Sri Lanka Institute of Information Technology

Emergence of Deepfake Technology

Individual Assignment

IE2022 - Introduction to Cyber Security

Student Registration Number	Student Name
[redacted]	Aththanayaka P.A.G.P.B.

26/05/2021

Table of Contents

Abstract

The present-day digital technologies make it harder and harder for the human brain capacity to detect what are the real technologies and what are fake. Deepfake technology is one of the most recently developed and emerging technology which marked a turning point in the creation of fake content. Deepfake technology is powered by the most advanced technological methodologies of Artificial intelligence and machine learning. The main task of this technology is, generating hyper-realistic fake content based on artificial intelligence such as videos, pictures, and audio clips containing the things someone never said or did. The worst effect of this deepfake technology is the possibility of deceiving a targeted person or entire society is endless because of the hardness of detection. With the latest social media technologies, a deepfake content can be spread among millions of people within few seconds which will implement a negative impact on society undoubtedly. Therefore, the society must be aware of this kind of technology and must be prepared for an attack like this. This report is a comprehensive review of deepfake technology and an overview of its underlying technology. In this report, there will be in detail content such as what is deepfake, which kind of parties generate deepfakes, what are the beneficiaries and negative impacts of deepfake technology, Examples of deepfakes, countermeasures against deepfake and the future development of deepfake technology, which analyzed and gained utilizing publicly available researches, books magazines and related video sources.

1. Introduction

Deepfake is an Artificial intelligence and machine learning based technology made by highly enthusiastic techniques that can basically replace a face with another. The main purpose of this technology is, to swap the face of a targeted person to a video acted by someone else. That acting person preventing like the targeted person and he or she is saying and doing things the targeted person does. This encloses one variant of deepfake technology which named face swap. There are another few categories of deepfake such as lip-sync which modified the mouth movement of a video along with a completely different audio recording and puppet master which makes an animated video of a targeted person and merges it with the eye and mouth movements of another person [1].

There are a few major methodologies to generate deepfake content. Some deepfake content can be created using computer graphics and visual effects also known as CG and VFX. Some deepfake content can be created using sophisticated models such as auto encoders and generative adversarial network, which are widely used in computer vision and neural network domains [2]. These technologies are used understand or study the facial expressions and emotions of a person and merge those expressions and movements with another person's images or videos and sync accurately [3]. Usually, creating deepfake content requires a large number of pictures or videos of the targeted person since the accuracy and realism of the deepfake content depends on the number of expressions that can be made using real images or videos. There is a large number of photos and videos of politician and actors, which are freely available on the internet. Therefore, they become the main target of deepfake content. It is a vicious threat to the world when deepfake technology is used to generate videos of world leaders. Deepfakes can be made as they appear and address the public saying false information and purposes [4]. And deepfake content can be made to create unnecessary political or religious tension between countries and religions. False content like that can also cause a disorder in financial markets. Deepfake technology also can be used to create fake satellite images or videos. By creating fake satellite images which are containing objects that really do not exist on the earth. That can confuse military analysts and mislead the military troops in wars [5].

There can be found some positive impacts of deepfake technology as well such as applications in visual effects as known as VFX, camera filters, and digital avatars. But the malicious use of deepfake is rising above the positive impact. development of advanced deep neural networks and machine learning made the process of creating deepfake simpler. And those content becoming indistinguishable to humans and computer algorithms. The process of creating deepfake creation is becoming simpler. In some instances, it only needs a small picture or a video of the targeted person [6]. Therefore, deepfake content can harm not only public figures but ordinary people. For example, in recent history, the CEO of a UK energy firm was scammed using a deepfake audio and stolen 243 thousand dollars [7].

2. Evolution of Deepfake Technology

2.1 History of Deepfake Technology

The basic purpose of deepfake technology, which is to swap a real fake with another fake face was firstly raised around 1865. It was the first known attempt to change the face of someone with another face using a painting. In this deepfake former US president, Abraham Lincoln's face was swapped with the body of John Calhoun who was a politician in the southern US. After Abraham Lincoln's death, demand for that deepfake painting was surprisingly increased. In the modern world, photo manipulation technology was first introduced in the 19th century, and the same kind of technologies developed later to manipulate videos as well as images. During the 20th century, deepfake technologies improved along with the development of digital media [8].

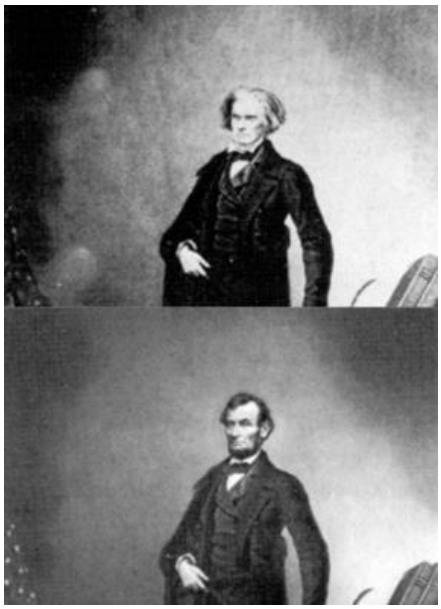

Fig. 1. – Deepfake painting of Abraham Lincoln

2.2 Deepfake Creation Process

Deepfakes have emerged as famous since the superiority of content and additionally, the convenience of use of the applications to a huge range of users with diverse technical skills from expert to apprentice. These deepfake content-generating applications are normally evolved primarily based on deep learning technologies which are widely known for their functionality of representing high dimensional data. One of the most popular tools with that functionality is deep autoencoders. which have been extensively carried out for photo compressions and dimensionality depletion [9].

The first strive of creating deepfake content was the module named Faceswap which developed utilizing the technology called autoencoder-decoder pairing structure [10]. Autoencoder is used to capture the features of the face and auto decoder is used to reconstruct that face image. To swap the source person's face and targeted person's face, two more encoder-decoder pair is needed. All the parameters of encoders are shared among two pairs of encoders which enable to find the similarity of the two faces. This process is relatively simple since the basic shapes and properties are the same in every face such as the position of the eyes, mouth, and nose. Fig. 2 explains the process of autoencoder-decoder.

