2023 IEEE IAS Global Conference on Emerging Technologies (GlobConET)
Loughborough University, London, United Kingdom. May 19-21, 2023

2023 IEEE IAS Global Conference on Emerging Technologies (GlobConET) | 979-8-3503-3179-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/GlobConET56651.2023.10150143

Deep Learning Technique for Recognition of Deep Fake Videos

Fahad Mira
Collage of computer and information Technology
Department of computer engineering
University of Bedfordshire
Almira3000@hotmail.com

***Abstract*—New computing methods and digital content have been created thanks to recent advancements in digital media technology. They have also contributed to advancing recent AI-based innovations and provide straightforward instruments for producing real video changes. These "Deep Fakes" or fraudulent films might seriously jeopardise the public's perceptions of a case or society. These films' consequences on spreading fake news, particularly, are significant when they act as accurate depictions. These false films may, however, be created by manipulating software. Data protection, identifying deep fakes, and preventing media manipulation are just a few ways deep fake detection contributes to cybersecurity. In light of this, it is essential and mandatory to be able to spot this sort of misleading data. This paper examines the most promising new approaches to deep fake video detection by analysing the latest findings from the research community. It analysed the results from two research and proposed using convolutional neural networks and long short-term memory to distinguish fake from real video frames. The report suggested using these and other detection methods and the unique method for identifying deep fakes that used the YOLO face detector to distinguish facial video frames (YOLO-CNN-XGBoost) and suggested investigating other novel detection methods.**

***Index Terms*—Deep Learning, Deep Fake, Deep Fake Video, Video Recognition, Yolo, Fake Detection**

I. INTRODUCTION

Deep learning is a branch of machine learning that deals with artificial neural network techniques that are impacted by the structure and operation of the brain. Approaches to deep learning increase the complexity of the technology used in creating and distributing multimedia content. Deep Fakes are a new technology that has started to surface recently. It is quite simple to create influential films in which actors' faces—or even their lips and eyes—have been altered [1]. Additionally, deep learning (autoencoders and networks) with generative adversaries have been widely used to address various issues [1]. Deep fake algorithms have also used these models, which examine a person's expressions and movements to generate fake images of their faces [2]. Thus, a large amount of picture and video data is usually required when training models to generate real results using deep fake techniques. Deepfakes typically begin with famous persons because of the abundance of publicly available information about them online. Pornographic photos and movies with recognisable faces were altered using deep fakes.

Since the advent of the internet, the pursuit of truth has taken on an even larger significance. Because deep fakes can be created by practically anybody using today's deepfake

technologies, and because they are typically employed for nefarious reasons, combating them is significantly more difficult. Many different techniques have been put out so far to find deep fakes. The majority of them also use deep learning.

In addition, several deepfake movies have been uploaded on social media due to the widespread availability of relevant technologies. The term "deepfake" refers to any digital media in which the subject's likeness has been altered through editing. One of the greatest challenges facing contemporary society is deepfake. Famous Hollywood stars' faces have frequently been added using Deepfake to pornographic images and videos. Additionally, Deepfake has been used to disseminate rumours and misleading information to politicians. [1] [2] [3].

The author has been comparing and evaluating current papers on deep learning algorithms for deep fake video identification to identify the most effective new approaches. It analysed the results of two experiments that compared results from using LSTM and CNN to identify fake and authentic video frames.

A. Problem Statement

The researcher observed from the literature review that deepfake images and videos began to increase. According to the study, a phoney film of Barack Obama was created in 2018 using quotes he never spoke [4]. Even more disturbingly, deep fakes were used to edit recordings of Joseph Biden's lips before the 2020 US election. These malicious uses of deep fakes might harm society by spreading misinformation, especially through social media; therefore, this research determines the exact forms of identity theft. And compare it with other countries to develop recommendations to help resolve this big issue. This agrees with [5], who confirms that deepfake images and videos have increased globally.

B. Research Question

What are the various Deep Learning Techniques for Deep Fake Video Recognition?

