DOI: 10.24996/ijs.2023.64.10.39

ISSN: 0067-2904

The Defensive Methods Against Deepfake: Review

Ebtehal Talib[1], Nidaa Flaih Hassan[2], Abeer Salim Jamil[3]

[1] *Ministry of Higher Education and Scientific Research, Baghdad, Iraq*
[2] *Department of Computer Science, University of Technology, Baghdad, Iraq*
[3]*Department of Computer Technology Engineering, Al-Mansour University College, Iraq*

Received: 28/5/2022 Accepted: 7/12/2022 Published: 30/10/2023

Abstract

Due to the spread of “Deepfake” in our society and the impact of this phenomenon on politicians, celebrities, and the privacy of individuals in particular, as well as, on the other hand, its impact on the electoral process as well as financial fraud, all these reasons prompted us to present a research paper dealing with this phenomenon. This paper presents a comprehensive review of Deepfake, how it is created, and who has produced it. This paper can be used as a reference and information source for the methods used to limit deepfake by detecting forgeries and minimizing its impact on society by preventing it. This paper reviews the results of much research in the field of deepfake, as well as the advantages of each method, as well as finding points of weakness such as complexity, time consumption, and the environment that is applied to it. Methods to avoid or reduce these obstacles are also suggested.

Keywords: Deepfake, Positive Defensive, Negative Defensive, Cryptography, Deep Learning.

الطرق الدفاعية ضد التزييف العميق: دراسة مسحية

ابتهال طالب خضير [1], نداء فليح حسن[2], عبير سالم جميل[3]

[1]وزارة التعليم العالي والبحث العلمي, بغداد, العراق
[2]قسم علوم الحاسوب, الجامعة, التكنولوجية , بغداد, العراق
[3]قسم, هندسة تكنولوجيا الحاسوب ,كلية المنصور الجامعة, بغداد, العراق

الخلاصة

نظرا لانتشار ظاهرة التزييف العميق (Deepfake) في مجتمعنا وتأثير هذه الظاهرة على السياسيين والمشاهير وتهديد خصوصية الأفراد بشكل خاص، ومن ناحية أخرى تأثيرها على العملية الانتخابية وكذلك الاحتيال المالي، كل هذه الأسباب دفعتنا إلى تقديم ورقة بحثية تتناول هذه الظاهرة. تقدم هذه الورقة مراجعة شاملة للتزييف العميق، وكيف يتم إنشاؤه ، ومن عمل على أنتاجه. يمكن استعمال هذه الورقة كمرجع ومصدر معلومات للطرق المستعملة للحد من التزييف العميق من خلال الكشف عن التزوير، وتقليل تأثيره على المجتمع من خلال منعه. من خلال مراجعة الكثير من الأبحاث في مجال التزييف العميق, تستعرض هذه الورقة النتائج التي توصلوا إليها ومزايا كل طريقة ، بالإضافة إلى إيجاد نقاط الضعف مثل التعقيد واستهلاك الوقت والبيئة التي يتم تطبيقها عليها. بالإضافة إلى الأساليب المقترحة لتلافي هذه المعوقات أو تقليلها.

*Email: cs.20.07@grad.uotechnology.edu.iq

1. Introduction

Deepfake first appeared in 2017 when one Reddit user posted a video showing some celebrities in sex scenes that were not real. Deepfake means the use of deep learning techniques to manipulate an image, audio, or video to publish something that is not real [1]. Many categories have been targeted by Deepfake, the most famous of which are actors, artists, and politicians, and they were targeted by integrating their faces into porn videos [2]. In the future, Deepfake will target evidence in court, political propaganda, blackmail, false news, and market manipulation [3]. Thus, Deepfake is considered a frightening phenomenon that can falsify facts and threaten society. And in 2018, it became easy to use this technology to spread fake facts; since then, Deepfake has made remarkable progress [4]. There are many research articles (about 570 articles, 54 articles for deepfake generation, and 216 articles for deepfake detection) directed at combating the phenomenon of deepfake, especially in the last three years, as shown in the figure below [5].

