A Comprehensive Study on Multimedia DeepFakes

Conference Paper · March 2023

DOI: 10.1109/ICAECCS56710.2023.10104814

4 authors:

Amal Boutadjine
University of Ferhat Abbas -Setif1
2 PUBLICATIONS **3** CITATIONS
SEE PROFILE

Fouzi Harrag
Ferhat Abbas University of Setif
54 PUBLICATIONS **586** CITATIONS
SEE PROFILE

Khaled Shaalan
British University in Dubai
380 PUBLICATIONS **11,944** CITATIONS
SEE PROFILE

Sabrina Karboua
Ferhat Abbas University of Setif
2 PUBLICATIONS **3** CITATIONS
SEE PROFILE

A comprehensive study on multimedia DeepFakes

1st Amal Boutadjine
Computer Science Departement
Ferhat Abbas University Setif 1
Setif, Algeria
boutadjine.amal@univ-setif.dz

2nd Fouzi Harrag
Computer Science Departement
Ferhat Abbas University Setif 1
Setif, Algeria
fouzi.harrag@univ-setif.dz

3rd Khaled Shaalan
Faculty of Engineering and IT
The British University in Dubai
Dubai, United Arab Emirates
khaled.shaalan@buid.ac.ae

4th Sabrina Karboua
Computer Science Departement
Ferhat Abbas University Setif 1
Setif, Algeria
sabrina.karboua@univ-setif.dz

***Abstract*—Since the entry of so-called DeepFakes in the development of fake multimedia, this late has marked a turning point and emerged as a major issue, although visual and aural media manipulations date back to the beginning of media itself. Thanks to this technology, the detection of altered and generated material has recently received more attention since the human ability to identify DeepFakes has significantly been far less effective than that of deep learning models. Organizations need to be ready as there are countless ways to deceive using convincingly altered photos, videos, and audio, such as perpetrating fraud, damaging reputations, extorting money or influencing public opinion during elections, which undoubtedly impacts society. In this regard, there is a critical need for automated solutions that can identify fake multimedia material and prevent the spread of dangerous misinformation. This article aims to give a comprehensive review of DeepFakes and a summary of the technology that underpins it. We provide information on various DeepFake detection algorithms, identify potential dangers of this frightening modern phenomenon, and highlight future research challenges.**

***Index Terms*—DeepFake, artificial intelligence, deep learning, DeepFake detection.**

I. Introduction

DeepFake multimedia, or DeepFakes, is one of the recent techniques used to generate manipulated media (images, videos, and audio) that look highly realistic to human eyes, created through powerful deep learning tools. DeepFake images consist of performing facial manipulations, such as identity swapping—where the target person's face is switched for another character in the original image—or by creating highly realistic-looking faces bearing identities that do not actually exist. In addition to the image synthesis-driven capabilities demonstrated by DeepFakes, the audio media area has also demonstrated impressive capabilities for generating synthetic data. Although they are automatically created by a computer via machine learning approaches or signal processing techniques, making Automatic Speaker Verification (ASV) systems vulnerable as they cannot easily identify them as rigged. Some studies have produced generated voice waveforms converted from source text using auto-encoders [4]. Other methods utilize generative adversarial networks (GANs) as the audio synthesizing component to produce high-fidelity audio by converting text to speech (TTS) [12]. DeepFake videos, on the other hand, depict individuals saying or acting in ways they never did. DeepFake videos, based on the incorporation of both image and sound modalities, require the combination of several methods in the deep learning field, and it has achieved significant improvements in creating highly realistic videos alongside paired speech. Whilst synthetic data can be generated by computer-graphics approaches or traditional visual effects from many years ago, Deep learning models are the more recent widespread underlying method for the generation of deep fakes. The field of computer vision has adopted a number of deep learning networks, ranging from auto-encoders to GANs to solve various problems, and a wide variety of DeepFake algorithms that use GANs have been proposed for the purpose of duplicating a person's facial expressions and motions and swap them with those of another person. For deep learning models to be trained to create photo-realistic images and videos using DeepFake techniques, a huge amount of picture and video data is typically needed. Therefore, the fact that celebrities and politicians typically have a lot of films and pictures available online makes them popular targets for DeepFakes. In practical real-world applications, this technology has both beneficial and detrimental effects. Although it can be utilized in virtual reality, video games, and movie productions, it is nevertheless often exploited for malevolent purposes like fabricating fake news, fraud, pornography for blackmailing, and violating privacy by creating images, voice clips, or videos of people without their consent. The alarming issue of online disinformation has recently taken a new turn thanks to this emerging AI threat with a substantial capability for manipulating public opinion with the potential to make fake news harder to spot with the naked eye. Over

Mechatronics Laboratory, Optics and Precision Mechanics Institute, Ferhat Abbas University Setif 1, Setif, Algeria.