C. Objectives of the study

To analyse recent research to investigate how deep learning can help to recognise the real fake videos from the fake ones using analytical review from old to new trends.

D. Significance of the research

- This study emphasises the significance of the Deep Learning Method for detecting Deep Fake Videos.
- It can assist other academics in researching the most re- cent tools for detection in various contexts and environments.

E. Definition of the Terms

1) Deep learning is a branch of an extended group of machine learning techniques concentrating on convolutional neural networks or representational learning in ar- artificial neural networks (CNN). Deep learning, a machine learning system, is very good at dealing with unstructured data. As opposed to traditional machine learning techniques, deep learning is more effective. It allows computer models of varying complexity to learn from the input gradually. A modern variation known as deep learning takes on an unlimited number of bounded-size layers, enabling functional deployment and optimal execution while maintaining theoretical universality under moderate response conditions [6].

2) Deepfake: It is a deep learning and false synthetic media phenomenon in which the image of a person in a real photo or video is changed to that of another person. Deepfakes result from recent advancements in deep learning, a category of artificial intelligence. Neural network algorithms discover rules and replicate patterns by combing enormous data sets. For instance, Google has developed fully qualified domain algorithms using this method. Deepfakes are unique because algorithms are pitted against one another in "generative adversarial networks" within a GAN that produces content based on source code data. [7].

F. Limitations of the study

The current research was limited to all research mentioned in the literature review. Therefore, the research's most obvious flaw is that it depends on already-existing data and literature to answer the research question. As a result, the conclusions and data are limited to those previously presented in the literature and from which data was gathered. This significantly impacted the analysis's ability to employ information from the most current research because it wasn't always available. Therefore, after it was discovered, the more current information gathered was used. The current research was done during the first semester of 2022.

II. LITERATURE REVIEW

The broad adoption of Deep Fakes is attributable to the high quality of the faked movies and the ease with which their programmes may be used by a wide variety of users, from professionals to novices with varied degrees of programming ability. The creation of these apps typically involves the use of deep learning methods. It is well-established that deep learning can successfully represent complex and high-dimensional data. For dimensionality reduction, a specific type of deep network called deep autoencoders has been frequently used and image compression [8]. The first effort at deep-fake creation was FakeApp, developed by an Internet user utilising the auto encoder-decoder pairing structure (Figure. 1).

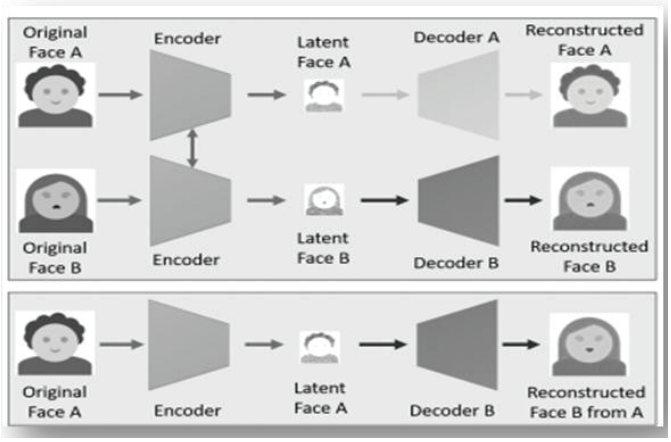

Fig. 1. :

However, [8] created a stunning Deep Fake data set that is made up completely of 620 videos. They used the GAN model and the Deep Fake data set. Deep Fake film was created using low and high-quality Faceswap-GAN Open Source Code Videos from the publicly available VidTIMIT website [9], which can faithfully mimic facial gestures, lip movements, and eye blinking. These films were also used to test several deep false detection techniques. When used to identify Deep Fake films from this freshly created data set, different approaches, such as lip-syncing methods and support vector machine (SVM) picture quality metrics [10], produce exceptionally high mistake rates.