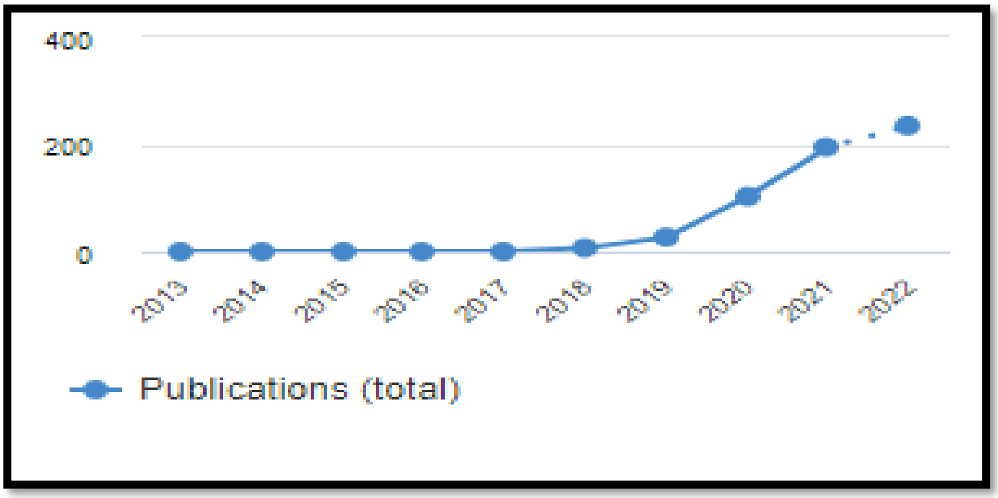

Figure 1: Number of Deepfake papers published 2012-2021**.**

To address this phenomenon, efforts must be directed towards ways that limit its spread and work to prevent it by using methods of cryptography and steganography to protect ownership and rights to publish and to prevent counterfeiting [6] [7] [8]. Since the deepfake has recently appeared in our society, studies and research in this field are considered to be few [1]**.**

The organization of this paper is as follows: After the introduction, the paper explains the concept of “deepfake” in Section 2. The paper then introduces sections that review deepfakes, deepfake producers in section 3, deepfake combat methods in section 4, and methods for creating deepfakes in section 5. Next, two sections are provided for anti-deepfake (detection in section 6 and prevention in section 7) and a discussion of these methods. After that, the results of some research are analyzed, and the most important recommendations are explained in the order addressed in Section 8. Finally, the study concludes with limitations and suggestions for future work in Section 9.

2. Deepfake Concept

Deepfake combines the terms "deep learning" and "fake," and its goal is to allow users to manipulate images, audio, and video to create fake facts that did not occur in reality by relying on neural networks that can deal with large data sets to train to mimic a person's behavior, facial expression, and way of speaking [1].The purpose of this is to feed a deep learning algorithm with data for two people and train them to change faces and voices [9]. That's why Deepfake is difficult to detect due to its use of deep learning techniques and real data. But deepfake technology can be considered a product of generative adversarial networks (GANs). It consists of two networks, the first one called "generator" and the second "discriminator," which are trained on the same set of data. This article provides a comprehensive deepfakes review and provides AI entrepreneurs and cybersecurity professionals with business opportunities in combating media fraud and fake news [1].

3. Deepfake Producers

Many deepfake producers have worked on its emergence for their purposes, some for fun and some for undermining the democratic process, such as presidential elections, some for fraud, and some for legitimate purposes, such as radio and television stations. There are four major producers of deepfake, as shown in figure 2:

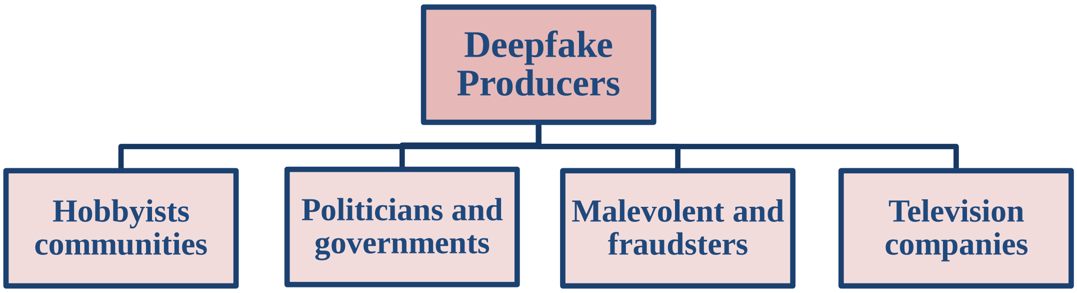

Figure 2: Deepfake Producers Types.

A. Deepfake hobbyist communities: Only a few months after one Reddit user faked a video of a celebrity in 2017, community members had become 90,000 enthusiasts for this newly established phenomenon. This community is divided into amateur forgers of pornographic videos and another group that works to trap actors and celebrities with artworks in which they did not participate in reality; some others consider it a form of humor, and some use it as a means to deceive and threaten people [1]**.**

B. Politicians and governments: Some governments and politicians use Deepfake to manipulate public opinion or reduce the trust of countries in certain state institutions and companies as a kind of war, as well as using it as a tool to undermine the democratic process, especially in elections [1].

C. Malevolent and fraudsters: They use Deepfake for financial fraud and stock market manipulation, for example, by faking the voice of an official to make a call intended to confirm the money transfer [10].

D. Television companies: They can be considered legitimate actors who practice deepfake [1]**.**

4. Deepfake Combat Methods

There are many harmful effects of Deepfake on society in general and individuals in particular; therefore, many legislative and legal institutions, in cooperation with research institutions, have found ways to limit the spread of this phenomenon or prevent its illegal use. By reviewing some articles, there are four methods to fight deepfake, as shown in figure 3:

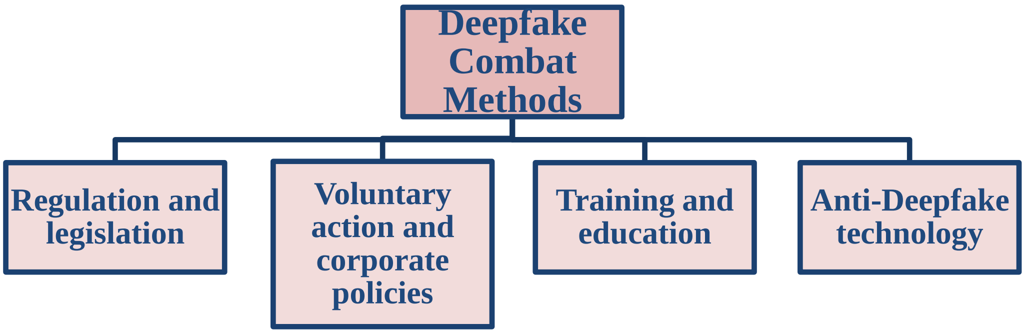

Figure 3: Deepfake Combat Methods

A. Regulation and legislation: Both of them work to reduce deepfake, but the current laws and legislation do not deal with it and do not consider it a felony or misdemeanor. Despite the spread of the phenomenon of impersonation and defamation of some well-known personalities, as well as publishing rights and privacy, where the law considers property and freedom of expression, some use Deepfake under the Freedom of Expression Act. The solution will be to enact new laws that prevent the harmful use of Deepfake and not prohibit the use of this technology in general [1].

B. Voluntary action and corporate policies: where more flexible tools are used to combat Deepfake, such as requiring politicians not to defame each other during elections using Deepfake techniques, or requiring social media site designers not to use forgery techniques to increase the number of followers or get more ads [10].

C. Training and education: This method is very important in the fight against Deepfake. The culture of distinguishing real from fake digital content must be spread through training and education because of the lack of awareness of technological development and inability to properly understand digital content, especially among children, as well as the spreading of that culture among the elderly and those with little experience with computer skills and digital development, as well as teaching digital literacy in schools [1].

D. Anti-Deepfake Technology: These are ways to combat Deepfake and are divided into two types: Deepfake detection (**a negative method**) and Deepfake prevention (**a positive method**) [1]. There are many opportunities provided by cybersecurity companies to combat fraud and digital forgery by detecting them. In cooperation with forensic medicine, they have identified indicators such as fluctuation and distortion of the face, the disappearance of certain features such as a mole, a tooth, etc., differences in lighting, a reflection, and a contradiction between mouth movements and spoken speech as defects to discover Deepfake using artificial intelligence techniques [11]. Some techniques, on the other hand, prevent the use of Deepfake by creating noise, such as in video clips or images. These technologies enable politicians and celebrities to protect the content of their digital content before it is published to ensure that it is not misused against them. Understanding the problem is the first step toward solving the Deepfake problem, according to cyber security [1].

We must find a mixture of solutions that combine technical solutions and legal legislation to ensure that the content is protected from tampering. This paper explains the negative and positive defensive methods for Deepfake.

5. Deepfake Creation

There is a great deal of diversity in the media targeted by Deepfake, and this has led to a diversity of techniques for Deepfake detection, as shown in figure 4. All this diversity has created great challenges for researchers to find new, fast, and high-accuracy methods in deepfake detection that target social media platforms and material published on the Internet, as will be clarified in the following paragraphs [12].

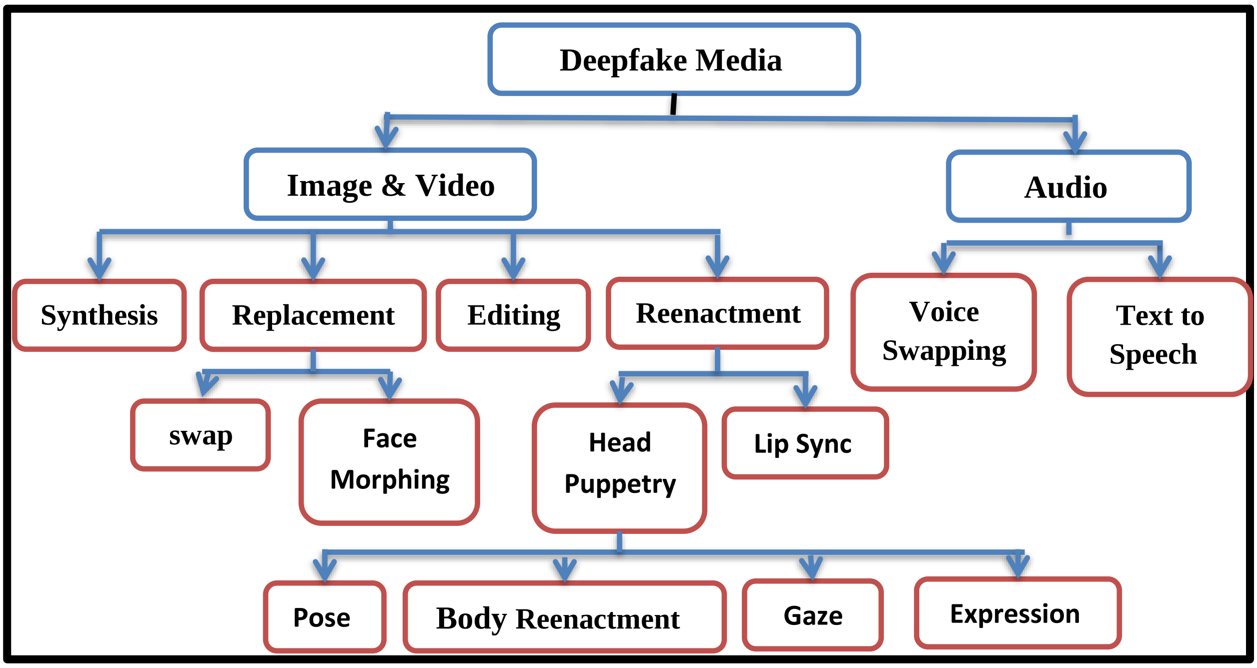

Figure 4: Current Deepfake Media Types and Detection Techniques.