978-1-6654-6310-2/23/$31.00 ©2023 IEEE

time, it might also undermine trust in journalism, including reputable sources. Therefore, there is a growing need for developing algorithms that are capable of effective detection of such technology. With the rapid spread of this threat among the public, typical citizens are growing more and more conscious of the potential harm posed by such a technology. In addition, news media, law enforcement organizations, and even Governmental bodies and other public entities are all becoming more aware of this as well. Thus, the development of reliable detectors that are able to automatically detect fake multimedia is a challenging task for the scientific community. Philosophers as well as scientists are concerned about the impact of such technologies and whether they will improve our health, feeling of social connection, and overall well-being [22]. The new generations tend to lose the ability to distinguish between real life and fake life as they become increasingly immersed in the digital world, believing that the world can be divided into followers and being followed [8], which creates significant problems that need to be addressed and resolved. The main tenet of this paper has been divided into five sections, including this introduction. First, Section 2 presents commonly used techniques for DeepFake generation. Section 3 is devoted to different methods of DeepFakes detection. after citing the good and the bad impact of DeepFakes in section 4, we discuss the threats, together with the rising challenges of DeepFake recognition in section 5. Section 6 concludes by reiterating the main points of this paper.

II. GENERATING DEEPFAKE MEDIA CONTENT

With the aid of the DeepFake algorithms, a user can photo-realistically replace a video's main actor's face with that of another actor. New manipulation algorithms, the majority of which are based on generative networks, are frequently proposed since the initial release of DeepFake movies. These techniques reveal a significant negative impact on society as DeepFake algorithms can be used to fabricate material and violate individual privacy. The technology of Face manipulation is not brand-new innovation that just emerged. The famous 1865 portrait of US President Abraham Lincoln contains the earliest effort at face manipulation in the literature. Editing StyleGAN latent codes to reflect the six common emotions—original, anger, disgust, fear, happiness, sorrow, and surprise— led to the creation of DeepFake images in Fig. 1. The remarkable advancement of computer vision algorithms has made Digital picture alteration an easy task that anyone with simple knowledge and skills can do. Existing DeepFake algorithms can be categorized in general into face swapping and face re-enactment, each of which serves a particular purpose in face manipulation. Currently, generative adversarial networks (GANs), autoencoders, and neural radiation fields (NeRF) are the three most popular deep learning-based techniques for creating synthetic human faces. Nevertheless, traditional CGI image generation techniques have been around since the 1970s. NeRF, is a late entrant that can also recreate the full human shape, yet it has the most primitive facial-generating skills of any of them. Autoencoder frameworks are mostly limited to the inner areas of the face and need host footage, which results in adding the challenge of selecting a target that is remarkably close to the injected identity. The most convincing faces can be produced by Generative adversarial networks, or GANs, a family of machine learning models that can create excellent deep fakes. Ian Goodfellow and his colleagues [11] created the GAN, which consists of two neural networks that compete with one another: the discriminator and the generator. Using the training image data set, the first network (the generator) creates fake images. While the discriminator in the second network tries to differentiate between real and artificially created images. The generator's goal is to deceive the discriminator into thinking the visuals are real. In this approach, as the discriminator becomes more adept at spotting AI-generated images, the generator creates ever-more realistic images in an effort to deceive them. During training, a Generative Adversarial Network captures high-level information from tens of thousands of photos in order to develop the capacity to duplicate similar images in the dataset's domain. Although DeepFakes initially focused on computer vision, it swiftly expanded to other fields, such as natural language text. Currently, lengthy documents are produced in which it is impossible to tell whether it was actually written by an AI program or a genuine human.

III. DETECTION METHODS

Open Internet-based platforms and programs like FaceSwap, FakeApp, DeepFaceLab, DeepNude and others, have democratized DeepFake algorithms and helped to promote their widespread production by making it surprisingly simple to produce high-quality DeepFakes [30], and better generative models make it harder to distinguish between DeepFakes and authentic content, therefore many studies tackled this problem in the few recent years. Nevertheless, even if DeepFakes are of very high quality, close examination reveals that the produced photos and videos contain certain artifacts that can be used to identify manipulated information. Using artifact hints from DeepFakes, several DeepFakes detection strategies have been put forth. Numerous methods focus on analyzing video inconsistencies, changes in temporal and spatial information [9], [19], and optical flow [7].

