International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net

Vol-19 Issue-01 Feb 2024

DEEPFAKE DETECTION A SYSTEMATIC LITERATURE REVIEW

[1] Dasari Shirisha, [2] M Harshitha Goud,[3] Vaddeman Swetha, [4] Ganta Vijay Kumar

[1,2,3] Assistant Professors,Department of Computer Science and Engineering, Kasireddy Narayanreddy College Of Engineering And Research, Abdullapur (V), Abdullapurmet(M), Rangareddy (D), Hyderabad - 501 505

[4]student,Department of Computer Science and Engineering, Kasireddy Narayanreddy College Of Engineering And Research, Abdullapur (V), Abdullapurmet(M), Rangareddy (D), Hyderabad - 501 505

ABSTRACT

Artificial intelligence (AI), machine learning (ML), and deep learning (DL) have all made tremendous strides in the recent few decades, opening up new possibilities for the manipulation of multimedia. Even while most people have utilised the technology for good, such in entertainment and education, some bad apples have found a way to get their hands on it and do evil. As an example, individuals have made convincingly phoney movies, photos, or audios to annoy and blackmail others, propagate false information and propaganda, or incite political strife and hatred. Recently, the term "Deepfake" has been used to describe these doctored, high-quality, and convincing films. In response to Deepfake's concerns, other methods have been detailed in the literature. In this study, we do a systematic literature review (SLR) to provide a current synopsis of Deepfake detection research. The publications range in date from 2018 to 2020 and cover a variety of approaches. Methods based on deep learning, methods based on classical machine learning, methods based on statistics, and approaches based on blockchain are the four main groups into which we classify them for analysis. We also compare the systems' detection capabilities across multiple datasets and find that deep learning-based approaches are superior for Deepfake detection.

I. INTRODUCTION

When it comes to manipulating audiovisual material, the significant advancements in technology based on artificial neural networks (ANNs) are crucial. For instance, photo and video face swapping using AI-powered apps like Fake App [2] and Face App [1] has been done convincingly. Thanks to this switching system, anyone may change their appearance, including their hairdo, gender, age, and more.

attributes. These false films have gone viral under the alias "Deep Fake," and they're causing a lot of worry. "Deep fake" is an amalgamation

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-19 Issue-01 Feb 2024

of "Deep Learning (DL)" and "fake," and it refers to particular photorealistic video or picture material that has been generated with the help of DL. Named after a Reddit user who, in late 2017, used deep learning techniques to make photorealistic false pornographic films by swapping out one person's face with another's. These fake films were created using two types of neural networks: one that generates content automatically and another that uses a Face Swap approach [3, 4]. By using an encoder and a decoder, the generative network is able to generate synthetic visuals. The new pictures' veracity is determined by the discriminative network. Ian Good, a colleague of mine, presented a hybrid of the two networks he dubbed Generative Adversarial Networks (GANs). [5].

Deep fake published an annual report [6] detailing several relevant generative modelling accomplishments performed by DL researchers. As an example, a technique called Face2Face [7] was suggested by computer vision experts for face re-enactment.

A digital "avatar" may have their facial expressions sent to them in real-time using this technique. Cycle GAN [8], developed by UC Berkeley researchers in 2017, may change the look of films and photos. Researches from the University of Washington put out an alternative approach to synchronising lip movements in video with audio from another source [9]. Lastly, the phrase "Deep fake" came up in November 2017 to describe the distribution of porn films that had the original faces of celebrities changed out. In January 2018, a number of websites, supported by private sponsors, introduced a Deepfake creation service. Gfycat[10], Pornhub, and Twitter were among the websites that banned these services after one month. However, research into Deep Fake developed quite quickly in light of the dangers and hazards associated with privacy vulnerabilities. In March 2018, the Face Forensic video dataset was released by Rossler et al. [11] to train media forensic and deep fake detection systems. Months later, a technique called "Deep video portraits" [12] was released by Stanford University researchers. It allows for the photorealistic re-animation of portrait movies. In order to replicate a person's motions on screen, researchers from the University of California, Berkeley came up with an alternative method [13]. For the purpose of creating synthetic images, NVIDIA unveiled a style-based generator architecture for GANs [14]. Multiple web sites containing Deep Fake-related movies were found by the Google search engine, according to the [6] study (see Figure 1).In this study [6], we discovered the following further details:

Although pornhub.com has disabled searches for "Deep fakes," the top ten pornographic

portals nonetheless managed to release 1,790C Deep fake films.

_ 6,174 Deep bogus movies with false material are posted on adult sites.

3 New sites were created specifically for the distribution of deep fake porn.

_ There were 902 publications published in arXiv in 2018 that included the term GAN in some way, either in the title or the abstract.

_ Twelve of the twenty-five papers published on the topic—including those without peer reviews—were supported by DARPA.