This phenomenon has spread due to the increased availability of high-quality manipulated images and video clips, as well as the ease of use of deepfake applications that do not require advanced computer skills. Deepfake applications are developed using deep learning because of their ability to represent high-dimensional complex data, like autoencoders that are used for compression or dimensionality reduction [13].

FakeApp, which was created by Reddit users using an autoencoder pairing structure, was considered the first attempt to create Deepfake. Two pairs of encoders and decoders are used to replace the face between an original and a target image, where each pair is used to train a set of images, as shown in figure 5 [14] [15]. In other words, the process of Deepfake's creation used features of the source face along with the decoding of the target face to reconstruct the target face from the source face. This method was used in many applications like DFaker, DeepFaceLab, and tensor flow-based Deepfake (Deepfake tf) [16].

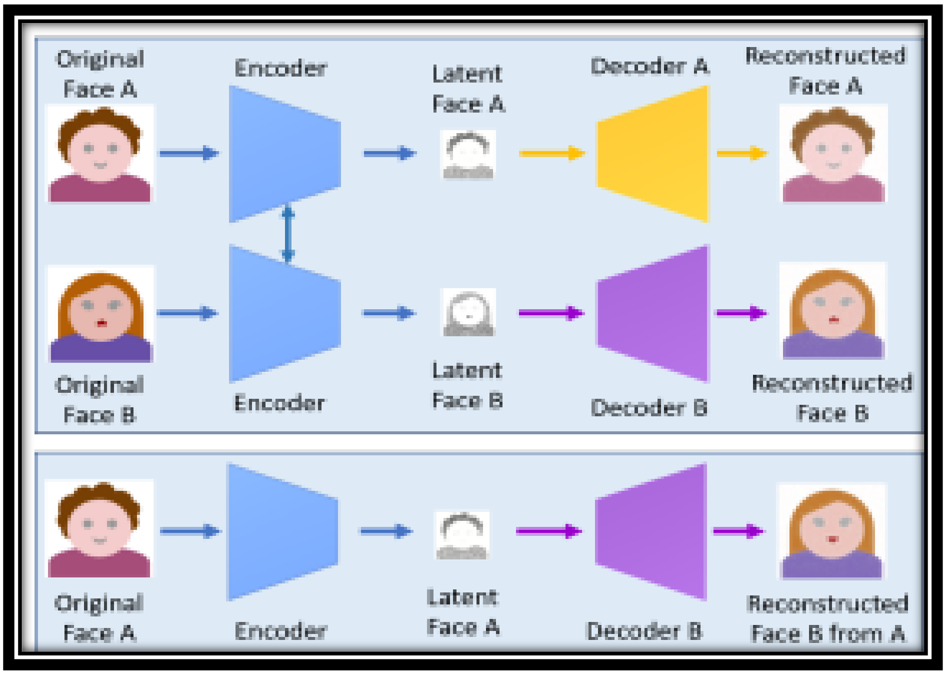

Figure 5: A Deepfake Creation Model using Two Pairs of Encoder-Decoder.

The perceptual and adversarial losses have been added to the encoder-decoder architecture to create the VGGFace application. After that, we proposed a face swap GAN application based on the generative adversarial network, which is improved by the Deepfake application [15]. Then, the FaceNet application used a multitasking convolutional neural network (CNN) to produce steady face detection and reliable alignment, and a generative network was used for the CycleGAN application [14].

6. Negative Defensive Methods (Deepfake Detection)

Deepfake has a significant and increasing impact on the security of society, the privacy of individuals, and democracy [17]. After this menace was introduced, many deepfake detection methods have been suggested [14]. At first, the detection methods were based on handcrafted features gained from contradictions in the fake video combination process [18]. Modern methods, on the other hand, use deep learning to detect deep fakes by automatically extracting discriminative and salient features [11]. This paper presents a survey of some deepfake detection methods that have been applied to images and videos, as shown in table 1.

