Journal of Science and Engineering Research
Vol. 2, No. 1, 2023, pp.29-40
eISSN: 2957-7187

Multimedia Video Generation and Detection using DeepFake with Deep Learning Techniques

Rahul Saha

Department of Computer Science, University of Padua, Italy
rahul.saha@unipd.it

DOI: http://dx.doi.org/10.56828/jser.2023.2.1.4
***Article history**: Received (January 7, 2023); Review Result (February 10, 2023); Accepted (April 14, 2023)*

Abstract: As technology has developed since then, it has provided cyberspace with resources that are exclusively available to computers, such as the capacity to create fictitious media. A flawless DeepFake media is created as a result of this process. Complications are used to affect the modifications. Because they're done in this manner, they're not apparent to the human eye. Making an algorithm that can automatically detect this kind of internet-based manipulation, on the other hand, is quite doable. Artificial intelligence is gaining popularity as robots take on more and more tasks. Because it is always learning new things, it grows smarter over time. A large number of new models are being created to increase the quality of DeepFakes, making it more difficult to distinguish between genuine and morphing materials.

Keywords: DeepFake media, Deep learning technique, Video manipulation technique, Artificial intelligence (AI), Multimedia

1. Introduction

Deepfake is an Artificial Intelligence (AI) [1] based video manipulation technique that involves digitally altering a person's face or body to mimic someone else with the intent of malicious use or harm through the dissemination of false information, as defined by the National Institute of Standards and Technology [2]. It was in April 2018 when a one-minute video clip of former United States President Barack Obama went viral, in which he was shown discussing themes on which he had never previously remarked [3]. The film was easily spotted as a fake by experienced specialists in the information technology area, but it was not perceived as such by those from the uneducated lower classes up to and including middle-class individuals [4]. The advancement of artificial intelligence, deep learning, machine learning, and image processing technologies has made the creation of deep fake films incredibly simple and realistic.

DeepFake 2 are created by training auto-encoding [5] algorithms over a vast collection of data while simultaneously compressing the resultant data [6] into save points known as data points, which are then used to create the fakes [7]. The difference is that the second method operates on the target video rather than on the source video. Both the target and the source then swap these save points to achieve a flawless lip-to-chin [8] effect, resulting in a successful DeepFake [1]. It is becoming increasingly easier to integrate and deploy this technology over time, thanks to the use of apps such as Faceup that make this possible [9].

Figure 1: Showing (a) the image of a normal person and (b) Images created using the DeepFake

These models analyze the facial expressions, features, contours, blood flow movements [10], and other characteristics of the source and synthesis the face characteristics of the target [11], resulting in lifelike facial expressions and positions that move. For most DeepFake algorithms to produce a naturalistic output, a large amount of source data is required for training and processing models [12], and this data must be collected in large quantities [13]. Because there are so many photographs of celebrities and famous people on the internet [14], it is extremely usual for these figures to become the primary targets of this type of assault. Consequently, there is an urgent [15] need for counteractive measures to be put into place so that fraudulent data sources can be identified and traced back to their sources of origin.

2. Background and Related Works

Deep learning-based solutions have been used to solve a variety of difficult issues, ranging from computer vision to human-level control [16]. Advances in deep learning are being used to build software that jeopardizes privacy, democracy, and national security. The deep learning program DeepFake was just released. DeepFake algorithms may create fictitious pictures and movies that fool the human eye. As a result, automatic detection and evaluation of the integrity of digital visual content is essential. This article goes through the strategies used to make DeepFake, as well as the detection techniques utilized in earlier research. We go through DeepFake problems in depth, as well as current research achievements and prospects. This research offers a thorough examination of DeepFake approaches, paving the way for the creation of new, more robust ways of dealing with increasingly difficult DeepFake.