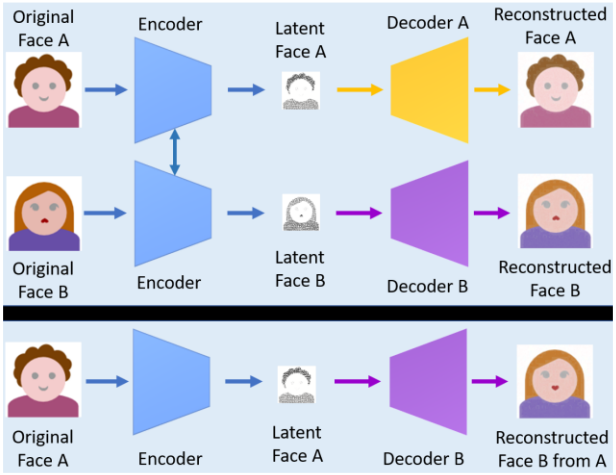

Fig. 2. – Process of autoencoders and decoders

By combining adversarial loss and perceptual loss techniques to the autoencoder-decoder architecture, an improved version based on the generative adversarial network known as GAN was developed. FaceSwapGAN is one example that developed under that improved architecture [11]. The major advantage developers attain by adding the adversarial loss and perceptual loss is this technique can make some face properties more realistic such as eye movements. And this technique helps to keep the consistency of input faces to smooth out the process which causes high-quality outputs with sizes of 64x64, 128x128, and 256x256. To make face detection more stable and face alignments more reliable, a Multi-task convolutional neural network known as CNN was introduced [12].

2.2.1 Deepfake Tools

i. Faceswap

Source - https://github.com/deepfakes/faceswap

When the face swap application was initially developed and released, that technology was unconventional. It has been an enormous step in artificial intelligence developments because the source code of this application was completely opposite of the current theories and it was confusing [13].

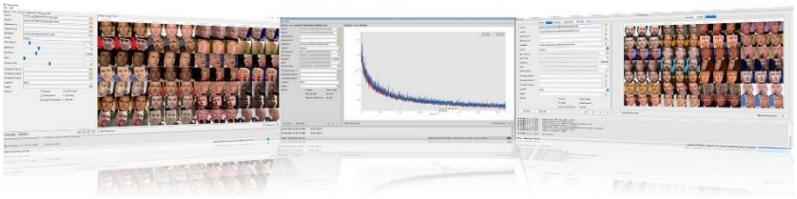

Fig. 3. – Faseswap tool

ii. Faceswap-GAN

Source - https://github.com/shaoanlu/faceswap-GAN

Faceswap-GAN is another application made for face swapping process. The major purpose of this model is change multiple faces utilizing single face image. As the founders states it can swap 5 or less targeted images using one image [11].

Fig. 4. – Process of swapping five faces using one face image

iii. Few-Shot Face Translation

Source - https://github.com/shaoanlu/fewshot-face-translation-GAN

Few shot face translation model is capable of creating deepfake content that has almost similar looking directions, spectacles, and hair of the source face. Developers of this module states that this module is more optimal for translating Asian faces [14].

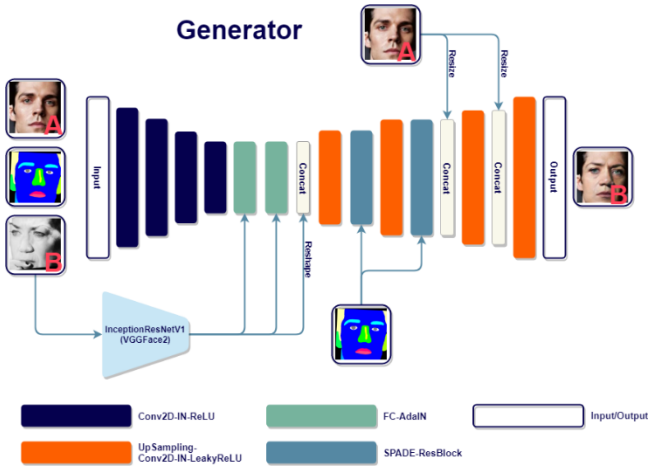

Fig. 5. – Process of few shot face translation

iv. DeepFaceLab

Source - https://github.com/iperov/DeepFaceLab

DeepFaceLab is one of the leading modules founded to create deepfakes. Developers of this module state that 95% of successful deepfakes are created using this module. And also, DeepFaceLab is also used by popular YouTube channels such as CorridorCrew and DeepFakeCreator. This module has some significant features such as manipulating lip movements and de-age faces [15].

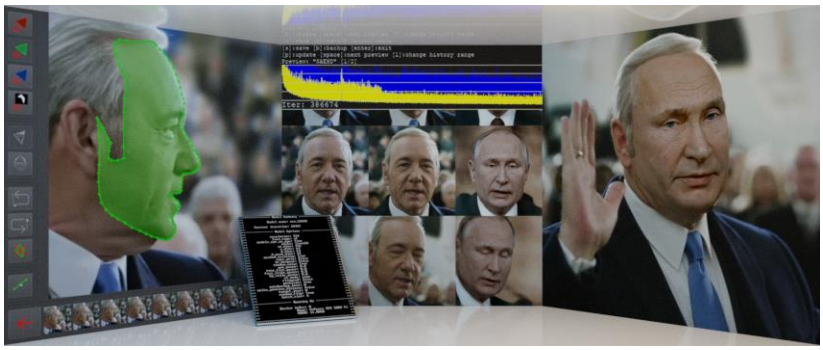

Fig. 6. – Interface of DeepFaceLab tool

v. DFaker

Source - https://github.com/dfaker/df

DFaker is another deepfake generating module. It is implemented based on the Keras library. Keras library is an open-source software library that can be utilized as a python interface for artificial neural networks. This module requires a 64x64 image of the targeted person. and it generates 128x128 pair of images one with RGB colors and one with black and white [16].

Fig. 7. – Face swapping process of DFaker

vi. Avatarme

Source - https://github.com/lattas/AvatarMe

Avatarme is a prominent technique that is capable of redeveloping hyper-realistic 3D faces from a single ordinary picture with improved details. As the developers say, they had to collect a large number of datasets of facial shapes to generate such kind of realistic 3D faces [17].

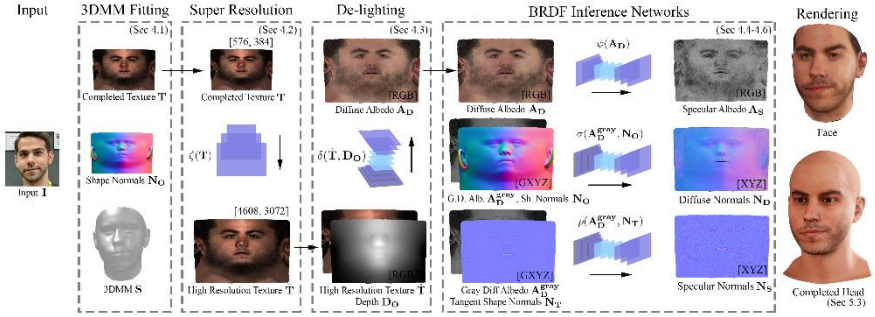

Fig. 8. – Process of Avatarme tool

vii. MarioNETte

Source - https://hyperconnect.github.io/MarioNETte

The main issue of most of the deepfake generating modules is the difficulties to find the identity of the deepfake. Especially with fewer source images. To overcome that issue an improved architecture named MarioNETte was introduced. MarioNETte has the capability to maintain the identity of the targeted person in the deepfake creation [18]

.