Table 1: Negative Defensive Methods

No.	Reference	Year	Dataset Type and Name	Technique	Methodology	Accuracy
1	[19]	2017	Images (The LFW face database)	• support vector machines (SVM). • Random forest (RF). • Multi-layer perceptron (MLP)	• The bag of words method is used for extracting features • Entering features into a classifier (SVM, RF, and MLP)	92%
2	[20]	2018	Videos (HOHA dataset)	• Convolutional Neural Network (CNN). • Recurrent neural network (RNN)	• CNN is used to extract frame-level features • Extracted features are used for	94%

					training a classifier (RNN)	
3	[21]	2018	Videos (Created by the researcher)	• Photo response nonuniformity analysis (PRNU)	• Tested effectiveness of PRNU analysis to Deepfake detection,	95%
4	[22]	2018	Videos (CEW Dataset)	• Long-term Recurrent Convolutional Neural Networks (LRCN)	• LRCN is used for the detection of eye blinking (a physiological signal).	98%
5	[23]	2019	Videos (Created by the researcher)	• Neural network • Logistic regression	• Neural network and Logistic regression are used for classifying video (real or fake)	87%
6	[24]	2020	Videos (Created by the researcher)	• Convolutional Neural Network (CNN)	• CNN is used for detecting fake videos, by exploiting spoken sounds that are sometimes conflicted with mouth shape.	93%
7	[25]	2020	Images (CELEBA dataset)	• K-Nearest Neighbors (KNN). • Support vector machines (SVM). • Linear discriminant analysis (LDA)	• Local features are extracted by using the expectation-maximization (EM) algorithm • KNN, SVM, and LDA used for detecting fakes.	93%
8	[26]	2020	Images (CelebA dataset)	• Convolutional Neural Network (CNN). • Common fake feature network (CFFN)	• CFFN was used for the extraction of the features • CNN was used for image classification.	94%
9	[27]	2020	Images (FaceForensics++ (FF++) dataset)	• Convolutional Neural Network (CNN)	• CNN has been used for classifying images	98%
10	[28]	2020	Videos (FF++ and DFDC datasets)	• Ensemble of convolutional neural network (CNN)	• Aggregating different trained CNN models used for classifying images.	84%
11	[29]	2020	Images (CelebA dataset)	• Logistic Regression (LR). • Support Vector Machines (SVM). • K-Means Clustering.	• Pipeline classification is used for the detection of the artificial face by using LR and SVM as supervised algorithms for the classification of labeled data • K-means clustering is used to check the performance of the classification of unlabeled data.	90%

12	[30]	2021	Videos (DeepFake Detection Challenge Dataset (DFDC))	• Convolutional Neural Network (CNN). • Vision Transformer	• CNN is used to extract learnable features • Vision Transformer took the acquired features as input and categorized them using the attention mechanism.	92%

7. Positive Defensive Methods

There are many cryptography and steganography methods used to prevent deepfake. Some of the papers used watermarking (in the video, image, or audio), while others used blockchain technology, and others used artificial intelligence techniques [31-33]. This paper presents a survey of some deepfake prevention methods that have been applied to images and videos, as shown in table (2).

Table 2: positive Defensive Methods

No.	Reference	Year	Dataset Type and Name	Technique	Methodology	Accuracy
1	[34]	2019	Images (UADFV dataset)	• Inconsistent head poses • Support Vector Machine (SVM)	• linking the synthesized face area to the original image, by identifying errors when estimating the 3D head poses from the images. • SVM is used as a classifier	85%
2	[35]	2020	Videos (Created by the researcher)	• Haar Filter • LBPH algorithm • watermarking • AES algorithm	• Haar Filter was used for facial features detection • LBPH algorithm for recognition • Semi-fragile watermarking and AES encryption algorithm was used to prevent fake.	92%
3	[36]	2020	Images (CelebA dataset)	• Conditional image translation • Disruption of spread spectrum	• Suggested a method to disrupt the conditional image translation • Suggested a method to GANs for adversarial training • Suggested a method to disrupt the spread spectrum for evading a vast range of blurring.	95%
4	[37]	2020	Images (CelebA dataset)	• Deep Tag to prevent fake	• Prevent Deepfake by embedding the watermarking message into the face image.	90%

5	[38]	2020	Videos (the database name was not mentioned)	• Digimarc robust watermark • The InterPlanetary File System (IPFS) • Ethereum blockchain	• Digimarc robust watermark technique for preventing audio and image forgery. • IPFS technology is used for storing video • Ethereum blockchain technology is used for storing video metadata	90%
6	[39]	2021	Images (CelebA dataset)	• smart watermark to prevent fake	• Extracts the semantic facial information by using a convolution network. • Specific positions were used as hosts (embedded in the watermark).	92%
7	[40]	2021	Images and Videos (the database name was not mentioned)	• Blockchain (Smart contracts) to prevent fake	• Proposed Smart contracts which keep the integrity of the original data content, thus denying Deepfake.	90%
8	[41]	2021	Images (CelebA-HQ dataset)	• Fake Tagger via Provenance Tracking to prevent fake	• Simple encryption system and deep learning were used for image tagged to trace their source	95%
9	[42]	2021	Images (CelebA dataset)	• A Cross-Model Universal Adversarial Watermark (CMUA-Watermark) to prevent fake	• The CMUA-Watermark module used to protect face image from the Multi - Deepfake	92%

As mentioned above, digital watermarks are used for countering deepfakes; the following is a brief description of these techniques.

A. Video Watermarking Techniques

The video is protected from unauthorized entities by using a watermark. In some research, a watermark is added to an image; in others, it is added to a sound; and in other research, it is added to both a sound and an image, and they are linked tightly to increase the level of safety [43]. Thus, the video is protected from unauthorized access, as it can only be changed by authorized entities [33], [38], and [44].