Image falsification has grown in popularity with the introduction of DeepFake videos [17]. DeepFakes are created by replacing a person's face, facial expressions, and voice with those of another person. In these flicks, the editing is nearly impenetrable. They have a major influence on both people and society. Defamation, extortion, and character assassination are all common uses of social media platforms. On social media platforms, there has been minimal effort put toward detecting DeepFake videos. The identification of DeepFake videos is the first step in preventing them from being spread on social media. The proposed new technique for identifying fraudulent movies based on neural networks that use a video frame extraction method to speed up DeepFake identification. The method uses a CNN and a classifier network to create a model. The "Exception" network has replaced the "InceptionV3" and "Resnet50" networks. This is a visual artifact detection method. The CNN module

provides feature vectors to the classifier network to categorize the video. This method is capable of accurately detecting DeepFake videos on social media while using little computing resources. Using the Face Forensics++ and Deepfake Detection Challenge datasets, it was able to obtain 98.5 percent accuracy. Every video generated by an auto encoder is recognized by our model. Almost all phony films, according to our approach, include a large number of crucial frames. When depending on a single crucial video frame, we report on the accuracy. Anyone can check the validity of a film due to its simplicity. The social and financial repercussions of fraudulent films posted on social media are the only subject of our investigation. We do not need to train the model with a huge quantity of data for this study. In contrast to previous studies, our key frame extraction technique is very efficient.

Artificial intelligence-generated fraudulent videos and audio logical audios have cast doubt on the trustworthiness of records and audio as definitive evidence of events in certain instances (also known as deep fakes). Some of these challenges are discussed in this article, as well as research prospects in this area [18].

DeepFake videos are made by superimposing pictures and video samples to create a convincing-looking spoof. The main problem is the widespread dissemination of sexual material through DeepFake videos. While the bulk of these films are sexual, there are a few that are just hilarious. Celebrities and other well-known people's faces have been grafted onto pornographic artists' bodies. This technique undermines video evidence's credibility, making it unfit for use in court. This article covers technology's current and future capabilities, the significance of treatment planning, and technology's current and future effects on video evidence authentication. With technological advances, new parallel technologies will be required to aid in the detection and exposure of fraudulent films [19].

This is a method that takes use of artificial intelligence, such as Neural Networks (NNs). Realistic human face swaps are generated using neural networks. A combination of artificial intelligence and trained algorithms are being used to spread disinformation, circumvent privacy, and conceal the truth. Social media users' reputation has been trashed in an attempt to discredit famous faces. Candidates, corporations, and celebrities are often targeted by DeepFake. For DeepFake movies of politicians, a novel method for detecting DeepFake films is suggested. The faked video is used to train a depth-based long short-term memory model, which is then fed into a convolutional long short-term memory model to identify fake frames. The suggested method was also studied using our latest ground truth dataset of faked videos. Our plan has worked very well [20].

DeepFakes, a growing issue, have lately become very common on the internet. They're meant to show the dangers of intrusions of privacy, falsification of documents, and so on. DeepFake films may be difficult to tell apart from real-life footage at times. Researchers will need to develop techniques for detecting them. We provide the most convincing evidence pointing to the possibility that photographs were exchanged for each other's faces in this post. This study aims to provide a technological solution that can detect if a picture has been digitally altered using DeepFake technology [21].

These films were produced utilizing contemporary machine-learning methods to ensure maximum realism [22]. This may have a significant impact on how you get digital information. With an emphasis on DeepFakes in particular, this article will try to uncover the most sophisticated and deep learning technique utilized by academics to date. While considerable effort was made to ensure that the videos were genuine, the algorithm provides data based on the geographical and temporal features of the recordings. The writers additionally provide an architecture that makes use of low-level features and regions of

interest to enhance the chosen frame even more. The test precision of a technique developed for the 470 GB DeepFake Detection Challenge dataset was 97.6%.

3. Materials and Methods

3.1. State of Art

The techniques used to differentiate between a false video produced by artificial intelligence (AI) and a genuine video have been examined in this study, and the results have been presented. We have analyzed and tested a variety of distinct techniques, each with its own set of characteristics, on a variety of publicly accessible DeepFake media, all of which show the efficiency and efficacy of each algorithm in real-world situations. The validation of the model would be the primary emphasis of the first stages. As a result of comparing different detection models and exploring their limitations, we were able to identify models that were able to surpass the constraints.

We have also described the fundamental design that nearly every model has followed as a result of our observations. The development of certain DeepFake Detection techniques and models is still in its early stages; in addition, several methods have been suggested and tested, albeit only on fragments of data sets. DeepFake technology is always evolving, and each new algorithm introduces a new branch of vulnerabilities that must be identified and neutralized repeatedly to remain effective.

