International Journal of Medical Toxicology & Legal Medicine
e-ISSN: 0974-4614
p-ISSN: 0972-0448
Volume 27, No. 1, 2024
DOI: https://doi.org/10.52710/ijmtlm.15

Article Submitted: 01-03-2024; Revised: 25-04-2024; Accepted: 10-05-2024

Recent Advancement in Deepfakes Generation and Detection Tools: A Review

Drishty Chaudhary[1*], Abhinav Singh[2], Prashant Agrawal[3]

[1]Department of Forensic Science, School of Allied Health Sciences, Sharda University, Greater Noida, 201310, India

*Corresponding author:
Dr. Prashant Agrawal,
[3]Department of Forensic Science,
School of Allied Health Sciences,
Sharda University, Greater Noida, 201310, India
Tel: +918375887095
E-mail: Prashant.agrawal@sharda.ac.in

Abstract:

Deepfakes have emerged as a major concern in today's digital age. These artificially manipulated videos can be produced with relative ease, and can be used for a variety of purposes, including entertainment, political propaganda, and even fraud. With the rise of deepfakes, there is a growing need for effective detection methods. In this article, we explored the basics of deepfakes, their generation, detection and their impact in the society. This articles also provides an overview of the existing literature on deepfakes and their detection methods. Our analysis suggests that while deepfakes are a significant threat to society, there is need for effective detection and mitigation using advanced forensic techniques. Additionally, the article also spreads awareness about the deepfakes to overcome the chaos that happens due to spreading fake news, revenge porn, manipulating videos and other criminal acts causing harm to individual or organizations reputation.

Keywords: Deepfakes, Face swap, Deepfakes detection, Deep learning, Artificial Intelligence.

1. Introduction

With the establishment of camera, digital media, and social media platforms, Images and videos are becoming an integral part of our lives. Images and videos record the present scenario and capture the moment that could be used further as per the requirement. It could be anything whether for capturing a memory, recording the crime scene, recording documents, recording scenery, etc..,

Nowadays, with the advancement in digital applications images are becoming more interesting, as these digital applications provide a feature that changes the entire face into a kid's face, aged, or any other type of entertainment. But, these applications are not limited to entertainment only, as this digital application and their advance version allows a perpetrator to tarnish the individual's reputation or committing crime against the individual. In general, the perpetrator tries to tampered the original image by swapping the face of the individual with the target, by morphing their voice, and by altering their facial expressions. Applications that are used to create fakes or manipulated images or videos includes Deepfakes, Deep-face Lab, Adobe Photoshop, Adobe After Effects, etc. As these software programs use sophisticated algorithms to blend the features of the two individuals into a seamless images and videos. Hence, it becomes a daunting task for an investigator or any digital media person to distinguish whether the provided image or video is real or fake.

2. Deepfakes

Deepfake is one of the trending topics across the world. As deepfakes includes the use of machine learning and artificial intelligence to manipulate the image and videos. The manipulation in the videos or images is done with accuracy that it becomes nearly impossible to determine whether the video or image is real or fake.

Figure1: Showing the original image on the left-hand side and its deepfake on the right-hand side (Image courtesy: https://www.technologyreview.com/2019/12/20/131462/this-startup-claims-its-deepfakes-will-protect-your-privacy/)

2.1. How Deepfakes works

To generate these fake deep learning algorithms is used, hence called as *Deep Fakes*.

Deepfakes are example of artificial images that have been manipulated to look like real ones usually by swapping the face of one person with another. This technology relies on deep learning algorithms like deep neural networks to create realistic fakes[Error! Reference source not found.].

Figure2: Showing examples of celebrities and their versions of deepfakes . Image courtesy: Doshi, Savla, Dholakia, Gandhi, Suratkar, & Kazi, 2022

- ***Deep Learning Algorithms***

In general, machine learning algorithms involves training of the neural network that has many hidden layers and hence called as deep learning. These networks are trained and tested by learning the features or patterns in the provided data by analyzing large sets data. The training of these neural network is done based on the three learning algorithms i.e., Supervised, Unsupervised, and Reinforcement learning.

In simple terms, deep learning algorithms process the data like neuron process the data in human brain. The large amounts of data help neural network to recognize complex pattern, classify objects, and make predictions based on their learning algorithms.

Deep learning has been applied to a wide range of tasks, including image recognition, speech recognition, natural language processing, and game playing. Nowadays, its popularity is increasing day-by-day due to its high-level performance, also surpassing human intelligence in some cases[Error! Reference source not found., Error! Reference source not found.].