B. Image Watermark Techniques

After dividing a video (whether compressed or uncompressed) into a set of frames, these video frames are protected by inserting a watermark (synchronization or payload signal) into those frames, where the synchronization signal is embedded in the frequency domain and the payload signal is embedded in the spatial domain [38] [32].

c. Audio Watermark Techniques

To protect the audio from being separated from the video or tampered with, a watermark is added to it. Authorized entities can protect the video from tampering and detect any change in the audio track by dividing the audio into a set of samples (typically between 44 and 48 kHz) and then adding an imperceptible watermark [38].

8. Analysis and Recommendations

There are great concerns due to the rapid development of the GAN, which works to produce deepfake with high-resolution images. Figure 6 shows the evolution of deepfake generation techniques. This is considered a challenge because it makes deepfake content detection very difficult and requires new and intelligent algorithms that can extract features quickly and with high accuracy without the need for training on large datasets for tracking and identifying patterns for fake content production.

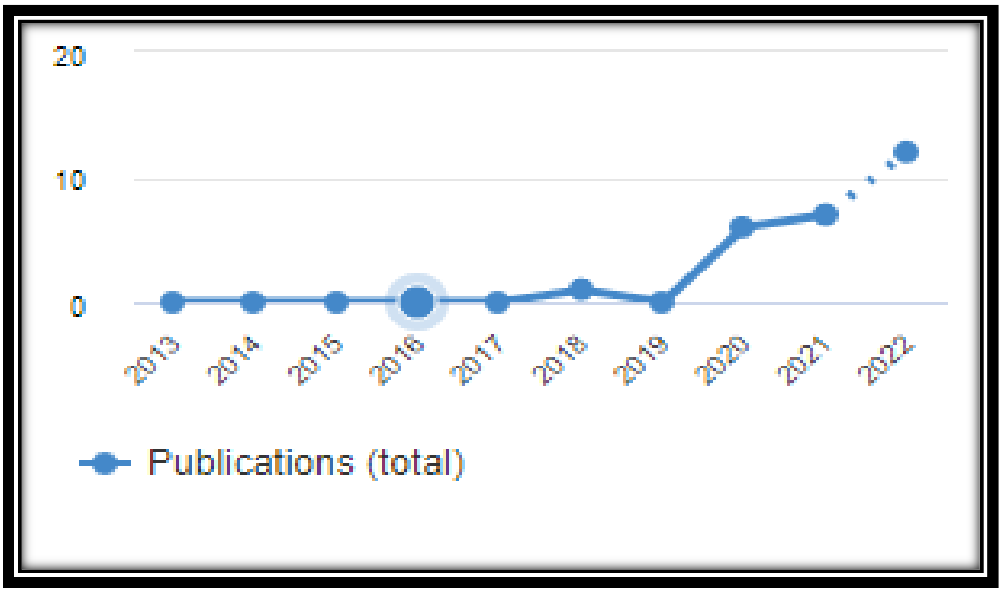

Figure 6: The Evolution of Deepfake Generation Techniques

As for the researchers that work to prevent the deepfake of digital content, they are few, but they have proven their effectiveness in preventing deepfake, with some weaknesses that can be remedied in the future.

Many challenges have been considered while studying deep-fake research, some of which are as follows:

Complexity and Time Consumption: Some of the methods that work to prevent Deepfake are complex and require a large amount of computational time, especially when they are applied to the video. The researcher demonstrated in [35] that the method is accurate in preventing deepfakes and is resistant to all types of noise, but it is a complex method with a long computational time because it works on video datasets, so it is suggested that to avoid these flaws, more accurate face detection methods be chosen and applied to selected frames to reduce the time.

Robustness Factor: After looking at the research, it is found that some of them are resistant to Deepfake, and some of the research distorts their image in the event of being counterfeited. In [36], it is shown that it is accurate and effective in preventing deepfake, and thus, a wide range of blur defenses can be avoided through the use of spread spectrum disruption. The results of the research [39] showed that when manipulating images of a document protected with a smart watermark, they become blurry, and even a person can distinguish them, not just a machine. The results of this research were compared with the results of [36] and proved that it is more accurate and effective than it. The results of a search [40] showed that it is good and preserves the digital content from Deepfake by tracing its content to a trusted source. According to the findings of the study [42], it reduced the quality of the manipulated image so that it could be easily detected by Deepfake.

The Deepfake Application Environment: After reviewing research in the field of Deepfake, it was discovered that the environment in which the research is applied varies depending on the techniques used and the type of data. Researcher in [38] combined the watermarking method with blockchains and proved high accuracy. To improve these methods in the future, they propose combining other encryption methods with blockchain technology. The results of a search [37] showed that it is 90% accurate in recovering the embedded message in the image. While the results of a search [41] showed that it is 95% accurate in recovering the embedded message in the image.