DeepFake may be created via the use of a variety of methods and implementations. To function properly, most excellent Generators rely on GANs (Generative adversarial networks), which need the involvement of two distinct neural networks that compete with and against one another. This appropriate Police and Thief pursuit increases the likelihood that the product will be on a more realistic level. GANs are an intriguing choice since they are capable of autonomously generating pictures and features without the need for human intervention. The usage of simple applications such as FakeApp makes it simple for even a novice to utilize and abuse the frameworks of DeepFake, assuming that the novice understands how the frameworks operate. In exchange for providing just the target and source image of the data, the program automatically downloads the source and target data, as well as pre-processing and filtering the pictures. In addition to Face Forensics++ and Face2Face, many more techniques and algorithms operate in a similar manner, such as FaceSwap and FaceSwap++. Using features such as skin color, eye tracking, and lip tracking, among other things, each of these techniques and technologies works by matching features between the target and the source materials.

3.2. DeepFake Detection

DeepFake are a secondary cause for data privacy issues, and as a result, techniques to detect deep fakes were brought to the public's notice almost as soon as the issue of fraud was identified. Earlier methods relied on human edging edge by edge, but recent advances in AI have enabled automation of feature extraction and delivery, as well as the delivery of features. Unassuming DeepFake discovery models began off as a binary classification issue that necessitated the use of huge datasets that included both source and target data. However, all that is required today is the extraction of characteristics from internet sources.

In Table 1, the first step in deepFake detection is face detection, which can be accomplished using a variety of methods, including knowledge-based methods, in which rules

are formed by the researchers based on personal knowledge, feature invariant methods, which can be used to detect when the orientation, pose, or angle of the face is varied, and template matching methods. The template matching technique makes use of the edge contours of a basic face shape or appearance-based techniques, which may be utilized to distinguish between face and non-facial pictures by classifying them as such. Deep learning techniques, which serve as the foundation for the deep fake detection software, may be used to extract features from faces after they have been detected using a face recognition algorithm were shown in Table 2.

Table 1: Showing the additional methods and dataset usage for DeepFake detection

Methods		
	Key Features	Data Sets Used
Using spatial and temporal signatures	This is how the CNN model gets features and does audio embeddings. It stacks a lot of modules and uses loss functions like KullbackLeibler divergence, among other things.	FaceForensics+ dataset and 5,600 deepfake audio & video datasets.
Using audio-visual affective cues	Face and voice modality as well as emotion embedding vectors can be found during training. This makes it easier to spot fakes.	TIMIT dataset and DFDC
Eye, teach and facial texture	As you watch this video, you'll see how facial texture, eye shape, and teeth shape change, as well as how reflections don't show up in deepfake. - The process of classifying is done with the help of neural networks and logistic regression tools.	A set of morphed videos downloaded from YouTube.
Using phoneme viseme mismatches	There are a lot of different dynamics and positions of the mouth shape, but the M, B, and P phonemes get a lot of attention in this method. Deepfakes often make it inefficient way.	Lip-sync deepfakes videos are created using artificial and synthesis techniques, i.e. (A2V) and (T2V)

In Table 3, the methods that are available to identify deep fake videos, as well as their limitations, are described in the following comparison table. All of the papers that we have mentioned so far have the same architecture; the only thing that distinguishes it as unique and efficient in comparison to other techniques is the feature that has been implemented. Deep fakes are corroding the uniqueness of media since seeing is no longer something that people can put their faith in anymore. It has been utilized in every technique and algorithm to include characteristics such as optical flow, face-warping artefacts, Heart Rate Variations, and other similar ones. High-definition movies, multiple face detection, and feature extraction models that take a long time to execute are some of the difficulties that the techniques and algorithms described above must overcome. Aside from the fact that deep learning's application sectors now include a wide range of domains of work, it has also paved the way for the development of intelligent automation systems capable of creating these repeated images.