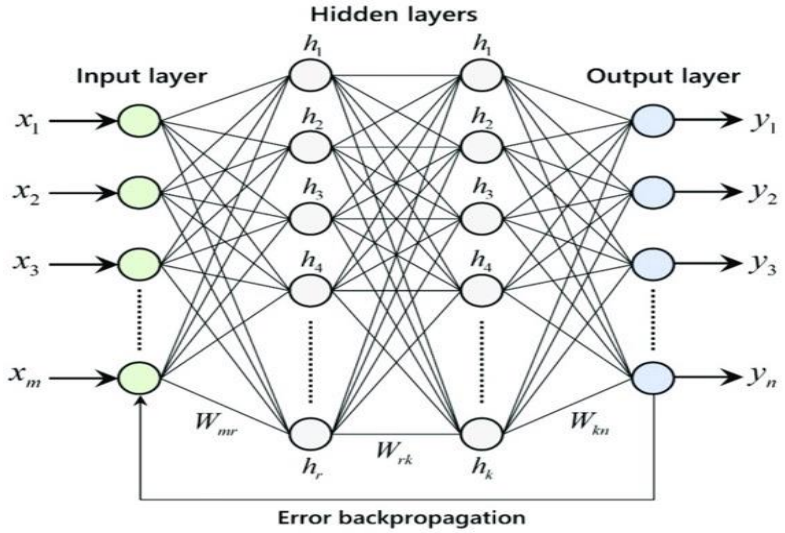

Figure 3: Representing a build neural network with x_1, x_2, $x_3 \ldots x_m$ as an input and h_1, h_2, $h_3 \ldots h_r$ are the hidden layers, w_{mr}, w_{rk}, w_{kn} are the weight adjust and y_1, y_2, $y_3 \ldots y_n$ are the output. Image courtesy: Bre, Gimenez, & Fachinotti, 2018

2.2. *Uses of the deepfakes*

Deepfakes are used for different purposes as per the intention of the user, from entertainment to political propaganda. While deepfakes can be used for harmless purposes, such as adding filters to mobile videos or for creating funny videos, they can also be used for malicious purposes. Deepfakes can be used to spread fake news, manipulate public opinion, or even commit fraud. For example, deepfakes could be used to create videos that falsely show a person confessing to a crime or making a statement they never actually made**Error! Reference source not found.**.

2.3. Deepfakes detection in Forensic Investigations

In recent years, deepfakes have become a growing concern for governments and law enforcement agencies around the world. The detection of deepfakes is an area of active research. Forensic techniques have the potential to be of great use in the detection of deepfakes. These techniques involve analyzing videos or images to identify inconsistencies in the data that may indicate manipulation.

As criminal activities with deepfakes are rising tremendously hence, it requires forensic expert with great knowledge of artificial intelligence to detect deepfake related crimes.

Deepfakes are detected by analyzing the metadata, and the moments of the facial expressions, body posture and other minute details. One of the important aspects for the deepfake detection is the movement of subject's eyes, as deepfakes often fail to accurately capture the natural eye movements. Further, voice identification and classification also play a pivotal role in deepfake detection.

However, these techniques are still in their development phase, as deepfakes are becoming more sophisticated. There is need to build more effective techniques for detecting them as forensic field is not that much explored with artificial intelligence and machine learning algorithms.

2.4. Recent advancement in Deepfakes development and their detection tools

Suwajanakorn, et al., 2017 presented a tool, Synthesizing Obama, that can create lip-synced videos of a person speaking. The tool uses deep learning techniques to analyze an audio recording of a speech and synthesized a video of the person's mouth movements that match the speech. They created realistic and high quality deepfakes of Obama's speech. Further, their study showed that the tool could accurately capture the nuances of lip movements, providing a potential solution for creating accurate lip-synced videos for animation or dubbing**Error! Reference source not found.**.

Moreover, Zakharov, Shysheya, Burkov, Lempitsky, & Sementsov,2019 introduced a tool called Neural Talking Head that employs deep learning techniques and adversarial training to synthesize realistic talking head videos from a few images of the target. The tool generates facial expressions and movements that matches with the audio, producing highly realistic and convincing deepfake videos. It also showed that Neural Talking Head can generate realistic talking head videos with minimum user input, demonstrating its potential for various applications in creating video content. However, paper

provides an in-depth overview of the Neural Talking Head software tool and its contributions to the field of computer graphics, particularly in generating highly realistic deepfake videos[Error! Reference source not found.].

Kim, et al., 2018 proposed a software tool called Deep Video Portraits that uses deep learning to create realistic human portraits from a single still image. The tool works by using a reference video to synthesize facial expressions and movements, and applied to the target image. The authors showed the effectiveness of their approach by generating high-quality and convincing deepfakes. The study showed that the tool could create realistic facial expressions, head poses, and eye movements, providing a potential solution for creating animated avatars or video game characters. Overall, this study provides valuable contributions to the field of computer graphics, particularly in generating high-quality deepfake portraits[Error! Reference source not found.].