9. Conclusion

Many deepfake types of research have been reviewed, and it has been found that in many of them it depends on the use of a method to determine whether the digital content is real or fake, which helps the user determine the credibility of the publisher. These methods are the most commonly used by researchers. Other research, which is less well covered by researchers, shows techniques to prevent the forgery of digital content by relying on techniques of artificial intelligence and cryptography methods. This research discusses a wide range of research in the field of deepfake by presenting the methods used and discussing the challenges, so it will have scientific importance in the research field of deepfake. The proposed Deepfake prevention methods are usually complex algorithms that require high computational power, but they are effective in preserving the privacy of digital content. To keep our society safe when the use of advanced technology is involved, we suggest that cybersecurity companies go to support researchers in the field of combating deepfake. Future studies will require ways that are a combination of methods of artificial intelligence (supervised and unsupervised learning) and tools for cryptography and have the ability to protect digital content with a high level of accuracy and low computational cost.

References

[1] M. Westerlund, " The Emergence of Deepfake Technology: A Review," Technology innovation management review, vol. 9, no. 11, 2019 .

[2] H. R. a. S. K. Hasan, "Combating Deepfake Videos Using Blockchain and Smart Contracts.," IEEE Access., vol. 7, p. 41596–41606, 2019 .

[3] M. H. a. A. A. Maras, "Determining the authenticity of video evidence in the age of artificial intelligence and the wake of Deepfake videos.," International Journal of Evidence & Proof, vol. 23, no. 3, p. 255–262, 2019 .

[4] M. a. W. L. Yisroel, "The Creation and Detection of Deepfake: A Survey," ACM Computing Surveys, vol. 54, Issue 1, Article No.: 7, pp. 1–41, January 2022. https://doi.org/10.1145/3425780

[5] T.-N. Le, H. H. Nguyen, J. Yamagishi and I. and Echizen, "Robust Deepfake On Unrestricted Media: Generation And Detection," In: Khosravy, M., Echizen, I., Babaguchi, N. (eds) Frontiers in Fake Media Generation and Detection. Studies in Autonomic, Data-driven and Industrial Computing. Springer, Singapore, 2022. https://doi.org/10.1007/978-981-19-1524-6_4

[6] T. H. Obaida and A. S. a. H. N. F. Jamil, " A Review: Video Encryption Techniques, Advantages, And Disadvantages," Webology (ISSN: 1735-188X) , vol. 19, no. 1, 2022 .

[7] N. Flash, A. Ali and T. a. A.-a. A. Khairi, "Video mosaic watermarking using plasma key.," Indonesian Journal of Electrical Engineering and Computer Science, vol. 22, no. 2, pp. 11-20, 2021 .

[8] A. S. a. H. N. F. a. A. R. A. Jamil, " Face Encryption based on Feature Extraction Supported by Canny Edge Detector," Webology, vol. 19, no. 1, 2022 .

[9] L. A. Passos, D. Judas, K. A. P. d. Costa, L. A. S. Junior, D. Colombo and J. P. and Papa, "A Review of Deep Learning-based Approaches for Deepfake Content Detection," arXiv:2202.06095v1 [cs.CV], 2022 .

[10] J. Fletcher, " Deepfake, Artificial Intelligence, and Some Kind of Dystopia: The New Faces of Online Post-Fact Performance," Theatre Journal, vol. 70, no. 4, 2018 .

[11] I. a. C. R. Amerini, "Exploiting prediction error inconsistencies through LSTM-based classifiers to detect deepfake videos," Proceedings of the 2020 ACM Workshop on Information Hiding and Multimedia Security, pp. 97-102, 2020 .

[12] M. Taeb and H. Chi, "Comparison of Deepfake Detection Techniques through Deep Learning," Journal of Cybersecurity and Privacy, pp. 89-106, 2022 .

[13] A. a. B. M. S. Punnap purath, "Learning raw image reconstruction-aware deep image compressors.," IEEE 12 Transactions on Pattern Analysis and Machine Intelligence. , 2019 .

[14] T. T. Nguyen, V. H. Quoc, M. Cuong, N. Dung, T. Duc and N. and Saeid, "Deep Learning for Deepfake Creation and Detection: A Survey," Fellow, IEEE, arXiv:1909.11573v3 [cs.CV], 2021 .

[15] T. T. Nguyen, V. H. Quoc, M. Cuong, N. Dung, T. Duc and N. and Saeid, "Deep Learning for Deepfake Creation and Detection: A Survey," arXiv:1909.11573v4 [cs.CV], 2022 .

[16] J. W. R. J. B. S. a. O. A. V. D. Chorowski, " Unsupervised speech representation learning using wavelet autoencoders.," IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 27, no. 12, pp. 2041-2053, 2019 .

[17] R. a. C. D. K. Chesney, "Deepfake: a looming challenge for privacy, democracy, and national security.," https://dx.doi.org/10.2139/ssrn.3213954., 2018 .

[18] O. F. S. B. S. K. B. a. G. A. De Lima, "Deepfake detection using spatiotemporal convolutional networks.," arXiv preprint arXiv:2006.14749. , 2020 .

[19] Y. Z. L. a. T. V. L. Zhang, "Automated face swapping and its detection.," IEEE 2nd International Conference on Signal and Image Processing (ICSIP), pp. 15-19, 2017 .

[20] D. a. D. E. J. Guera, "Deepfake video detection using recurrent neural networks," 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) , pp. 1-6, 2018 .

[21] M. R. A. M. a. G. Z. Koopman, "Detection of deepfake video manipulation.," In The 20th Irish Machine Vision and Image Processing Conference (IMVIP) , pp. 133-136, 2018 .