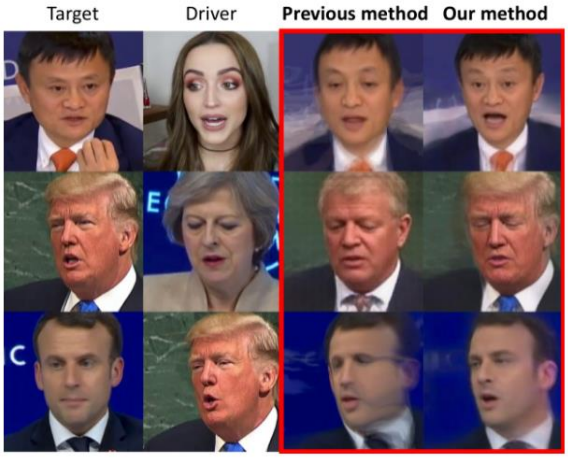

Fig. 9. – Comparison between other tools and MarioNETte

viii. DiscoFaceGan

Source - https://github.com/microsoft/DiscoFaceGAN

DiscoFaceGan can generate face images by treating some properties of the original picture as independent variables. Therefore, changing one property will not be affected to other properties. Identity, expressions, pose and illumination are the four main properties [19].

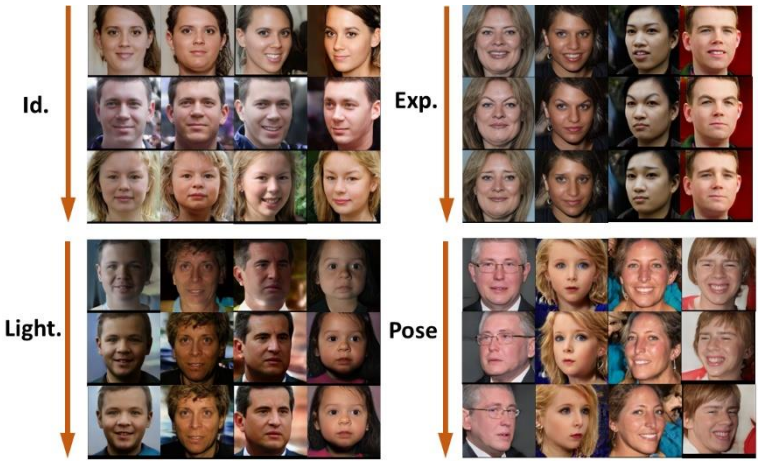

Fig. 10. – Process of DiscoFaceGAN tool

ix. StyleRIG

Source - https://gvv.mpi-inf.mpg.de/projects/StyleRig

StyleRIG is another popular module for deepfake creations. It can generate hyper-realistic portrait face images with eyes, hair, neck, teeth, and shoulders. But it lacks the controlling over the properties like pose, expressions, and lighting [20].

x. FaceShifter

Source - https://lingzhili.com/FaceShifterPage

FaceShifter is a deepfake generating module that swaps faces with high constancy by exploiting and syncing the attributes of the targeted person conscientiously. This module contains a tool that direct to recovery anomaly regions named Heuristic Error Acknowledging Refinement Network (HEAR-Net) [21].

Fig. 11. – Shifted face images using FaceShifter

xi. FSGAN

Source - https://github.com/YuvalNirkin/fsgan

Face swapping GAN known as FSGAN is another module capable of face swapping and re-arranging. The specialty of FSGAN is it can be applied to two face images and swap them without train those faces. This module utilizing Recurrent Neural Network (RNN) to rearrange faces which can adjust both pose directions and expressions. This module can be applied to both images and videos [22].

xii. Transformable Bottleneck Networks

Source - https://github.com/kyleolsz/TB-Networks

Transformable Bottleneck Networks which known as TBNs have the capability to manipulate fine-grained 3D manipulations. TBN has arbitrary non-linear transformation which enables the creation of creative object manipulations for its users [23].

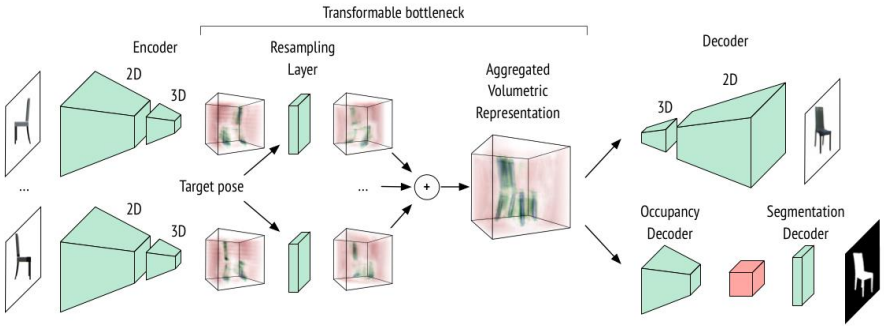

Fig. 12. – Process of Transformable Bottleneck Network

xiii. Neural Voice Puppetry

Source - https://justusthies.github.io/posts/neural-voice-puppetry

Neural voice puppetry is a latest and advance audio driven deepfake creator. By inputting an audio sequence of a source person or a digital assistant, it can generate a hyper-realistic output video of targeted person including lip sync with the audio of source person which inputted. This deepfake creation module is use deep neural network systems [24].

Fig. 13. – Process of Neural Voice Puppetry

xiv. Do as I Do Motion Transfer

Source - https://github.com/carolineec/EverybodyDanceNow

Do as I do motion transfer is a different approach to create deepfake content. It can examine a given video of a source person dancing and transfer that motion to a targeted person's video who performing only a few standard moves. This technique is known as video-to-video transition. As the developers state, they extract every pose from the source person and apply a technology called pose-to-appearance to generate the targeted person's deepfake video [25].