Numerous initiatives have been taken to enhance the effectiveness of face forgery detection, and several deep learning networks such as LSTMs, CNN [29], [30], RNN [9], in addition to capsule Networks have been used to address the DeepFake issue [3]. In order to detect obvious semantic distortions in DeepFakes, early face forgery detection systems like [2], [7] often use the standard pipeline of learning convolutional neural networks (CNN) for image classification. These techniques directly accept a facial image as input and then categorize it as real or fake using pre-built CNN backbones. The fact that these plain CNNs tend to only look for fakes on a small subset of faces, however, shows that forging is beyond the detectors' comprehension. With the goal of focusing on the images' mesoscopic characteristics, Afchar et al. [2] suggested two distinct CNN architectures with just a few layers: (a) a CNN made

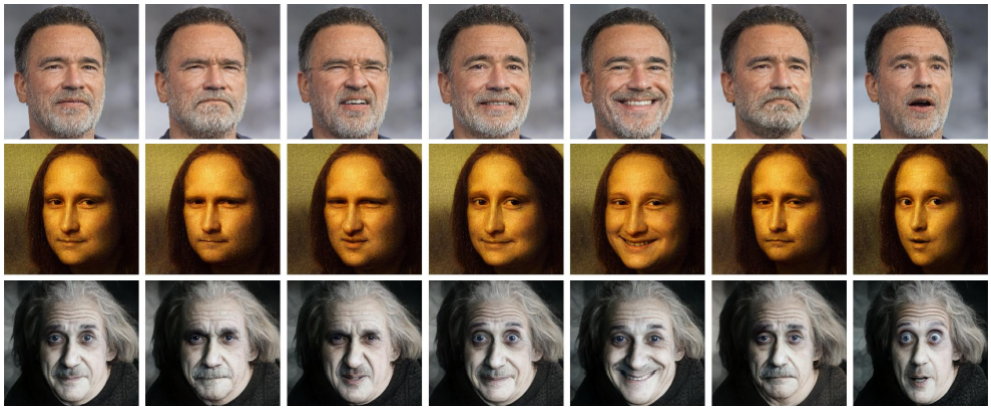

Fig. 1. DeepFake images generated by editing StyleGAN latent codes in the direction of the six prototypical emotions: original, anger, disgust, fear, happiness, sadness, surprise [22].

TABLE I
RECENT DEEPFAKE DETECTION ALGORITHMS.

No	Title of Paper	Year	method	Dataset	Best accuracy
[9]	Recurrent convolutional strategies for face manipulation detection in videos.	2019	CNN-RNN(GRU)	FaceForensics++ (FF++)	96.9%
[7]	DeepFake video detection through optical flow based CNN	2019	CNN with optical flow	FF++	81.61%
[13]	Effective and fast DeepFake detection method based on Haar wavelet transform	2020	Haar Wavelet Transform	UADFV	90.5%
[20]	Interpretable and trustworthy DeepFake detection via dynamic prototypes	2021	DNN	FF++ Celeb-DF	98.26%
[23]	FakeLocator: Robust localization of GAN-based face manipulations	2022	GAN based: Gray-scale fakeness prediction map; encoder-decoder architecture with attention mechanism		99.95%
[26]	M2tr: Multi-modal multi-scale transformers for DeepFake detection	2022		FaceSh	65.56%
[27]	Combining efficientnet and vision transformers for video DeepFake detection	2022	EfficientViT	DFDC	95%
[6]	Protecting world leaders against deep fakes	2019	SVM (Head Pose and Facial Features)	Own (FaceSwap, HQ)	96.3%

up of four convolutional layers followed by a fully-connected layer (Meso-4), and (b) an adaptation of Meso-4 constructed using a special Inception module called MesoInception-4.The dual CNN-introduced model demonstrated a 98% detection rate on DeepFake videos.