However, Kim, Liu, Cao, Cha, & Kautz, 2020 extended the version of Deep Video Portraits tool, and called it Deep Video Reenactment, that enables real-time video reenactment of people using a single camera. The tool uses a combination of deep learning and computer vision techniques to synthesize facial expressions and movements that match the input video stream. The developed tool helps to track the facial landmarks of the input video stream. Also, used to generate realistic facial expressions and head poses, providing a potential solution for creating real-time video reenactment or live avatar animation[Error! Reference source not found.].

Li, Li, & Wu, 2020 presented a comprehensive review of deep learning techniques used for detecting deepfakes. The authors discussed various types of deepfake techniques and explained how deep learning methods have been used to detect them. They categorized deepfake detection methods into three main categories: *classification-based methods*, *face reconstruction-based methods*, and image *quality assessment-based methods*. The authors then provided an in-depth discussion of each of these methods, highlighting their advantages, limitations, and performance. In addition, discussed datasets commonly used for evaluating the performance of deepfake detection methods[Error! Reference source not found.].

Nguyen, Nguyen, & Nguyen, 2020 conducted a comprehensive review of the current deep learning-based techniques used for deepfake detection. The authors provide an overview of generative models, convolutional neural networks, and auto-encoders used in deep learning for detecting deepfakes. The paper covers the strengths and limitations of each method and highlights the challenges faced in detecting deepfakes. The authors also present a comparative analysis of these deep learning techniques and discuss the potential of combining multiple approaches for improving deepfake detection accuracy. This paper provides an informative and impressive review of the latest techniques for deepfake detection based on deep learning[Error! Reference source not found.].

Dang, Guo, Zhang, & Yu, 2020 conducted comprehensive survey of deepfake detection techniques that have been developed in recent years. Also, highlighted traditional machine learning, deep learning, and hybrid approaches that are used for deepfake detection. These approaches include image and video-based, audio-based, and multi-modal transformation. Further, the strengths and limitations of these approaches were highlighted. The important benchmark datasets for evaluating the performance of deepfake detection algorithms and discuss the future directions for research in this area. Overall, this paper provides a useful resource for researchers and practitioners interested in deepfake detection[Error! Reference source not found.].

Zhang, Chen, Wen, & Li, 2020 provide a comprehensive review of deepfake detection techniques, focusing on both traditional and deep learning-based methods. The authors describe the most commonly used datasets and benchmarks for evaluating deepfake detection techniques, and provide an overview of the performance of different approaches. They discuss the limitations of current techniques, such as the difficulty of detecting more advanced deepfakes, and highlight the need for more diverse and challenging datasets to improve the robustness of deepfake detection models. Overall, this review paper provides valuable insights into the current state-of-the-art in deepfake detection and future directions in this rapidly evolving field[Error! Reference source not found.].

Wu, Wu, X, & Xu, 2020 present a review paper discussing the current trends and challenges in deepfake detection. The authors discuss the importance of large-scale datasets and the role they play in developing effective deepfake detection methods. In addition, highlighting the potential for adversarial attacks used in deepfake generation. Further, the need for more robust detection methods to combat and concluding the potential for future research in the areas of developing more sophisticated deep learning models and improving the interpretability of these models[Error! Reference source not found.].

Additionally, Wang, Lu, & Zhang, 2021 provide a comprehensive overview of deepfake detection techniques, from traditional methods to deep learning-based techniques. Moreover, discussing the importance of large-scale datasets and the

need for robustness against adversarial attacks and provide insights to the potential use of emerging technologies such as blockchain and explainable AI for enhancing the effectiveness of deepfake detection methods[Error! Reference source not found.].

Ning, & Zhan, 2021 calls attention to comprehensive overview of recent progress in deepfake detection. Majorly, focusing on the use of deep learning-based techniques for deepfake detection, by generative models, convolutional neural networks, and autoencoders. Also, discussing the importance of multi-modal analysis, that involves analyzing of visual content of a video and accompanying audio and text data. Further, highlighting future research scope on developing more sophisticated deep learning models that can effectively distinguish between real and manipulated videos, as well as on developing more comprehensive datasets that capture a wider range of deepfake scenarios[Error! Reference source not found.].

Sun, Li, & Liu, 2021 highlighted the current state of deepfake detection, covering both traditional and deep learning methods. Further, discussing about the challenges faced in the field, such as the increasing sophistication of deepfake technology that requires a large and diverse dataset for training, testing and their future advancement.[Error! Reference source not found.]