Many other harmful or unlawful applications of Deep fake exist outside of deep fake pornography, including the dissemination of disinformation, the instigation of political instability, and the commission of many forms of cybercrime. Academics and industry professionals have focused heavily on deep fake detection in recent years, leading to a proliferation of deep fake detection approaches, as a means of countering these dangers. Efforts to survey specific literature on detection techniques or performance analysis have also been made. Previous surveys conspicuously omitted information regarding accessible datasets while summarising Deep fake in all its parts; so, a more thorough assessment of this study topic would be helpful in supporting the academic and practitioner communities. With that goal in mind, this article presents a Deep fake detection SLR. Our goal is to outline and evaluate the various techniques to Deep fake detection as well as the commonalities among them.

We have summarised our contributions below.

_ We do an exhaustive literature review in the Deep fake field. While offering certain research topics, we provide a rundown of the present state of the art in Deep fake detection methods, strategies, and datasets.

Our groundbreaking taxonomy is the first of its kind; it divides Deep fake detection algorithms into four distinct classes and provides an overview of each class along with its associated properties.

_ The experimental evidence from the main research is examined thoroughly. In addition, we use a variety of indicators to assess how well different Deep fake detection approaches work.

To aid in the development of future studies and methods in this area, we provide some recommendations for deep fake detection and draw attention to a few key points.

The rest of the article is structured like this: Interest research questions are defined in Section II to provide the review method. We go over the results of all the research in Section III. The study's main findings are summarised in Section IV, and the difficulties and restrictions are detailed in Section V. The paper is concluded in Section VI.

International journal of imaging science and engineering

ISSN: 1934--9955 www.ijise.net

Vol-19 Issue-01 Feb 2024

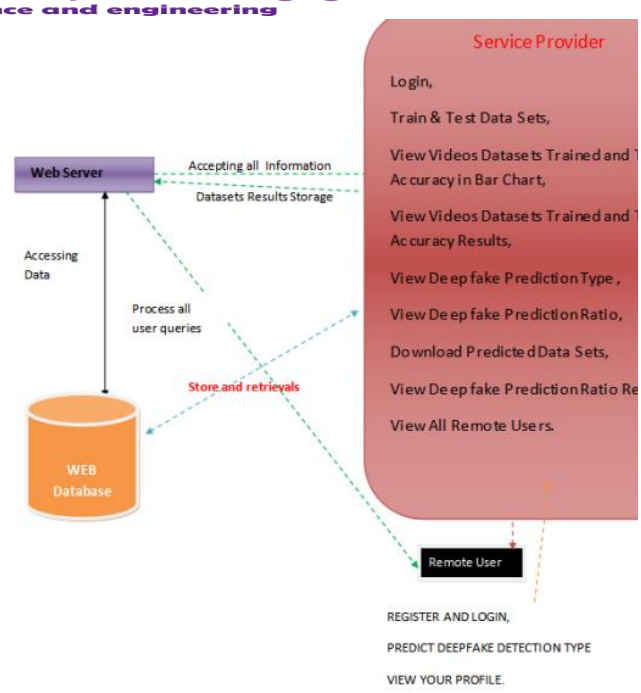

Fig1 : System Architecture

II. EXISTING SYSTEM

In order to verify the authenticity of GAN-generated films or pictures, the authors of [22] employ metrics for biological sign consistency in conjunction with spatial and temporal [23]_[25] directions to identify distinctive facial characteristics (such as the eyes, nose, mouth, etc.) that may be used as landmarks [26]. You may find Deepfake films with similar features by estimating the 3D head position [27].

Head motions are usually the first thing that come to mind when thinking about face expressions. Habee b.

In order to identify Deepfake videos using little processing resources, et al. [88] used MLP to take advantage of visual artefacts in the face area. In terms of performance, Deepfake algorithms based on machine learning have been shown to attain a detection accuracy of up to 98%. Nevertheless, the dataset type, feature selection, and test-set alignment are the three most important factors determining performance. By dividing the dataset into a train set and a test set at a certain ratio—for instance, 80% for the former and 20% for the latter—the research may get a better outcome in the experiment. It is an assertion without basis in reality that the unrelated dataset reduces performance by about half.

A GAN simulator was presented by Zhang et al. [33] that uses a classifier to detect Deepfake by simulating collective GAN-image artefacts and feeding them into the system. For the purpose of

International journal of imaging science and engineering
ISSN: 1934--9955 www.ijise.net
Vol-19 Issue-01 Feb 2024

typical feature extraction from RGB data, Zhou et al. [34] suggested a network, and [35] offered a comparable but general resolution. Researchers have suggested a novel detection framework in [36]_[38] that relies on physiological measurements like heartbeat.

For Deepfake video identification, the first proposal was to use a deep learning-based approach [40]. Their proposed network was built using two inception modules, Meso-4 and MesoInception-4. This method trains using a loss function that is the mean squared error (MSE) of the actual labels minus the predicted labels. Meso-4 has been suggested for improvement in [41].

Disadvantages

Binary coding does not maintain high-dimensional characteristics.