[22] Y. C. M. C. a. L. S. Li, "In ictu oculi: Exposing AI-created fake videos by detecting eye blinking," IEEE International Workshop on Information Forensics and Security (WIFS) , pp. 1-7, 2018 .

[23] F. R. C. a. S. M. Matern, " Exploiting visual artifacts to expose Deepfake and face manipulations.," IEEE Winter Applications of Computer Vision Workshops (WACVW) , pp. 83-92, 2019 .

[24] S. F. H. F. O. a. A. M. Agarwal, "Detecting deep-fake videos from phoneme-viseme mismatches.," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 660-661, 2020 .

[25] L. G. O. a. B. S. Guarnera, "Deepfake detection by analyzing convolutional trace," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops , pp. 666-667, 2020 .

[26] C. C. Z. Y. X. a. L. C. Y. Hsu, "Deepfake image detection based on pairwise learning.," Applied Sciences, vol. 10, no. 1, p. 370, 2020 .

[27] L. B. J. Z. T. Y. H. C. D. W. F. &. G. B. Li, " Face X-ray for more general face forgery detection.," In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition , pp. 5001-5010, 2020 .

[28] N. Bonettini, E. D. Cannas, S. Mandelli, L. Bondi, P. Bestagini and S. and Tubaro, " Video Face Manipulation Detection Through Ensemble of CNNs," Politecnico di Milano Milano, Italy, arXiv:2004.07676v1 [cs.CV] , 2020 .

[29] R. Durall, M. Keuper and F.-J. a. K. J. Pfreundt, "Unmasking Deepfake with simple Features," arXiv:1911.00686v3 [cs.LG] , Germany. , 2020 .

[30] D. a. A. S. Wodajo, "Deepfake Video Detection Using Convolutional Vision Transformer," 2021, https://arxiv.org/abs/2102.11126 .

[31] N. a. A. R. N. Flaih, "Proposed Video Watermarking Algorithm based on Edge or Corner Regions," Engineering and Technology Journal, http://dx.doi.org/10.30684/etj.36.1B.4, vol. 36, no. 1, 2018 .

[32] T. N. a. F. N. Hummadia, "Survey of Recent Video Watermarking Techniques," Engineering and Technology Journal , vol. 39, no. 1, pp. 165-174, 2021 .

[33] T. H. Obaida and A. S. a. H. N. F. Jamil, "Real-time face detection in digital video-based on Viola-Jones supported by convolutional neural networks," International Journal of Electrical and Computer Engineering (IJECE) , 2022 .

[34] X. L. Y. a. L. S. Yang, "Exposing Deepfake using inconsistent head poses.," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 8261-8265, 2019 .

[35] L. Sethi, A. Dave, R. Bhagwani and A. and Biwalkar, "Video security against Deepfake and other forgeries," Journal of Discrete Mathematical Sciences and Cryptography, , vol. 23, no. 2, p. 349–363. , 2020 .

[36] N. Ruiz, S. A. Bargal and S. and Stan, "Disrupting Deepfake: Adversarial Attacks Against Conditional Image Translation Networks and Facial Manipulation Systems," Boston University, Boston, MA, USA. , 2020 .

[37] R. Wang, J.-X. Felix, H. Yihao, M. Lei, L. Yang and L. and Wang, "Deep Tag: Robust Image Tagging for DeepFake Provenance," Virtual Event, Association for Computing Machinery, China, 2020 .

[38] A. Alattar, S. Ravi and S. and John, "A System for Mitigating the Problem of Deepfake News Videos Using Watermarking, .," IS&T International Symposium on Electronic Imaging, Media Watermarking, Security, and Forensics, 2020 .

[39] L. Luochen, "Smart Watermark to Defend against Deepfake Image Manipulation," IEEE 6th International Conference on Computer and Communication Systems (ICCCS), 2021 .

[40] M. M. Rashid, S.-H. Lee and K.-R. and Kwon, "Blockchain Technology for Combating Deepfake and Protect Video/Image Integrity," Journal of Korea Multimedia Society, vol. 24, no. 8, pp. 1044-1058, 2021 .

[41] R. Wang, J.-X. Felix, L. Meng, L. Yang and L. and Wang, "FakeTagger: Robust Safeguards against DeepFake Dissemination via Provenance Tracking," Virtual Event, Association for Computing Machinery, China., 2021 .

[42] H. Huang, Y. Wang, Z. Chen, Y. Zhang, Y. Li, Z. Tang, W. Chu, J. Chen, W. Lin and K.-K. and Ma, "CMUA-Watermark: A Cross-Model Universal Adversarial Watermark for Combating Deepfake," arXiv:2105.10872v2 [cscv], 2021 .

[43] N. a. H. T. N. Flaih, "Robust Video Watermarking Algorithm using Features Detection and Discrete Cosine Transform," Design Engineering, vol. 7, pp. 15279-15292, 2021 .

[44] N. a. J. R. K. Flaih, "Proposed Algorithm for Digital Image Watermarking Survival against JPEG Compression," Eng. & Tech. Journal, vol. 32, no. 1, 2014 .