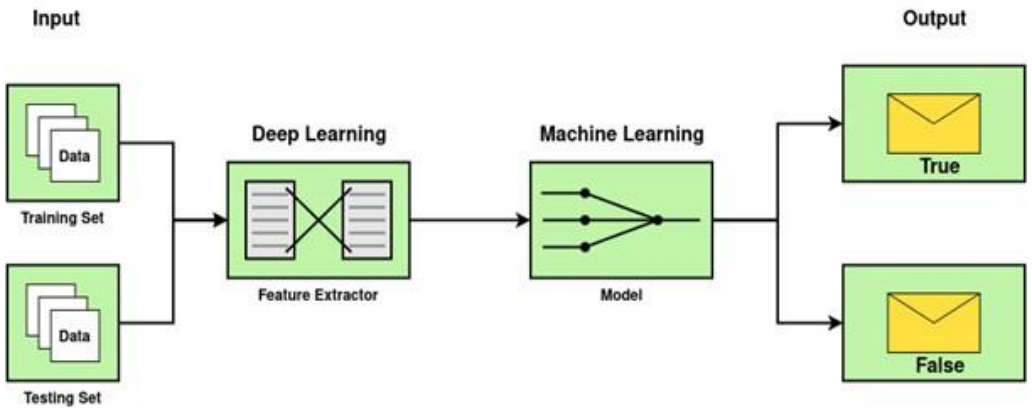

Figure 2: Proposed algorithm for classification of the DeepFake video

Figure 2 shows the essential architecture of DeepFake recognition classical. It is divided into three main parts:

1. The dataset into training and testing with 80% and 20%.
2. Deep Learning Models using Feature Synthesis.
3. Feature Synthesis was used for the model classifications.

Because this is a multi-classification model, the productions will either be completely correct or completely incorrect. There are YouTube videos included in our dataset. It contains both genuine and fabricated videos. Several other datasets are accessible, including the following: 1) The "DeepFake" video collection contains a lot of false information. This database exclusively contains movies in which the faces of the actors have been altered using an open-source GAN-based approach, which can be found here. Both sets of modified films are included in the dataset in their original sizes (64x64 and 128x128). Some films have a cast of around 32 people. Each subject receives an additional 10 videos in which their faces have been modified or replaced. Each of the videos has a resolution of 512 by 384 pixels and lasts around 4 seconds. A second dataset is the UADFV DeepFake video dataset, which contains around 100 videos, 50 of which are authentic and 50 of which have been altered. Each video is around 12 seconds in length and only discusses a single topic. We're looking into the DeepFake video-creation process to determine whether any faults may be exploited in the future. For time sequence analysis, the initial step is to get frame-level features via the usage of a CNN, which is then used to train an RNN. Using machine learning, it will be possible to train the model to distinguish between genuine and false movies by comparing the properties of video frames created by the model to the features of actual video frames. When determining if a video is genuine or fabricated, a classification model is utilized. The model produces a result that is either true (fake) or false (genuine) (original video).

4. Results and Discussions

To evaluate the efficacy of different classification algorithms, we perused a huge number of reference books and soaked in as much knowledge as we could on the best ways to find and produce DeepFake. We were able to derive a fundamental model from this. As a consequence of this, this model takes use of support vector machines in addition to logistic regression models, multi-level perceptron models, and a random forest classifier. As can be

shown in Table 4, all of them can learn quickly and make correct predictions, even when given just a limited amount of data. This specific model was chosen as the best option in light of these considerations.

Table 4: Showing the performance of the models

Classifiers	Problem Type	Train Speed	Predict Speed	Interpretability	Performance	performance with low data
Support Vector Machines	C	Moderate	modest	modest	moderate	high
Linear Regression	C	Exponential	dissolute	dissolute	lower	high
Random Forest	Both	short	modest	modest	high	low
K-Nearest	Both	moderate	moderate	low	moderate	low

Starting with data cleaning, we were able to improve the usability of our datasets by removing missing values and other extraneous information from them. The label distribution was examined to determine whether or not our dataset was balanced, as well as how it was balanced. As seen in Figure 2.