Agarwal et al. [5] employed a statistical framework [1] based on the information-theoretic investigation of authenticity as a hypothesis testing problem for locating the DeepFakes. The first thing this approach does is determine the minimum distance between the original and the GAN-generated image distributions. This distance calculates the detection capability based on the findings of the hypothesis. For instance, DeepFakes are easily identified when this distance is increased. If the GAN delivers less accuracy, the distance will typically increase. Additionally, in order to produce modified images with high resolution that are more difficult to detect, a very exact GAN is required.

A DeepFake detection technique based on Capsule Networks was mentioned by the authors in [3]. They outline a system that consists of a CNN for feature extraction, several convolutional layers leading to the primary capsules, a stats pooling layer that determines the mean and variance of each filter, the primary capsules, then two output capsules for the two different classes (real and fake). Each network primary capsule gathers a different artifact, which is subsequently integrated through the use of dynamic routing. According to the likelihood generated by the output block, each frame of the video series is judged to be fake or real. This network learned features that are simple to visualize and performed well even with changes that were not visible. Different Capsule Network designs were employed in [3], with the best one achieving a 93.11 classification accuracy using just 3.9 million parameters. The temporal directional shifts in videos were

explicitly considered when designing another set of methods. Sabir et al. [9] suggested utilizing an end-to-end trained Recurrent Convolutional Network rather than a pre-trained model. FaceForensics++ dataset was used to evaluate their suggested detection method, and AUC results of over 96% were obtained.

Since The majority of DeepFake detection techniques are primarily focused on recognizing a specific artifact and rely on a precise approach, they are unable to generalize to manipulations in novel datasets. Nevertheless, given the nature of these systems' training, their specialty in finding various artifacts could be taken advantage of by using multiple detection system strategies, in this context, the authors of [21] presented two distinct fusion techniques: score-level fusion of three independent classifiers and feature-level fusion of face patches from several facial areas, and the basic average was applied.

Ruther than using deep neural network for DeepFake video detection, Younus and Hasan [13] proposed a method that compares the blurriness and sharpness of the edges in the facial region with the background boundary using the Haar Wavelet Transform. On the UADFV dataset, the accuracy obtained using this approach is above 90

A dynamic prototype well-designed deep neural network (DPNet) was recently created by Trinh et al. [20] to learn prototypical video patches from deep-faked and real videos, by adopting a more general approach than concentrating on a specific camera artifact and highlighting noise artifacts produced by the entire acquisition process, independent of their particular origin. Because different cameras exhibit varying inherent noise, the measurement of local noise level could be used to identify splicings, as demonstrated in [20]. In [23], the FakeLocator method, a technique for the detection and localization of GAN-based face alterations is suggested. What motivates the authors' study is the observation that GAN-based face alterations introduce artificial textures to the image when low-quality images are upsampled to high resolution. To identify the modified areas in the input image, the authors create gray-scale maps using a GAN. Ground truth fakeness maps are created first. In order to produce gray-scale fakeness prediction maps, the encoder-decoder architecture networks are implemented with the attention mechanism, to input actual or fake images. Finally, the loss is computed by combining the ground truth fakeness maps with the gray-scale fakeness prediction maps. However, the application of this technology is limited to GAN-based face-altering techniques and cannot be utilized to identify fake faces or expressions.

Computer vision has made substantial use of multi-scale feature representation. Vision Transformer hasn't yet completely realized the potential of this usual arrangement. The authors of [26] recently developed the M2TR, a multi-modal and multi-scale transformer-designed feature representation based on different patch sizes to exploit both spatial and frequency domain artifacts. Multiple ViTs with various spatial embedding sizes were stacked in [26]. Based on ViT, CEViT [27] also tackles the problem of video DeepFake detection by integrating different kinds of Vision Transformers with a convolutional EfficientNet B0 utilized for feature extraction. To determine if the video taken was edited or not, the faces from various frames of the video are individually examined, sorted, and voted on at inference time. The suggested method makes it feasible to handle scenarios like the existence of multiple people in a video where just one has been altered, for example, more effectively. These techniques showed encouraging generalization ability in cross-database evaluation, however, the CNNs and Vision Transformers combination had a high computational complexity because of the huge number of parameters used.

A face forgery detection approach using frequency cues is proposed in certain publications, such [31], which observe the diversity of actual faces and false faces in the frequency domain.

In [6], a different methodology is used. The work aims to secure individuals by giving them soft traits that define them and are extremely difficult for a generator to imitate. Particularly, head motions and facial expressions were found to have a high correlation, and that changing the former without affecting the latter could reveal a manipulation. By capturing the peculiar behavioral traits of a specific individual, the time series of facial landmarks extracted from real films are used to identify DeepFakes. The drawback of this strategy is that it requires that all interested parties have access to a sizable and varied library of videos in a variety of settings.