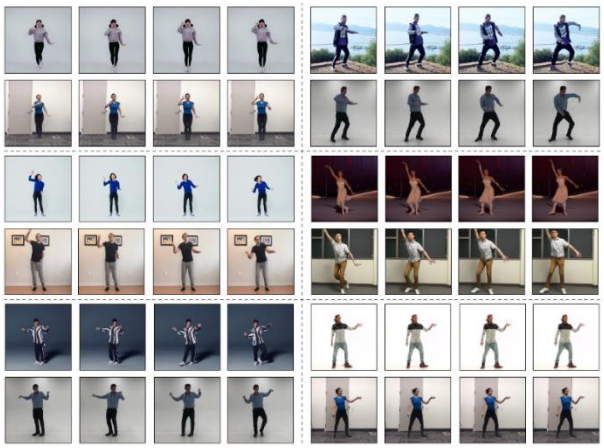

Fig. 14. – Dancing figures made using Do as I Do model

2.3 Deepfake Detection Process

In a situation deepfake keep damaging society and privacy, Methods for detecting deepfakes should have to be founded as soon as the deepfake creating application is introduced. The classical method to detect deepfake images and videos was a natural approach that examines manually and finds inconsistencies and artifacts of the fake content creation process. The Modern approach is to find deepfake content is an automatic process. It can detect major and discriminative features of deepfake contents [26].

Usually, deepfake detection tools using a binary classification system to examine and classify real content and fake content. This kind of strategy requires a database containing a huge number of real and fake videos to program the classification module. The quantity of fake content is highly available. But it is yet restricted as far as setting a benchmark for approving different identification strategies. To overcome this issue, Korshunov and Marcel [26] were able to create an outstanding deepfake dataset comprising 620 videos generated based on GAN model utilizing an open-source code named FaceSwapGAN [11] which explained in deepfake creation tools. VidMIT database which provides audio and video clips freely were used to create both low and high-quality deepfakes [27]. Those videos were utilized to test different deepfake identification techniques. Test outcomes show that famous face recognition frameworks dependent on VGG and facenet are unable to detect deepfakes effectively [28]. Other technologies like, lipsyncing and image quality measurements with help of vector machines known as SVM were able to produce a high error rate [29]. This raises concern about the basic need for future advancement of more strong strategies that can distinguish deepfakes from real. This survey helps to group all the deepfake detection methods into two major categories. Those are fake image detection and fake video detection.

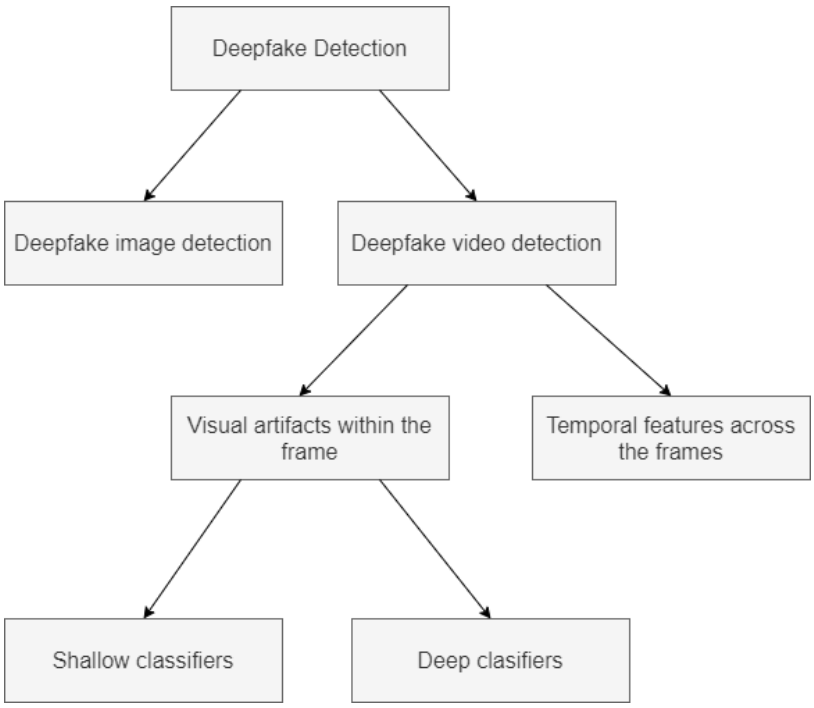

Fig. 15. – Categories of deepfake detecting methods

2.3.1 Fake Image Detection

Deepfake images are dangerous as deepfake videos. Basically, deepfake images can replace faces in photographs with a new face image from a collection of stock images. This is one of the most popular methods used by cyber attackers to exploit identification or authentication and gain access to unauthorized systems. By utilizing deep learning tools such as CNN and GAN explained in the deepfake tool section, all the properties of the images such as pose, facial expressions, and illumination [30]. Zhang and team [31] utilized the bag of words strategy to extract a set of compact features and fed them into deepfake detectors like SVM [29], Random forest known as RF [32], and Multi-Layer Perceptions known as MLP [33] to segregate fake or swapped face videos from real videos. Deepfake images which are made using GAN technology are the fake images that are most difficult to detect because of the high-quality outputs that GAN-based modules provide.

Most deepfake image detectors with recognition of images generated based on GAN technology, do not consider the speculation ability of the deepfake detection models even the

development of GAN technology is progressing, and numerous new augmentations of GAN technology are often introduced. Xuan and team [34] utilized an image preprocessing step, for example, Gaussian blur and Gaussian noise to eliminate low-level and high-frequency details in GAN-based fake images. That can make an increment of the pixel level similarity between fake images and real images and it requires forensic classifier to learn more about features which have better generalization abilities than previously introduced fake image detecting methods.

Besides that, that GAN-based detection method was a theory-testing issue where a measurable structure was presented utilizing the data hypothetical investigation of verification. The analytic outcomes show that the rate of oracle error increments when the GAN generator is less precise. Oracle error is the minimum distance between legitimate images and images generated using GAN-based tools. Therefore, in the case that the "oracle error" rate is increasing it becomes easier to detect deepfake images. But as the analysis states in a case of high-resolution image input, an extremely accurate GAN-based tool can generate deepfake images which are hard to detect still [35].

Lately, Hsu and team [36] presented a two-phase deep learning strategy for the recognition of deepfake images. The main stage of this method is a feature extractor which dependent on the common fake feature network known as CFFN. common fake feature network contains a set of units called dese blocks [37] in each unit including a different number of dense blocks in order to improve the representative ability for deepfake images. The number of dense blocks is relying upon the approval information being face or general images. The quantity of channels in each unit is changed up to a few hundred. Discriminatory changes between deepfake images and real images are extracted throughout this common fake feature network learning process. Then all of that features are sent to the second phase of this deep learning strategy which can consider as the last involvement layer of the common fake feature network in order to recognize fake images from real images.