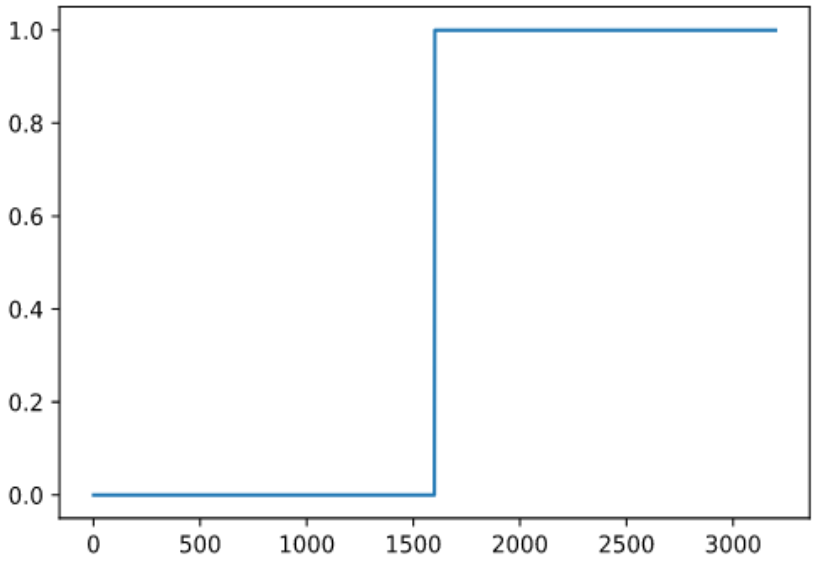

Figure 3: Label distribution of DeepFake video

Next, we segmented our dataset into training and testing sets as the next stage in the process. In this study, we tested a large number of classification models using three-fold cross-validation. AUC = 0.850, Logistic: f1 = 0.787 %.