In order to free them from dependence on one or more particular DeepFake production techniques, Coccomini et al. [28] put their focus on the difficulty of generalizing DeepFake identification approaches.

In order to categorize synthetically produced falsified audio, Khochare et al. [29] examined feature-based and image-based methods. This work made use of two brand-new Deep Leaning models, the Temporal CNN (TCN) and the Spatial Transformer (STN) via mel-spectrograms. While STN only managed an accuracy of 80%, TCN distinguished between phony and real audio with a 92 percent accuracy. The TCN works well with sequential data, but it cannot handle inputs with Short-Time Fourier Transform (STFT) or Mel Frequency Cepstral Coefficients (MFCC) features. These image-based algorithms outperformed feature-based ones that made use of features for frequency, bandwidth, energy, and short-term transform features like MFCCs for the detection of synthetic audio.

Table 1 analyzes the various techniques and features utilized for DeepFake detection. It comprises methods based on machine learning and deep learning. As it can be observed from this analysis table, Deep neural networks with attention mechanism produce superior outcomes and accuracy.

IV. DOUBLE-EDGED UTILISATION OF DEEPFAKES

A. The advantages

The entertainment aspect is one of the most important goals of DeepFakes. This synthesis technology is often used to produce a variety of YouTube videos and memes featuring famous people with their faces swapped solely for comedic

effects. Other applications of these techniques include the placement of puppet politicians' figures in various comedic and less amusing demonstrations [18]. In fact, the reenactment of historical events in classrooms is another way that deep fake may be utilized to educate students. This provides them with a dynamic, hilarious, and imaginative way of seeing, participating, and learning from history. Creativity is another driving force behind deep fakes; it can be applied to produce realistic museum displays, artwork, songs, and poetry. In films and ads, long-dead actors can also be brought back to life [8]. Hollywood can be a plight of DeepFakes (such as facial aging or de-aging utilized in many blockbuster movies) which can be employed to alter the appearance of older or younger actors [18].

B. The drawbacks

Despite increased knowledge and the introduction of financial incentives, studies such as [17] have demonstrated that consumers are unable to accurately identify deep fakes. As privacy is one of the most important risks of DeepFakes, additionally, the darkest makeup on a celebrity's face was applied against their will and in highly unethical circumstances [18]. While these traps have been the subject of ongoing philosophical scrutiny and are deployed with active consent, serious ethical problems still exist [14]. In Japan, the defendant was detained in the first criminal digital case involving deep fakes for utilizing artificial intelligence to improve the resolution of pornographic content films [15]. More than ever, it will be difficult to tell the difference between the real and the fake [16], where the virtual world is leaking into the real world through AI-synthetic content [14], as deep fake technologies become more adept at producing artificially manipulated visual content on the one hand and social media continue to take up more space in our lives on the other. The antagonistic ideologies and ideological movements that frequently populate the increasingly digital and hyper-visual environment in which we live [16] employ DeepFake to further their goals and disseminate propaganda. Through the weaponization of synthetic media, deep fakes can likewise endanger national security and the interests of nations [10].

V. THREATS OF DEEPFAKES AND RISING CHALLENGES OF DETECTION

Nowadays, many DeepFakes on social media platforms can be viewed as light-hearted humour or creative works utilizing famous people, both alive and deceased. However, there are also misuse cases of this powerful technology. DeepFakes cause cyber-security concerns for individuals, society, the political system, and the economy because of their ability to fake reputable, high-quality media content. A person's face in a video or voice in an audio file could once be reliably used to identify them, but this is no longer the case with the massive advances in DeepFake technology. Biometric systems that were once thought to be safe can now be easily defeated. Media content can particularly fool remote identification techniques like video identification or voice recognition over the phone. DeepFakes endanger even national security by compromising elections, disrupting candidates' campaigns, spreading propaganda, and it is even used in wars to spread rumors by showing politicians say things they haven't said [24]. Due to this technology, it is quite likely that the media sector will have to deal with a significant problem with customer confidence. Because DeepFakes are more difficult to detect and are now feasible even for laypeople who have a minimal understanding of technology, and individuals are more likely to assume the fake is real, these fake media offer a larger threat than classic fake news.