This method of detecting deepfake images is validated for both fake face images and fake general images. To detect face images this method utilizing datasets obtain from CelebA [38] which contain more than 200,000 face images of 10,177 celebrity identities with a variety of poses

and backgrounds. Five major generative adversarial networks were used to create deepfake images with the size of 64x64, including deep convolution GAN known as DCGAN, Wasserstein GAN known as WGAN, Wasserstein GAN with gradient penalty known as WGAN-GP, least-squares GAN, and progressive growth of GAN known and PGGAN. 358,198 training images and 10,000 test images of both real and fake faces are generated to test this method. To detect general fake images, they extracted a general dataset from ILSVRC 2012 [39]. Few GAN training models were used to generate these general fake images with sizes of 128x128. Those GAN models were, the large-scale GAN training model for high fidelity natural image synthesis known as BIGGAN, self-attention GAN, and spectral normalization GAN. This dataset consists of 600,000 fake and general images and 10,000 test images with both real and fake details. These experimental results showed a significant performance to detect fake images against the other methods.

2.3.2 Fake Video Detection

Most of the image detection methods and tools cannot be utilized for deepfake video detection. The reason for that is the details of one frame, debasement when it compresses as a video [40]. Besides, that videos have temporary properties which are changing rapidly after each frame. Therefore, detecting fake videos is slightly challenging for the methods that trained to detect deepfake images. This subsection explains the methods which trained to detect deepfake videos and it containing two major categories named employ temporal features and explore visual artifacts within frames.

i. Temporal Features across Video Frames

Based on the inspection that temporal properties of a video cannot use effectively to detect deepfake videos, Ekraam Sabir [41] explains in his research that spatio temporal feature of video can be used to detect the deepfake videos. Video creation is done on a frame-by-frame basis. Based on that theory, the Recurrent convolutional model which known as RCN was developed. It developed under the integration of the convolutional network densenet [37] and the gated recurrent unit cells. This model is capable of exploiting the temporal disparities among the frames of video. This strategy is tested on the FaceForensics++ Dataset which includes 1000

original video sequences created using deepfake creation tools such as FaceSwap, Neutra Textures, and Face2Face and the test outcomes show positive and promising results [42].

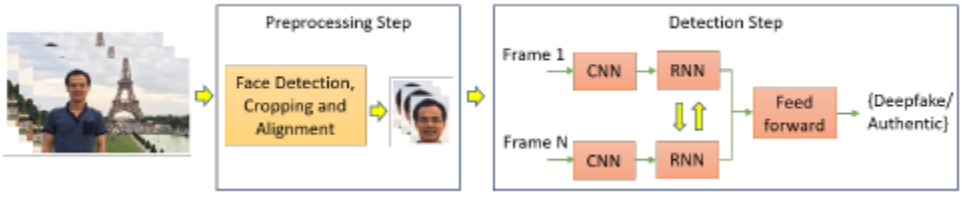

Fig. 16. – Process of detecting temporal features across the frames

David and Edward [8] state that deepfake videos can be detected from temporal inconsistencies since most of the deepfake videos are containing intra-frame inconsistencies. They point out a temporal-aware pipeline method based on CNN and long short-term memory know as LSTM can be used as a deepfake video detector. As explained in the deepfake image detecting method CNN is used to extract features in every frame. Long short-term memory is utilized to temporal sequence descriptor. A completely associated network is finally utilized to detect deepfake videos with the involvement of a sequence descriptor.

Besides these detecting models, there is one physiological sign to detect deepfakes. Eye blinking can be used as a major clue to detect deepfake videos. Dependent on the perception that a person in a deepfake video, has a significantly a smaller number of eye blinking. As written in medical researches a healthy person normally blinks 15 to 20 times each minute and each blink lasts 1.0 to 4.0 seconds [43]. As explained in the deep face creation methods part, most of the deepfake creation technologies using face images that are taken from the datasets to train the deepfake models. Most of the face images available in those datasets are the images of faces with opened eyes. Therefore, deepfake video creating tools are not having the capability to create fake videos with faces with blinking eyes that are blinking in the normal frequency. Yuezun Li and the team [44] attempted to utilize this property and detect deepfakes. They decode the video into frames then examine for parts where the faces appear and extracted eye areas in faces utilizing six eye landmarks which mark boundaries of the region where eyes aligned. The next step is creating a new sequence with cropped eye area frames after doing some alignments and scaling

the boundaries. This newly created sequence will be sent to long-term recurrent convolutional networks known as LRCN [45] to do the dynamic state prediction. LRCN contains a feature extractor made based on CNN, a sequence detector made based on Long Short-Term Memory known as LSTM which helps to detect temporal patterns and state predictor which utilize to predict the probability of eye blinking rate. To evaluate this method a dataset contains 49 interviews and presentations and corresponding fake videos made utilizing deepfake tools. These experimental outcomes were promising in detecting fake videos, which can further improve.

ii. Visual Artifacts within Video Frame

The previous chapter explained the methods to detect deepfake videos utilizing temporal patterns across the frames of videos. Almost all of that methods were based on a deep recurrent network model. This subsection explains a different approach to detecting deepfake videos that examine each frame of the video to find visual artifacts by decomposing the video. For further classifications, this kind of detection methods divided into two parts named deep classifiers and shallow classifiers.

a. Deep classifiers

Deepfake videos are regularly generated with restricted resolutions since they require performing face warping processes such as rotations and scaling-ups to match the properties of the original face. Because of this resolution difference, these deepfake videos creating methods, leave some clues in warped areas of the faces which can detect by the deepfake detecting models such as VGG16 which also known as oxfordnet developed by Visual Geometry Group from Oxford, and Residual neural network versions such as ResNet 50,101 and 152. A deepfake detection method to detect deepfakes by examining the artifacts of wrapped areas of deepfake videos is firstly explained by Yuezun and the team [46]. To evaluate this method, they use two datasets named UADFV [47] which contains 49 real videos and 49 deepfake videos with 32,752 total frames and DeepfakeTIMIT dataset which contains 64x64 size of low-quality videos and 128x128 size of high-quality videos with10,537 original images and 34,023 fake images extracted from 320 videos. The major advantage of this model is it not required to create deepfake videos as negative examples

before training the model. Instead of that what this model does is, generate images by extracting only the face region of the original image and rescale and warping back to the original image without applying the gaussian blur effect. This method reduces a lot of time and resources compared to other models.

b. Shallow classifiers:

Most of the deepfake methods are based on detecting artifacts or irregularity between fake and real videos. Xin Yang and the team [47] states another new approach to detect deepfake videos by examining the differences in the pose orientation and positions of fake 3d models which generate utilizing 68 landmarks in the original face image. The reason they choose 3d face models is they detected some weaknesses in 3d deepfake generating process. This detecting model is based on the Support Vector Machine algorithm known as SVM. As the previous models evaluate, this model also uses two datasets to evaluate. The first dataset is UADFV which contains 49 deepfake videos and 49 real videos. The second dataset is a subset of the DARPA dataset which contains 241 real images and 252 deepfake images. This model is requiring images or videos that contain certain properties such as an open mouth and eyes. That is one major disadvantage of this model.