Deep Fake is another technique cybercriminals use to get past authentication or identity checks and get unauthorised access. (CNN) and (GAN) are two examples of deep learning tools that have made preserving facial characteristics and posture more challenging for forensic models in switched-face images. [11] as well as the photographs' lighting. Zhang et al. [12] employed the bag of words method to extract a group of condensed traits, which they then fed into classification algorithms, including SVM, random forest, and multi-layer perceptrons (MLP) to distinguish from the real swapped face photographs. Since GAN models can learn how to disperse detailed input data, their synthesised images are accurate and high-quality, possibly the most challenging deep learning-generated images to categorise.

Recently [13] conducted a study and pointed out Artificial neural networks (ANNs) as it takes some of their fundamental ideas from how the human brain operates. The architecture of (ANNs) is shown in (Figure 2). Neural networks consist of many layers: an input layer, perhaps numerous hidden layers, and an output layer. The input to the neural network is a data set. Namely, neural networks are programmed to foresee and classify these data into specified buckets.

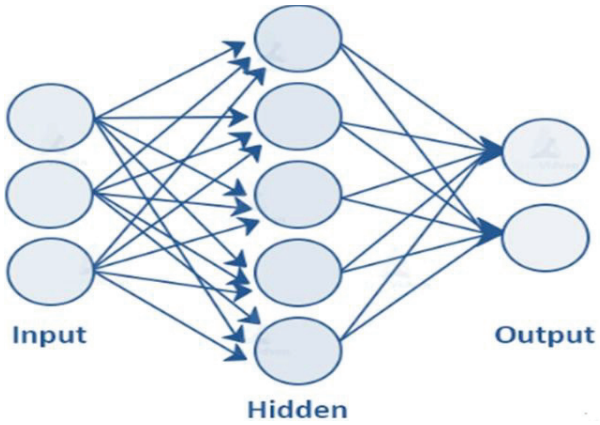

Fig. 2. :

III. METHODOLOGY AND PROCEDURES

However, the researcher in the current study took a descriptive method, reading broadly on the issue and drawing from prior studies to accomplish the study's goals. The researcher used current literature to perform a descriptive and comparative review of the relevant variables that may potentially help to explore the ways of identity fraud to conduct this research. The related literature was acquired via google scholar research papers.

IV. RESULTS AND DISCUSSIONS

The question "What are the various Deep Learning Methods for the Identification of Deep Fake Videos?" was posed for this study. Several deep learning-based strategies were outlined in the literature, including (CNN); (RNN); (and LSTM). The researcher briefly explains these strategies before explaining how they were used in deepfake discovery. It is crucial to note that the foundation of a deep learning machine learning technique is identical to that of a neural network. In deep learning, "deep" refers to utilising numerous hidden layers inside the network. With raw input data, the deep learning architecture (influenced by artificial neural networks) can extract higher-level information by using an unlimited number of hidden layers of finite size. The level of complexity present in the training data determines the number of hidden layers. More hidden layers are required for more complex data to offer good results correctly. During the past few years, deep learning has shown effective in various contexts, such as computer vision, audio processing, machine translation, and natural language processing. Among the many methods that rely on deep learning are:

A. Convolutional Neural Network (CNN)

The most often used model is one based on deep neural networks. Much like neural networks, CNN consists of an input layer, an output layer, and one or more hidden layers. First, the hidden layers in a CNN process the inputs from the first layer by a mathematical convolution operation. Convolution, here, refers to a dot product or matrix multiplication. CNN uses matrix multiplication, nonlinearity activation like the Rectified Linear Unit, and convolutional approaches like pooling layers.

B. Recurrent Neural Network (RNN)

RRN's strength lies in its ability to facilitate the identification of temporally dynamic activity. We offer a recurrent hidden state representing dependency across many time scales to deal with a temporal sequence.

C. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is an artificial recurrent neural network (RNN) for dealing with dependencies across time. Learning the whole data sequence with LSMT's feedback connections is possible. The fundamental architecture of an LSTM consists of input, forget, and output gates. The LSTM cell's state remembers the values from previous periods and stores them there.