VI. RISING CHALLENGES OF DETECTION

Images produced by GAN can be quite convincing. It is disturbing how well neural networks are now able to model human faces. It is critical for humans to be able to distinguish between authentic faces and auto-generated ones given the high stakes involved for both individual and societal security. DeepFake detection, despite the accuracy of some detection systems, is a never-ending "arms-race" between detectors and attackers. When the training and test sets exhibit identical data distributions these approaches perform admirably [2], [7], [13], [20]. However, real-world data in practice are often unseen and differ from those in the training set regarding the unseen generation methods, pre-processing applied, the compression rate, the camera used, and more. These gaps cause drastic performance decreases, which restricts the scope of available applications and poses challenges to DeepFake fighters in the real world. Motion blur produced by networks is another specific issue with videos. During the creation of a movie, it may be challenging to track some objects, which causes some areas of the image to appear fuzzy. Although this problem has been addressed in certain publications like [13], it remains unresolved. When a lesser-known character is depicted in the video with just the altered version being available publically, it becomes considerably more challenging to validate the film's digital integrity. To boost the efficacy in combating the widespread influence of DeepFakes, another research challenge involves integrating detection techniques into distribution channels like social media.

Additional challenges, such as the Audio Deep Synthesis Detection Challenge 2022 [25], will promote more research in this area in order to protect against IA risks and DeepFake attacks.

VII. CONCLUSION

Thanks to developments in artificial intelligence, the technology for fabricating multimedia forensics is currently in full growth and will get far better in the years to come. The method and function of image video modification are changing as a result of the continuous advancement in the modern world. With the help of GAN models and Convolutional Autoencoders, anybody with a Smart device can now easily produce believable images and videos, a task that was previously exclusive to skilled experts. These attacks have a negative psychological, political, economic, and individual impact. With

carefully collected research articles, we have provided in this survey an overall review and analysis of the research effort for combating DeepFakes, the proposed techniques, along with generation methods and the social impact as well as the open challenges of the field, and we have discussed the conflict between the attackers and the detectors. The identification of DeepFakes will undoubtedly get increasingly challenging and major funding sources are supporting significant research projects, initiatives are being launched in an effort to reduce DeepFake technology's unfavorable consequences. Future AI-based countermeasures must concentrate primarily on creating more reliable, scalable, and effective solutions.

REFERENCES

[1] U. M. Maurer, "Authentication theory and hypothesis testing," IEEE Transactions on Information theory, vol. 46, no. 4, pp. 1350–1356, 2000.

[2] Afchar D, Nozick V, Yamagishi J, Echizen I (2018, Dec) MesoNet a compact facial video forgery detection network. In: 2018 IEEE international workshop on information forensics and security (WIFS). IEEE, pp 1–7

[3] Nguyen HH, Yamagishi J, Echizen I (2019, May). Capsule-forensics: using capsule networks to detect forged images and videos. In: 2019 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE, pp 2307–2311

[4] A. Polyak, L. Wolf, and Y. Taigman, "TTS skins: speaker conversion via ASR," CoRR, vol. abs/1904.08983, 2019. [Online]. Available: http://arxiv.org/abs/1904.08983

[5] Agarwal S, Varshney LR (2019). Limits of deepfake detection: a robust estimation viewpoint. https://arXiv.com/1905.03493.

[6] Agarwal, S., Farid, H., Gu, Y., He, M., Nagano, K., and Li, H. (2019, June). Protecting world leaders against deep fakes. In Computer Vision and Pattern Recognition Workshops (pp. 38-45).

[7] I. Amerini,L. Galteri, R. Caldelli, A. Del Bimbo, Deepfake video detection through optical flow based CNN. IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea (South) (2019), pp. 1205–1207. https://doi.org/10.1109/ICCVW.2019.00152

[8] v. Doorn, M. Duivestein, S., and Pepping, T. (2019). The synthetic generation. Sogeti Labs.

[9] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, P. Natarajan, Recurrent convolutional strategies for face manipulation detection in videos. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (2019), pp. 80–87

[10] Bonfanti, M. E. (2020). The weaponisation of synthetic media: what threat does this pose to national security?. Ciber Elcano, (57).

[11] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

[12] M. Binkowski, J. Donahue, S. Dieleman, A. Clark, E. Elsen, ´ N. Casagrande, L. C. Cobo, and K. Simonyan, "High fidelity speech synthesis with adversarial networks," in International Conference on Learning Representations, 2020. [Online]. Available: https://openreview.net/forum?id=r1gfQgSFDr

[13] Younus, M. A., and Hasan, T. M. (2020, April). Effective and fast deepfake detection method based on Haar wavelet transform. In 2020 International Conference on Computer Science and Software Engineering (CSASE) (pp. 186-190). IEEE.