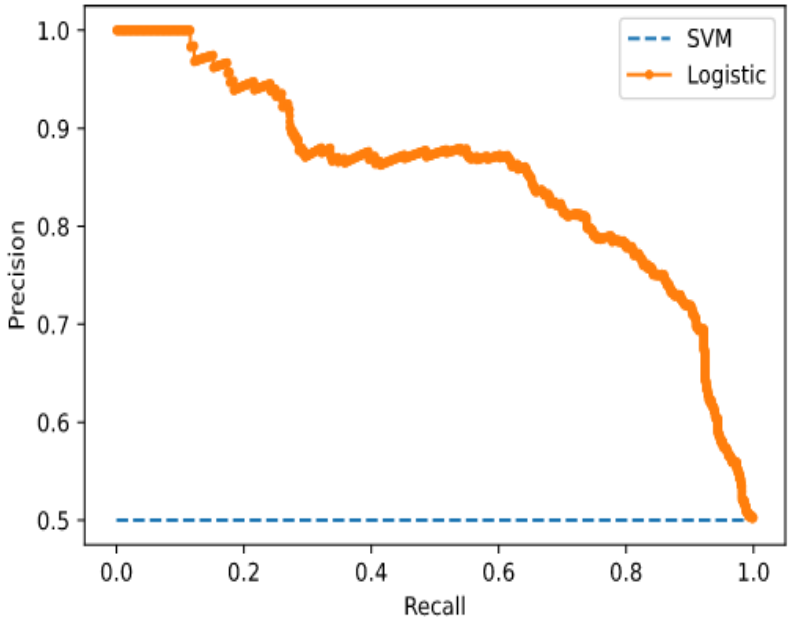

Figure 4: Recall-precision graph of the proposed classifier

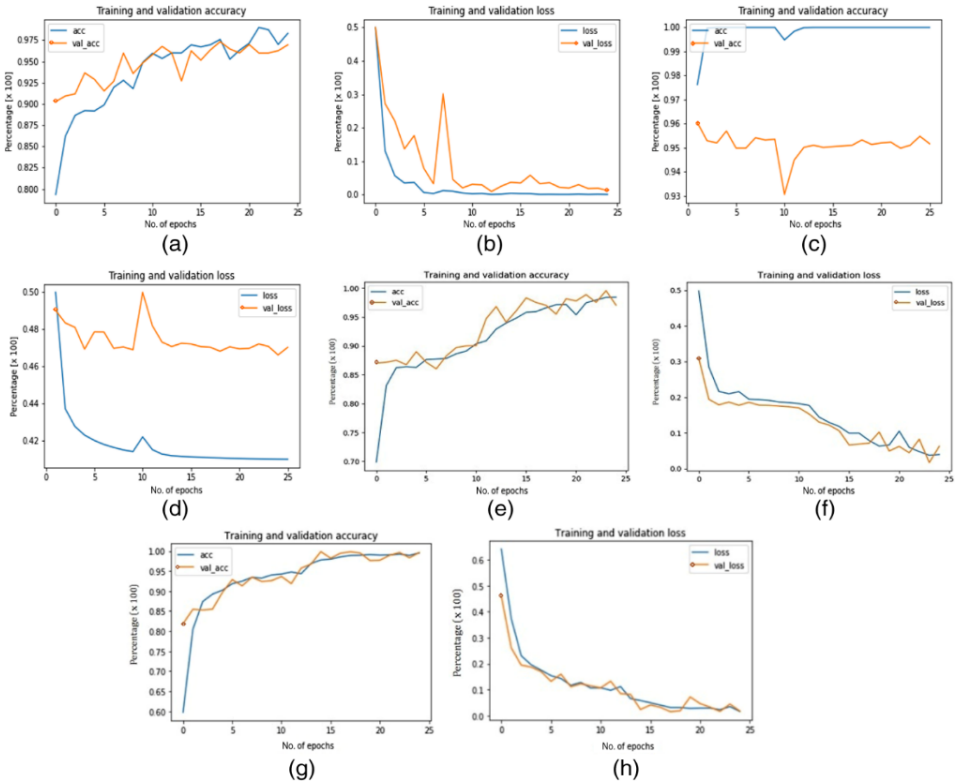

Figure 5: Showing the training and testing accuracy on various levels of parameters

According to the findings of this research, the most critical criteria in determining the accuracy and loss of a DeepFake video model are CNN training VS capsule model training, as well as validation accuracy and loss for the obtained DeepFake video dataset.

Table 5: Test accuracy per algorithm

Classifiers	Results	
	Accuracy	*Percentage*
Support Vector Machines	0.866	86.6
Linear Regression	0.788	78.8
Random Forest	0.753	75.3
K-Nearest	0.759	75.9

5. Conclusion

When determining if a video has been tampered with by Artificial Intelligence (AI), there are a variety of ways available, as we discovered throughout our examination. Different classification models and methodologies were put to the test using a wide variety of DeepFake Videos datasets, and the results showed that all of them worked very well in the real world. The validation of the model should be the major focus of the first phase in the process. Multiple detection models were compared and analyzed, and we were able to uncover models that may potentially work around the problems. Based on everything we have learned up to this point, we have focused our attention on the many different ways to detect a DeepFake. As an added benefit, it gives valuable guidance on how to approach a certain dataset. To develop a robust algorithm, you first need to construct a robust training set. As a direct result of our results, we have provided an overview of the fundamental structure that the vast majority of models have adhered to. For DeepFake technology to function properly, there are always going to be brand-new challenges that need to be uncovered and addressed. To stay up with the improvement in the quality of DeepFake, the algorithms that are used to recognize them must also improve. From our perspective, criminals might potentially be using deepfakes to harm the image of AI by spreading false information about the technology (AI). By using Artificial Intelligence (AI) and developing detection models that are based on AI, our goal is to educate people about the potential advantages that may be brought to them by technology. Researchers working in forensics and security may benefit from this study.

References

[1] Jung, T., Kim, S., & Kim, K. (2020). DeepVision: Deepfakes detection using human eye blinking pattern. In IEEE Access, 8, 83144-83154. DOI:10.1109/ACCESS.2020.2988660.

[2] Joshua, E. S. N., Bhattacharyya, D., Chakkravarthy, M. & Byun, Y. –C. (2021). 3D CNN with visual insights for early detection of lung cancer using gradient-weighted class activation. Journal of Healthcare Engineering. DOI:10.1155/2021/6695518.

[3] Joshua, E. S. N., Chakkravarthy, M., & Bhattacharyya, D. (2020). An extensive review on lung cancer detection using machine learning techniques: A systematic study. Revue d'Intelligence Artificielle, 34(3), 351-359. DOI:10.18280/ria.340314.