2.4 Uses of Deepfake

2.4.1 Blackmail

Blackmail is requesting some advantage from someone as a trade-off for not revealing sensitive information about them. Using deepfake content as blackmailing material can be effective since the realities of the deepfake videos and images. A deepfake content can be used for an activity like this. Blackmail utilizing deepfake content can be used to falsely incriminate a victim. On the other hand, utilizing deepfake content as blackmailing materials raises a new issue. Since the fake content cannot be reliably distinguished from real content, a victim who blackmailed can state that the even real content is fake. This is a major issue for the real blackmailers which is known as blackmail inflation.

2.4.2 Politics

Politics is the most targeted domain of the deepfake creators. What deepfake content can do is simply mispresent a politician in a video or in an image. There were so many incidents related to the politicians' deepfake videos and images in recent history as well. In 2018, A popular American actor, comedian, and filmmaker Jordan Peele made a deepfake video associated with an entertainment company named BuzzFeed. The target of this deepfake video was Barack Obama. Jordan Peele's voice was utilized for this video synced with Barack Obama's lip movements. In this deepfake video He explains the deepfake videos and how can it effect society [48]. In 2019, an employee who worked at the KCPQ channel which owned by the Fox television network made a deepfake video targeting Donald Trump. In that video, the fake character was pretending like Donald trump is addressing his staff at the Oval Office in the White House [49]. In 2020, a deepfake video was published that made targeting Belgian Prime Minister Sophie Wilmes. That video was about a connection between Covid-19 and deforestation [50]. Including the examples explained so far, there were so many cases raised about deepfakes of politicians.

2.4.3 Movie Industry

Utilizing deepfakes in the Digital entertainment industry can consider as a positive impact of deepfake technology. By utilizing deepfake technology digital content creators can fulfill their purposes that were impossible in the past. For example, they can create a younger version of an actor for the prequel movies that have been become very popular recently. On the other hand, movie producers can use deepfake technology for the actors who passed away during the production of that movie. Placing Harrison Ford's younger version as young Han Solo in “Solo: A Star Wars story" is one example of that incident [51]. In the same movie series, Carrie Fisher's face was deep faked as a young Princess Leila in Rough One movie which is a prequel movie of the Star Wars movie series [52]. The world-famous movie-making company: Disney is a leading company among the companies that use deepfake technologies in their movies [53]. They

developed a model with the capability to identify facial expressions and face-swapping which generate high-quality outputs with 1024x1024 resolutions [54].

Fig. 17. – Harrison Ford's younger face swapped with another actor

2.5 Responses taken against Deepfake

In the United States, there are some legal responses raises against the deepfake content-related cases. As the first step against deepfake content, in 2018, the Deepfake content prohibition act was acquainted with the US Senate [55]. In 2019, Act against the Deepfake liability was introduced for the first time at the house of Representatives which is the lower house of the US congress [56]. In several states in the US like New York, Virginia, and Texas, there are different kinds of rules against the deepfakes. As an example, Governor in California established two laws against deepfake and deepfake related activities. The first one is against the creators who create sexually explicit deepfake content without the targeted person's consent [57]. As in that law, it can cause action against the creator of that deepfake content. The second law prohibits the distribution of the deepfake content created targeting candidates running for office, within 60 days before the election [58].

Twitter is a huge social networking service that spreads over the world. As H. Tankovska states in 2020, there were 187 million daily active users worldwide who use Twitter [59]. Therefore, the possibility to spread deepfake contents is high. To overcome such issues twitter is taking few preventive measures against manipulated or fake content. Once Twitter detects fake or

manipulated content, they notice their users about that. Therefore, when a user tries to like, retweet, or share externally, Twitter will auto-generate a warning message which makes the user reconsider engaging with that tweet. Twitter also can remove the tweets which are containing fake and manipulated content to prevent the harm that can happen to its' users. To improve twitter's safety and detection level against fake and manipulated content, Twitter is letting its users contribute. They are inviting users who are interesting in detecting deepfake content to join with twitter and work on fake and manipulated content detection to make Twitter more secure [60].

Partnering with other industry leaders and academic experts, Facebook company was able to host a challenge named Deepfake Detection Challenge (DFDC). That was held in December 2019. It has enabled to all the tech experts all around the world who interested in deepfake detection to introduce their deepfake detection methods and modules. 2114 participants were joined with this competition and total number of deepfake detection modules they introduced was more than 35000. Most significant deepfake detection modules were selected for further researches and improvements [61].

3. Future of Deepfake Technology

Presently there are many tools developed mostly based on GAN technology that can be used to create content using deepfake technology, and much counterfeit content has been created using those tools. On the other hand, there are a number of modules that can distinguish deepfake content from genuine content. Almost every deepfake content is recognizable as fake content and every detecting model has some kind of issue that downturn the capability of detecting deepfake content. Considering the most recently developed tools, it can be concluded that the situation will change in the future. More efficient deepfake content recognition tools will be developed. Deepfake creating tools will be also developed in the future that can focus more on lighting, orientation pose, and the properties which could not track down yet.

Although it is now mandatory to have some kind of basic knowledge of the domain to use both deepfake contents generating tools and deepfake content detection tools, both types of tools will be released for general public use in the future. With the use of these tools by the public, the use of deepfake content will become very common. Also, the number of crimes committed using fake content will increase. On the other hand, new laws will be enacted to curb such crimes, and as mentioned in the blackmail chapter in this report, people will be tempted to say that the crime was committed using fake content, even if it was done with the genuine content in the future.

With the capability of creating more realistic deepfake content, the time, labor, and cost of creating digital content will be significantly reduced. When deepfake technology is combined with new technologies such as holographic technology, more realistic content will be created. Finally, with the continued use of deepfake technology, the value of factual information will be plummet. This is one of the biggest disadvantages of the Deepfake technology.