D. Generation and Detection of Deepfakes

It is a technology that creates fake images and videos using Generative Adversarial Networks (GANs) techniques. An encoder and a decoder are the two neural network components that make up the architecture of GANs. The model trains on a big data set using the encoder to generate fake data. The bogus data is then learned from actual data using the decoder. But for this model to produce faces that seem realistic, a lot of data, including photographs and videos, are needed. (Figure 3).

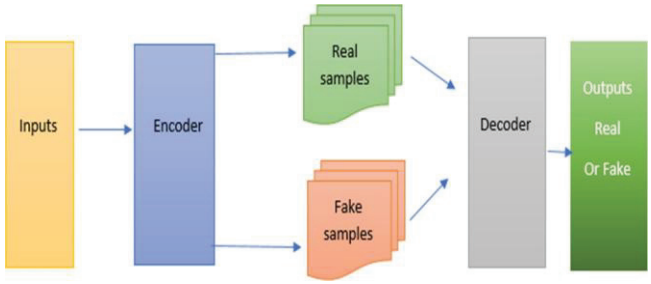

Fig. 3. :

1) Detecting deep fakes:

Machine learning has successfully identified deep fakes. It can recognise photos and videos. However, the researcher will focus on video detection for this investigation.

a) Detecting Deepfake Videos:

1) Analysis of biological singles:

To identify fake face videos, [7] introduced a novel method based on natural networks. In contrast to other research, this technique considers eye blinking, an important physical characteristic that may be utilised to identify fraudulent films. It uses a (CNN) and (RNN) combination to recognise physiological signals like blinking and eye movement. The next step is for the model to detect whether the eyes are open or closed using a binary classifier. An eye-blinking dataset downloaded from the internet is used to evaluate this strategy.

2) Analysis of Spatial and Temporal Features:

Most current methods for identifying deep fakes rely solely on a single still image. Recent research has shown that carefully scrutinising the temporal sequence between frames makes it feasible to tell the difference between a real and a fake video. A temporally-aware approach was proposed in a recent study to identify deepfake films. In the first stage, a

convolutional neural network extracts frame information for the model (CNN). Afterwards, they are sent to the Long Short-Term Memory (LSTM) layer to analyse a time series to detect facial expression changes between consecutive images. Finally, the video is categorised as either real or false using a softmax algorithm. [14] described a novel Recycle-GAN method that combines temporal and geographic data using conditional generative adversarial networks. The evaluation's findings demonstrate that integrating the time and geographical limitations can produce a useful output. A brand-new strategy based on recurrent convolutional networks is also proposed [15]. The technique is divided into face processing and face modification detection. In the processing phase, a spatial transformer network retrieves the cropped and aligned face (STN). The intermediate results are then fed into a recurrent convolutional network specifically designed for face modification detection, where temporal information across frames is analysed. (Figure 4).

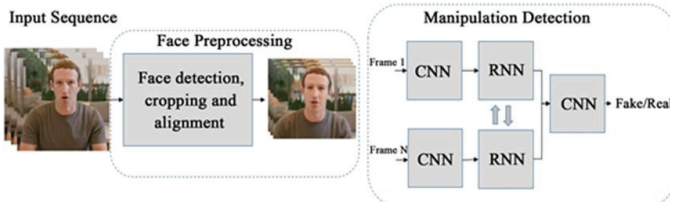

Fig. 4. :

The authors [15], [13] presented introduces a new technique for detecting deep fakes: severe gradient boosting, convolutional neural network, and you look once (YOLO-CNN-XGBoost). After extracting features using the YOLO face detector and video frames, these faces are fed into the InceptionResNetV2 CNN. The CNN network's top-level recogniser, XGBoost, receives these properties. The proposed method may achieve 90% AUROC in receiver operating characteristic plots. The experimental analysis validates the benefits of the proposed strategy over state-of-the-art solutions. However, Deepfake detection uses a variety of datasets, including 100K-Faces; FFHQ; CASIA-WebFace; DFFD; VG-GFace2; The eye-blinking dataset; and DeepfakeTIMIT.