[4] Park, T., Liu, M. Y., Wang, T. C., & Zhu, J. Y. (2019). Semantic image synthesis with spatially-adaptive normalization. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[5] DeepFaceLab: Explained and Usage Tutorial. Available at: https://mrdeepfakes.com/forums/thread-deepfacelab-explained-and-usage-tutorial.

[6] Jafar, M. T., Ababneh, M., Al-Zoube, M., & Elhassan, A. (2020). Forensics and analysis of deepfake videos. In The 11th International Conference on Information and Communication Systems (ICICS), 053-058, IEEE.

[7] Schroepfer, M. (2019). Creating a data set and a challenge for deepfakes. Available at https://ai.facebook.com/blog/deepfake-detectionchallenge.

[8] Rossler, A., Cozzolino, D., Verdoliva, L., Riess, C., Thies, J., & Niener, M. (2018). FaceForensics: A large-scale video dataset for forgery detection in human faces.

[9] Choi, Y., Choi, M., Kim, M., Ha, J. W., Kim, S., and Choo, J. (2018). StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation. InProceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[10] lang, C., Ding, L., Chen, Y., and Li, H. (2020). Defending against GAN-based deepfake attacks via transformation-aware adversarial faces pg 4.

[11] Kaliyar, R. K., Goswami, A., and Narang, P. (2020). Deepfake: Improving fake news detection using tensor decomposition-based deep neural network. Journal of Supercomputing, DOI:10.1007/s11227-020-03294-y.

[12] Tucker, P. (2019, March 31). The newest AI-enabled weapon: Deep-Faking photos of the Earth. Available at https://www.defenseone.com/technology/2019/03/next-phaseaideep-faking-whole-world-and-china-ahead/155944/.

[13] Fish, T. (2019, April 4). Deep fakes: AI-manipulated media will be weaponized to trick the military. Available at https://www.express.co.uk/news/science/1109783/deep-fakesaiartificial-intelligence-photos-video-weaponised-chi.

[14] Marr, B. (2019, July 22). The best (and scariest) examples of AI-enabled deepfakes. Available at https://www.forbes.com/sites/bernardmarr/2019/07/22/the-bestandscariest-examples-of-ai-enabled-deepfakes/.

[15] Zakharov, E., Shysheya, A., Burkov. (20193.)Few-shot adversarial learning of realistic neural talking head models," arXiv preprint arXiv:1905.08233. Damiani, J. A voice deepfake was used to scam a CEO out of $243,000, in 2019, Available at https://www.forbes.com/sites/jessedamiani/2019/09/03/avoicedeepfake-was-used-to-scam-a-ceo-out-of-243000/.

[16] Nguyen, T., Nguyen, C. M., Nguyen, T. (2019). Deep learning for Deepfakes creation and detection: A survey.

[17] Mitra, A., Mohanty, S. P., & Corcoran, P. (2021). A machine learning based approach for deep fake detection in social media through key video frame extraction. SN COMPUT. SCI. 2, 98. DOI10.1007/s42979-021-00495-x.

[18] Lyu, S. (2020). Deepfake detection: Current challenges and next steps. 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 1-6. DOI:10.1109/ICMEW46912.2020.9105991.

[19] Maras, M. –H. & Alexandrou, A. (2019). Determining the authenticity of video evidence in the age of artificial intelligence and the wake of Deepfake videos. The International Journal of Evidence & Proof, 23(3), 255-262. DOI:10.1177/1365712718807226.

[20] Kaur, S., Kumar, P., & Kumaraguru, P. (2020). Deepfakes: Temporal sequential analysis to detect face-swapped video clips using convolutional long short-term memory. Journal of Electronic Imaging, 29(3), 033013. DOI:10.1117/1.JEI.29.3.033013.

[21] Maksutov, A., Morozov, V. O., & Lavrenov, A. A. (2020). Methods of Deepfake detection based on machine learning. 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 408-411. DOI:10.1109/EIConRus49466.2020.9039057.

[22] Singh, A., Saimbhi, A. S., & Singh, N. (2020). DeepFake video detection: A time-distributed approach. SN COMPUT. SCI., 1, 212. DOI:10.1007/s42979-020-00225-9.

This page is empty by intention.