4. Conclusion

This report summarized the process of identity change utilizing fake media such as videos and images, known as deepfake technology that is still being developed, making new discoveries from the nineteenth century to the present day. The origins of deepfake technology, the positive and the negative impact of Deepfake technology, the tools used to create fake content using deepfake technology, the tools used to distinguish deepfake content from genuine content, and the suppression of deepfake content from various countries and public social media networks and the future of deepfake technology were the main topics of this report. As described under that topics, it is clear that deepfake technology is not a domain that will end up being built. The continuous evolvement of the deepfake content creating methods as well as deepfake detecting methods is the main reason for that. However, the disadvantages of deepfake technology are greater than the advantages of deepfake technology acquired by few domains such as the movie industry. If there is a chance to enhance more in the deepfake detection domain and if deepfake technology can be adapted in a more advantageous way, there is no doubt that there will be a revolution in digital media technology in the future.

5. References

[1] Shruti Agarwal, Hany Farid, Yuming Gu, Mingming He, Koki Nagano, Hao Li, "ProtectingWorld Leaders Against Deep Fakes," IEEE Xplore, 2019.

[2] Ming-Yu Liu , Xun Huang, Jiahui Yu, Ting-Chun Wang, Arun Mallya, "Generative Adversarial Networks for Image and Video Synthesis: Algorithms and Applications," 2020.

[3] S. Lyu, "Detecting ‘deepfake’ videos in the blink of an eye," 29 Aug 218. [Online]. Available: https://theconversation.com/detecting-deepfake-videos-in-the-blink-of-an-eye-101072. [Accessed 13 May 2021].

[4] T. Hwang, ""Deepfakes: A Grounded Threat Assessment"," Center for Security and Emerging Technology, 2020.

[5] P. Tucker, "The Newest AI-Enabled Weapon: ‘Deep-Faking’ Photos of the Earth," 31 Mar 2019. [Online]. Available: https://www.defenseone.com/technology/2019/03/next-phase-ai-deep-faking-whole-world-and-china-ahead/155944/. [Accessed 14 May 2021].

[6] Egor Zakharov, Aliaksandra Shysheya, Egor Burkov, Victor Lempitsky, "Few-Shot Adversarial Learning of Realistic Neural Talking Head Models," Moscow, 2019.

[7] C. Stupp, "Fraudsters Used AI to Mimic CEO’s Voice in Unusual Cybercrime Case," 19 Aug 2019. [Online]. Available: https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402. [Accessed 14 May 2021].

[8] David G¨uera, Edward J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," 2018.

[9] Abhijith Punnappurath, Michael S. Brown, "Learning Raw Image Reconstruction-Aware Deep Image Compressors," 2015.

[10] L. Guilloux, "fakeapp," [Online]. Available: https://www.malavida.com/en/soft/fakeapp/. [Accessed 14 May 2021].

[11] shaoanlu, "faceswap-GAN," [Online]. Available: https://github.com/shaoanlu/faceswap-GAN. [Accessed 14 May 2021].

[12] davidsandberg, "facenet," [Online]. Available: https://github.com/davidsandberg/facenet. [Accessed 14 May 2021].

[13] deepfakes, "faceswap," [Online]. Available: https://github.com/deepfakes/faceswap. [Accessed 14 May 2021].

[14] shaoanlu, "fewshot-face-translation-GAN," [Online]. Available: https://github.com/shaoanlu/fewshot-face-translation-GAN. [Accessed 14 May 2021].

[15] iperov, "DeepFaceLab," [Online]. Available: https://github.com/iperov/DeepFaceLab. [Accessed 14 May 2021].

[16] dfaker, "df," [Online]. Available: https://github.com/dfaker/df.

[17] lattas, "AvatarMe," [Online]. Available: https://github.com/lattas/AvatarMe. [Accessed 14 May 2021].

[18] Sungjoo Ha, Martin Kersner, Beomsu Kim, Seokjun Seo, Dongyoung Kim, "MarioNETte: Few-shot Face Reenactment," [Online]. Available: https://hyperconnect.github.io/MarioNETte/. [Accessed 14 May 2021].

[19] microsoft, "DiscoFaceGAN," [Online]. Available: https://github.com/microsoft/DiscoFaceGAN. [Accessed 14 May 2021].

[20] A. Tewari, M. Elgharib, G. Bharaj, F. Bernard, H-P. Seidel, P. Perez, M. Zollhöfer, C.Theobalt, "StyleRig: Rigging StyleGAN for 3D Control over Portrait Images," [Online]. Available: https://gvv.mpi-inf.mpg.de/projects/StyleRig/. [Accessed 14 May 2021].

[21] Lingzhi Li, Jianmin Bao, Hao Yang, Dong Chen, Fang Wen, "FaceShifter: Towards High Fidelity And Occlusion Aware Face Swapping," [Online]. Available: https://lingzhili.com/FaceShifterPage/. [Accessed 14 May 2021].

[22] YuvalNirkin, "fsgan," [Online]. Available: https://github.com/YuvalNirkin/fsgan. [Accessed 14 May 2021].

[23] kyleolsz, "TB-Networks," [Online]. Available: https://github.com/kyleolsz/TB-Networks. [Accessed 14 May 2021].

[24] Justus Thies, Mohamed Elgharib, Ayush Tewari, Christian Theobalt, Matthias Nießner, "Neural Voice Puppetry: Audio-driven Facial Reenactment," 08 Jan 2020. [Online].

Available: https://justusthies.github.io/posts/neural-voice-puppetry/. [Accessed 14 May 2021].

[25] carolineec, "EverybodyDanceNow," [Online]. Available: https://github.com/carolineec/EverybodyDanceNow. [Accessed 14 May 2021].

[26] Pavel Korshunov, S´ebastien Marcel, "Vulnerability assessment and detection of Deepfake videos".

[27] "VidTIMIT Audio-Video Dataset," [Online]. Available: https://conradsanderson.id.au/vidtimit/. [Accessed 14 May 2021].

[28] Omkar M. Parkhi, Andrea Vedaldi, Andrew Zisserman, "Deep Face Recognition," BMVA Press, 2015.

[29] Javier Galbally, S´ebastien Marcel, "Face Anti-Spoofing Based on General Image Quality Assessment".

[30] Iryna Korshunova, Wenzhe Shi, Joni Dambre, Lucas Theis, "Fast Face-swap Using Convolutional Neural Networks," IEEE Xplore.

[31] Ying Zhang, Lilei Zheng, Vrizlynn L. L. Thing, "Automated face swapping and its detection," IEEE, 2017.

