar et al [4] supports face detection as another technique for detecting manipulation, as it recognizes minor changes that are neglected by the deepfake program. Although this tampering and manipulation technology is mainly used to create unethical content, there are valid areas for the application of this technology. Gardiner [11] suggests that deepfake can be used in multimedia industries, such as the movie and gaming industries, and large social applications. However, it remains evident that this technology is widely used to create content that supports unethical beliefs.

Table 1: Systematic survey of deepfake detection methods

Ref	TITLE	METHOD	RESEARCH FINDINGS
[10]	How a pervert shook the world	Eye blinking	Using Artificial Intelligence to detect the blinking of an eye
[7]	Exploiting visual artefacts to expose deepfakes and face manipulations	Detecting differences in eye color and the refection in teeth	Using Computer Vision to detect slight color reflection in the eyes and missing details in teeth
[11]	Facial re-enactment, speech synthesis, as well as the rise of the Deepfake	Multimedia forensics	Using the detailed history of an image to detect any changes
[18]	An overview of image forensics	Watermarking	Detection was done by identifying hidden traces
[13]	Deepfake video detection using recurrent neural networks	CNN	Using powerful image analysis to detect any minor defects or changes that have been made to an image or video

Advances in science and technology supported the creation of deepfake and its increased effectiveness and efficiency over the past few years. Deepfake content has not only decreased the authenticity of evidence, but it has also affected the lives of many people.

Deepfake is a technology that can create a realistic altered image or video. Deepfakes have impacted the virtual world, where the validity of a video or image can now be questioned. Politicians use this technology to undo what they said or did while web users often use deepfake to create scandals of celebrities or to harass someone. Deepfakes are also used for entertainment for many users. The deepfake software is easy to use and this ease of use contributes to the increasing uptake of deepfake. In short, this techno1logy has harmed the social world. Despite all the drawbacks of deepfake, there are a few applications of deepfakes that are worth considering. These include the dubbing of characters in the animation and gaming industry. The use of artificial intelligence techniques has made it easier to detect and counter the dissemination of deepfake content. Various detection strategies like CNNs, multimedia forensics, and watermarking can be used to efficiently detect deepfakes. Detection allows appropriate action to be taken to either delete the content or to flag the content as tampered. To conclude, even though deepfake has impacted this world in a negative way, these negative aspects can be detected and countered leaving the technology available for multiple fruitful applications

REFRENCES:

[1] M. Koopman, A. M. Rodriguez, and Z Geradts, Detection of Deepfake Video Manipulation, *Conference: IMVIP*, 2018.

[2] D. K. Citron and and R Chesney, Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security. *Draft*.

[3] D. Harris, Deepfakes: False Pornography Is Here and the Law Can not Protect You, *Duke L. & Tech. Rev*, vol. 17, pp. 99–99, 2018.

[4] D. Afchar, V. Nozick, J. Yamagishi, and I Echizen, Mesonet: a compact facial video forgery detection network, *2018 IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1–7, 2018.

[5] P. Korshunov and S Marcel, Deepfakes: a new threat to face recognition? assessment and detection. *arXiv preprint arXiv:1812.08685*.

[6] M. H. Maras and A Alexandrou, Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos, *The International Journal of Evidence & Proof*, pp. 1 365 712 718 807 226–1 365 712 718 807 226, 2018.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[7] F. Matern, C. Riess, and M Stamminger, exploiting visual artifacts to expose deepfakes and face manipulations, *2019 IEEE Winter Applications of Computer Vision Workshops (WACVW)*, pp. 83–92.

[8] R. Delfino, Pornographic Deepfakes—Revenge Porn's Next Tragic Act–The Case for Federal Criminalization. *Available at SSRN 3341593*.2019.

[9] D. Stover, Garlin Gilchrist: Fighting fake news and the information apocalypse, *Bulletin of the Atomic Scientists*, vol. 74, no. 4, pp. 283–288, 2018.

[10] R. Chawla, Deepfakes: How a pervert shook the world, *International Journal of Advance Research and Development*, Vol 4, 2019.

[11] N. Gardiner, Facial re-enactment, speech synthesis and the rise of the Deepfake. *Edith Cowan University, Theses* 2019.

[12] B. J. Siekierski, Deep Fakes: What Can be Done About Synthetic Audio and Video. Library of Parliament.2019.

[13] D. Guera 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.

[14] Y. Li and S. Lyu. (2018). Exposing deepfake videos by detecting face warping artifacts. [Online].Available:https://arxiv.org/abs/1811.00656.

[15] J. Hui, how deep learning fakes videos (Deepfake) and how to detect it? Medium Corporation. Accessed from: https://medium.com/@jonathan_hui/how-deep-learning-fakes-videos-deepfakes-and-how-to-detect-it-c0b50fbf7cb9. 2018.

[16] A. Zucconi, How to Create the Perfect DeepFakes. Alan Zucconi. Accessed from: https://www.alanzucconi.com/2018/03/14/create-perfect-deepfakes/

[17] H. Alibašić and J. Rose, Fake News in Context: Truth and Untruths, Public Integrity, vol. 21, no. 5, pp. 463–468, Jun. 2019.

[18] A. Piva, An overview on image forensics, ISNR Signal Processing, pp. 1–22, October 2012.