[14] M. Cotton. (2021). Virtual Reality, Empathy and Ethics. Springer Nature.

[15] H. Montgomery. (2021, October 19) Man arrested for uncensoring Japanese porn with AI in first Deepfake case. https://www.vice.com/en/article/xgdq87/deepfakes-japan-arrest-japanese-porn?fbclid=IwAR0zE-mB2bOABgCY18AmC-FnlbDSNvp_UnEYKDoN9TYbgvmDAwWX0PhIV6s

[16] Zuev, D., and Bratchford, G. (2021). Visual sociology: practices and politics in contested spaces. Springer Nature.

[17] Köbis, N. C., Doležalová, B., and Soraperra, I. (2021). Fooled twice: people cannot detect deepfakes but think they can. Iscience, 24(11), 103364.

[18] Lanham, M. Generating a New Reality: from autoencoders and adversarial networks to Deepfakes, 2021.

[19] R. Durall, M. Keuper, F. Pfreundt, J. Keuper, Unmasking deepfakes with simple features. Retrieved 17 April 2021, from (2021) https://arxiv.org/abs/1911.00686

[20] Trinh, L., Tsang, M., Rambhatla, S., & Liu, Y. (2021). Interpretable and trustworthy deepfake detection via dynamic prototypes. In Proceedings of the IEEE/CVF winter conference on applications of computer vision (pp. 1973–1983).

[21] Tolosana, R., RomeroTapiador, S., VeraRodriguez, R., Gonzalez-Sosa, E., & Fierrez, J. (2022). DeepFakes detection across generations: analysis of facial regions, fusion, and performance evaluation. Engineering Applications of Artificial Intelligence, 110, 104673.

[22] Haas, René, Stella Graßhof, and Sami S. Brandt. "Tensor-based Emotion Editing in the StyleGAN latent space." arXiv preprint arXiv:2205.06102 (2022).

[23] Huang, Y., JuefeiXu, F., Guo, Q., Liu, Y., & Pu, G. (2022). FakeLocator Robust localization of GANbased face manipulations. IEEE Transactions on Information Forensics and Security.

[24] B. ALLYN. (2022, March 16). Deepfake video of Zelenskyy could be 'tip of the iceberg' in info war, experts warn. npr. https://www.npr.org/2022/03/16/1087062648/deepfake-video-zelenskyy-experts-war-manipulation-ukraine-russia

[25] Yi, J., Fu, R., Tao, J., Nie, S., Ma, H., Wang, C., ... & Li, H. (2022, May). Add 2022: the first audio deep synthesis detection challenge. In ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 9216-9220). IEEE.

[26] Wang, J., Wu, Z., Ouyang, W., Han, X., Chen, J., Jiang, Y. G., & Li, S. N. (2022, June). M2tr: multi-modal multi-scale transformers for deepfake detection. In Proceedings of the 2022 International Conference on Multimedia Retrieval (pp. 615-623).

[27] Coccomini, D. A., Messina, N., Gennaro, C., & Falchi, F. (2022). Combining efficient net and vision transformers for video deepfake detection. In International Conference on Image Analysis and Processing (pp. 219-229). Springer, Cham.

[28] Coccomini, D. A., Caldelli, R., Falchi, F., Gennaro, C., & Amato, G. (2022, June). Cross-forgery analysis of vision transformers and CNNs for Deepfake Image detection. In Proceedings of the 1st International Workshop on Multimedia AI against Disinformation (pp. 52-58).

[29] Khochare, J., Joshi, C., Yenarkar, B., Suratkar, S., & Kazi, F. (2022). A deep learning framework for audio deepfake detection. Arabian Journal for Science and Engineering, 47(3), 3447-3458.

[30] Kolagati, S., Priyadharshini, T., & Rajam, V. M. A. (2022). Exposing deepfakes using a deep multilayer perceptron–convolutional neural network model. International Journal of Information Management Data Insights, 2(1), 100054.

[31] Sun, K., Yao, T., Chen, S., Ding, S., Li, J., & Ji, R. (2022, June). Dual contrastive learning for general face forgery detection. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 36, No. 2, pp. 2316-2324).