Additionally, there are Google Colab, Jupyter Notebook, the Python programming language, the Keras deep learning library, the Python Imaging Library (Pillow), and Google Drive: for storing datasets. This is especially crucial now since social networking sites make it simple for users to spread and share such fake information, and deep fake-making tools are becoming more widely available. Numerous fields have shown much interest in deep learning techniques. Numerous deep learning-based approaches have recently been implemented to address this problem and effectively identify fake photos and videos.

V. RECOMMENDATIONS

Despite deep learning's impressive success in identifying deep fakes, the quality of deep fakes has been rising. Improving existing deep learning approaches is important to recognise fraudulent movies and pictures successfully. The study's author recommends a new method for spotting deep fakes. The YOLO face detector is used in this method to locate faces in videos. InceptionResNetV2 CNN is used to extract discriminant spatial properties of these faces, which aids in detecting visual artefacts in the video frames. These visual attributes are distributed over the XGBoost classifier to help distinguish between real and deepfake movies. In conclusion, the study suggests applying cutting-edge research techniques.

VI. CONCLUSION

Deepfake's rise to prominence can be attributed to the proliferation of visual content on social media platforms. This is especially important now, as social media platforms facilitate the dissemination and sharing of such false information, and deep fake-creating tools become more readily available. The application of deep learning methods has attracted considerable attention from many disciplines. As was previously noted, many deep learning-based algorithms have recently been released to address this issue and reliably detect phoney photographs and videos.

REFERENCES

[1] E. Vezzetti, F. Marcolin, S. Tornincasa, and P. Maroso, "Application of geometry to rgb images for facial landmark localisation-a preliminary approach," *Trilling, Bernie & Fadel. 21st Century Skills:. Learning for Life in Our Times*, vol. 8, pp. 978–978, 2013.

[2] L. Nataraj, "Detecting GAN generated fake images using co-occurrence matrices," *Electronic Imaging*, vol. 5, pp. 532–533, 2019.

[3] Sheng-Yu Wang, "CNN-generated images are surprisingly easy to spot... for now," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020.

[4] C. Vaccari and A. Chadwick, "Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news," *Social Media+ Society*, vol. 6, pp. 2 056 305 120 903 408–2 056 305 120 903 408, 2020.

[5] M. Masood, "Deepfakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward," *Applied Intelligence*, pp. 1–53, 2022.

[6] D. Güera and E. J. Delp, "Deepfake video detection using recurrent neural networks," *15th IEEE international conference on advanced video and signal based surveillance (AVSS)*, 2018.

[7] X. Yang, Y. Li, and S. Lyu, "Exposing deep fakes using inconsistent head poses," *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2019.

[8] G. Grekousis, "Artificial neural networks and deep learning in urban geography: A systematic review and meta-analysis," *Computers, Environment and Urban Systems*, vol. 74, pp. 244–256, 2019.

[9] S. Pouyanfar, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Computing Surveys (CSUR)*, vol. 51, pp. 1–36, 2018.

[10] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.

[11] Hochreiter, S., and J. Schmidhuber. "Long short-term memory Neural computation. 1997. 9 (8): 1735–1780."

[12] Schuster, Mike, and K. Paliwa Kuldip. "Bidirectional recurrent neural networks (1997)."

[13] Ismail, Aya, et al. "A New Deep Learning-Based Methodology for Video Deepfake Detection Using XGBoost." Sensors 21.16 (2021): 5413.

[14] Bansal, Aayush, et al. "Recycle-gan: Unsupervised video retargeting." Proceedings of the European conference on computer vision (ECCV). 2018.

[15] Sabir, Ekraam, et al. "Recurrent convolutional strategies for face manipulation detection in videos." Interfaces (GUI) 3.1 (2019): 80-87.