3. Conclusion & Discussion

The generation of high-level deepfakes has been the subject of extensive research, resulting in the development of various tools to generate such content. However, the potential harm and dangers associated with deepfakes cannot be overstated. These digital forgeries have the potential to manipulate public opinion, discredit individuals, and even cause financial losses. Enormously, used in spreading false information, damaging the reputation, cyberbullying, revenge porn, and identity theft against the individual or organization. Therefore, detecting deepfakes is crucial to reduce their impact. Fortunately, several tools such as Deep Face Lab[Error! Reference source not found.], Face Swap[Error! Reference source not found.] and Xception Net[Error! Reference source not found.] have been developed for this purpose. These tools use machine learning algorithms to analyze images and videos to identify inconsistencies that indicate tampering. While these tools are effective, it is important to note that they are not full proof and can be bypassed by new deepfake techniques. However, it is important to note that the technology behind deepfakes is continuously evolving, and the detection methods must keep up with these advancements. Therefore, people should be aware that the images and videos they see online may not be genuine. It is essential to verify the authenticity of the source and the content before sharing or acting on it. Moreover, policymakers should take measures to address the threat of deepfakes, including regulation and education. By being aware of the risks and taking proactive steps, we can mitigate the impact of deepfakes and protect ourselves from their harmful effects.

4. Acknowledgement:

We would like to acknowledge Dr. Sally Lukose for her constant support and guidance.

5. Conflict Of Interest:

There is no conflict of interest.

References:

[1] Coccomini DA, Caldelli R, Falchi F, Gennaro C. On the Generalization of Deep Learning Models in Video Deepfake Detection. J Imaging. 2023 Apr 29;9(5):89.

[2] Zhu Y, Wang M, Yin X, Zhang J, Meijering E, Hu J. Deep Learning in Diverse Intelligent Sensor Based Systems. Sensors (Basel). 2022 Dec 21;23(1):62.

[3] Bre, F., Gimenez, J., Fachinotti, V., Prediction of wind pressure coefficients on building surfaces using artificial neural networks. Energy and Buildings,2018;158:1429-41.

[4] Shahzad HF, Rustam F, Flores ES, Luís Vidal Mazón J, de la Torre Diez I, Ashraf I. A Review of Image Processing Techniques for Deepfakes. Sensors (Basel). 2022 Jun 16;22(12):4556.

[5] Suwajanakorn, S., Seitz, S. M., Kemelmacher-Shlizerman, I., Synthesizing Obama: Learning Lip Sync from Audio.. ACM Transactions on Graphics,2017;36:1-13.

[6] Zakharov, E., Shysheya, A., Burkov, E. and Lempitsky, V.,. Few-shot adversarial learning of realistic neural talking head models. In Proceedings of the IEEE/CVF international conference on computer vision 2019; 9459-68.

[7] Kim, H., Garrido, P., Tewari, A., Xu, W., Thies, J., Niessner, M., Pérez, P., Richardt, C., Zollhöfer, M., Theobalt, C., Deep video portraits. ACM Transactions on Graphics (TOG), 2018;37(4):1-14.

[8] Han, Y., Wang, Z., Xu, F., Learning a 3D Morphable Face Reflectance Model from Low-cost Data. arXiv preprint arXiv: 2023; 2303:11686.

[9] Li, J., Li, Y., Wu, X., Deep learning for deepfake detection: A review. IEEE Access, 2020; 8:71683-71695.

[10] Nguyen, T. M., Nguyen, T. P., Nguyen, K. T., A survey of deep learning-based deepfake detection techniques. Journal of Ambient Intelligence and Humanized Computing,2020; 11:4493-4505.

[11] Dang, H., Guo, J., Zhang, Y., Yu, T., Deepfake detection: A survey. Information Fusion,2020;64:145-162.
[12] Zhang, Y., Chen, Y., Wen, F., Li, B., Deepfake detection: A review.. Neurocomputing, 2020;412:305-326.
[13] Wu, Q., Wu, X., Xu, X., A survey on deepfake detection: Trends and challenges. Information Sciences,2020; 546:137-157.
[14] Wang, J., Lu, Y., Zhang, C., A survey on deepfake detection: From traditional methods to deep learning-based techniques. Multimedia Tools and Applications,2021;80:4617-47.
[15] Ning, Y., Zhan, Y., Deepfake detection: A review of recent progress. IEEE Transactions on Multimedia,2021;23:2385-2401.
[16] Sun, L., Li, C., Liu, X., A review of deepfake detection and related research. Future Generation Computer Systems,2021;117:255-274.
[17] Perov, I., Gao, D., Chervoniy, N., Liu, K., Marangonda, S., Umé, C., Dpfks, M., Facenheim, C.S., RP, L., Jiang, J., Zhang, S., DeepFaceLab: Integrated, flexible and extensible face-swapping framework. arXiv preprint arXiv: 2020. ; 2005.05535.
[18] Korshunova, I., Shi, W., Dambre, J., Theis, L., Fast face-swap using convolutional neural networks. In Proceedings of the IEEE international conference on computer vision,2017; 3677-85.
[19] Kusniadi, I., Setyanto, A., . Fake video detection using modified XceptionNet. In 2021 4th International Conference on Information and Communications Technology (ICOIACT) ,2021; 104-107.