[32] Kasthurirangan Gopalakrishnan, Siddhartha K. Khaitan, Alok Choudhary, Ankit Agrawal, "Deep Convolutional Neural Networks with Transfer Learning for Computer Vision-Based Data-Driven Pavement Distress Detection," 2017.

[33] Lilei Zheng, Stefan Duffner, Khalid Idrissi, Christophe Garcia, Atilla Baskurt, "Siamese Multi-layer Perceptrons for Dimensionality Reduction and Face Identification," 2015.

[34] Xinsheng Xuan, Bo Peng, Wei Wang, Jing Dong, "On the generalization of GAN image forensics," 2019.

[35] Sakshi Agarwal, Lav R. Varshney, "Limits of Deepfake Detection: A Robust Estimation Viewpoint," 2019.

[36] Chih-Chung Hsu , Yi-Xiu Zhuang, Chia-Yen Lee, "Deep Fake Image Detection Based on Pairwise Learning," 2020.

[37] Gao Huang, Zhuang Liu, Laurens van der Maaten, "Densely Connected Convolutional Networks," IEEE Xplore.

[38] Ziwei Liu, Ping Luo, Xiaogang Wang, Xiaoou Tang, "Deep Learning Face Attributes in the Wild," 2015.

[39] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," 2015.

[40] Darius Afchar, Vincent Nozick, Junichi Yamagishi, Isao Echizen, "MesoNet: a Compact Facial Video Forgery Detection Network," 2018.

[41] Ekraam Sabir, Jiaxin Cheng, Ayush Jaiswal, Wael AbdAlmageed, Iacopo Masi, Prem Natarajan, "Recurrent Convolutional Strategies for Face Manipulation Detection in Videos".

[42] Andreas R¨ossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, Matthias Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," IEE Xplore.

[43] E. Hersh, "How Many Times Do You Blink in a Day?," 24 Sep 2020. [Online]. Available: https://www.healthline.com/health/how-many-times-do-you-blink-a-day. [Accessed 15 May 2021].

[44] Yuezun Li, Ming-Ching Chang and Siwei Lyu, "In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking," 2018.

[45] Jeff Donahue, Lisa Anne Hendricks, Marcus Rohrbach, Subhashini Venugopalan, Sergio Guadarrama, Kate Saenko, Trevor Darrell, "Long-term Recurrent Convolutional Networks for Visual Recognition and Description," 2016.

[46] Yuezun Li, Siwei Lyu, "Exposing DeepFake Videos By Detecting Face Warping Artifacts," 2018.

[47] Xin Yang, Yuezun Li, Siwei Lyu, "Exposing Deep Fakes Using Inconsistent Head Poses," 2018.

[48] A. Romano, "Jordan Peele’s simulated Obama PSA is a double-edged warning against fake news," 18 Apr 2018. [Online]. Available: https://www.vox.com/2018/4/18/17252410/jordan-peele-obama-deepfake-buzzfeed. [Accessed 16 May 2021].

[49] K. Swenson, "A Seattle TV station aired doctored footage of Trump's Oval Office speech. The employee has been fired.," 11 Jan 2019. [Online]. Available: https://www.washingtonpost.com/nation/2019/01/11/seattle-tv-station-aired-doctored-footage-trumps-oval-office-speech-employee-has-been-fired/. [Accessed 16 May 2021].

[50] G. Holubowicz, "Extinction Rebellion takes over deepfakes," 2020 15 Apr. [Online]. Available: https://journalism.design/deepfakes/extinction-rebellion-sempare-des-deepfakes/. [Accessed 16 May 2021].

[51] P. Radulovic, "Harrison Ford is the star of Solo: A Star Wars Story thanks to deepfake technology," 18 Oct 2018. [Online]. Available: https://www.polygon.com/2018/10/17/17989214/harrison-ford-solo-movie-deepfake-technology. [Accessed 16 May 2021].

[52] E. W. page, "How acting as Carrie Fisher's puppet made a career for Rogue One's Princess Leia," 16 Oct 2018. [Online]. Available: https://www.technologyreview.com/2018/10/16/139739/how-acting-as-carrie-fishers-puppet-made-a-career-for-rogue-ones-princess-leia/. [Accessed 16 May 2021].

[53] Jacek Naruniec, Leonhard Helminger, Christopher Schroers, Romann M. Weber, "High-Resolution Neural Face Swapping for Visual Effects," 29 Jun 2020. [Online]. Available: https://studios.disneyresearch.com/2020/06/29/high-resolution-neural-face-swapping-for-visual-effects/. [Accessed 16 May 2021].

[54] B. Mitchell, "Disney's deepfake technology could be used in film and TV," 21 Jul 2020. [Online]. Available: https://blooloop.com/technology/news/disney-deepfake-face-swap-technology/. [Accessed 16 May 2021].

[55] "S.3805 - Malicious Deep Fake Prohibition Act of 2018," 21 Dec 2018. [Online]. Available: https://www.congress.gov/bill/115th-congress/senate-bill/3805. [Accessed 17 May 2021].

[56] "H.R.3230 - Defending Each and Every Person from False Appearances by Keeping Exploitation Subject to Accountability Act of 2019," 06 Dec 2019. [Online]. Available: https://www.congress.gov/bill/116th-congress/house-bill/3230. [Accessed 17 May 2021].

[57] "AB-602 Depiction of individual using digital or electronic technology: sexually explicit material: cause of action.," 10 Apr 2019. [Online]. Available:

https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB602. [Accessed 17 May 2021].

[58] "AB-730 Elections: deceptive audio or visual media," 10 Apr 2019. [Online]. Available: https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201920200AB730. [Accessed 17 May 2021].

[59] H. Tankovska, "Twitter - Statistics & Facts," 04 May 2021. [Online]. Available: https://www.statista.com/topics/737/twitter/#:~:text=According%20to%20recent%20social%20media,active%20users%20(mDAU)%20worldwide.. [Accessed 17 May 2021].

[60] Delbius, "Help us shape our approach to synthetic and manipulated media," twitter, 11 Nov 2019. [Online]. Available: https://blog.twitter.com/en_us/topics/company/2019/synthetic_manipulated_media_policy_feedback.html. [Accessed 17 May 2021].

[61] Cristian Canton Ferrer, Brian Dolhansky, Ben Pflaum, Joanna Bitton, Jacqueline Pan, Jikuo Lu, "Deepfake Detection Challenge Results: An open initiative to advance AI," facebook, 12 Jun 2020. [Online]. Available: https://ai.facebook.com/blog/deepfake-detection-challenge-results-an-open-initiative-to-advance-ai/. [Accessed 17 May 2021].