The owner retains control over the contents of the information since it is not kept in a permission-based Blockchain.

III. PROPOSED SYSTEM

_ We do an exhaustive literature review in the Deepfake field. By offering various research topics, we apprise readers of the existing tools, approaches, and datasets for study pertaining to Deepfake identification.

_ We provide a groundbreaking taxonomy that organises Deepfake detection algorithms into four distinct groups, outlining each group and their associated properties.

_ We examine the experimental data from the main research thoroughly. We also compare the effectiveness of several Deepfake detection algorithms based on a variety of criteria.

_ We draw attention to a few points and provide some recommendations on Deepfake identification that may be useful for future studies and applications in this area.

Advantages

Introduces a general Blockchain-based architecture that establishes a means of verifying the validity of digital material to its verified source. Details the architecture and design of the proposed solution to manage and oversee participant interactions and transactions. Brings together the essentials of Ethereum's Name service with IPFS's [114] decentralised storage capability.

IV. MODULES

Service Provider

A valid username and password are required for the Service Provider to access this module. Upon successful login, he will be able to do actions such as logging in, accessing the train and test data sets, You can see the trained and tested accuracy of video datasets in a bar chart. You can also see the types of deep fake predictions, the deep fake prediction ratio, the predicted data sets you can download, and the

ISSN: 1934--9955 www.ijise.net
Vol-19 Issue-01 Feb 2024

results of the deep fake prediction ratio. You can also see all the remote users.

View and Authorize Users

Here the administrator may get a complete rundown of all the registered users. Here, the administrator may see the user's information (name, email, and address) and grant them access.

Remote User

All all, there are n users in this module. Prior to performing any actions, the user must register. The user's information will be entered into the database after they register. He will be prompted to provide his authorised user name and password upon successful registration. After logging in, users will be able to do things like see their profile, make predictions about the sort of deepfake detection, and register and login.

V.CONCLUSION

From 2018 to 2020, 112 research published several state-of-the-art approaches for Deep fake detection, which are presented in this SLR. In this study, we introduce fundamental methods and examine the performance of several detection models.

Here is a brief overview of the study:

_ The FF++ dataset makes up the bulk of the trials. _ Deep fake detection algorithms based on deep learning are frequently employed. _ A considerable portion of all models are deep learning (mostly CNN) models. Detection accuracy is the most used performance measure. _ Deep fake can be successfully detected using deep learning approaches, according to the experimental findings. It may also be said that deep learning models generally perform better than non-deep learning models. Deep fake detection is still facing a lot of problems because to the fast advancements in multimedia technology and the abundance of tools and apps. We are hopeful that this SLR will be a helpful tool for researchers in their pursuit of more efficient detection approaches and mitigation strategies.

VI.REFERENCES

[1] *FaceApp*. Accessed: Jan. 4, 2021. [Online]. Available: https://www. faceapp.com/

[2] *FakeApp*. Accessed: Jan. 4, 2021. [Online]. Available: https://www. fakeapp.org/

[3] G. Oberoi. *Exploring DeepFakes*. Accessed: Jan. 4, 2021. [Online]. Available: https://goberoi.com/exploring-deepfakes-20c9947c22d9

[4] J. Hui. *How Deep Learning Fakes Videos (Deepfake) and How to Detect it*. Accessed: Jan. 4, 2021. [Online]. Available: https://medium. com/how-deep-learning-fakes-videos-deepfakes-and-how-to-detect-itc0b50fbf7cb9

[5] I. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D. W. Farley, S. Ozair, A. Courville, and Y.

Bengio, ``Generative adversarial nets," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, vol. 2. Cambridge, MA, USA: MIT Press, 2014, pp. 2672_2680.

[6] G. Patrini, F. Cavalli, and H. Ajder, ``The state of deepfakes: Reality under attack," Deeptrace B.V., Amsterdam, The Netherlands, Annu. Rep. v.2.3., 2018. [Online]. Available: https://s3.eu-west-2.amazonaws.com/rep2018/2018-the-state-of-deepfakes.pdf[7] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Niessner, ``Face2Face: Real-time face capture and reenactment of RGB videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2387_2395, doi: 10.1109/CVPR.2016.262.

[8] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, ``Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Venice, Oct. 2017, pp. 2242_2251, doi: 10.1109/ICCV.2017.244.

[9] S. Suwajanakorn, S. M. Seitz, and I. K. Shlizerman, ``Synthesizing Obama: Learning lip sync from audio," *ACM Trans. Graph.*, vol. 36, no. 4, p. 95, 2017.

[10] L. Matsakis. *Arti_cial Intelligence is Now Fighting Fake Porn*. Accessed: Jan. 4, 2021. [Online]. Available: https://www.wired.com/story/gfycatarti _cial-intelligence-deepfakes/

[11] A. Rössler, D. Cozzolino, L.Verdoliva, C. Riess, J. Thies, and M. Nieÿner, ``FaceForensics: A large-scale video dataset for forgery detection in human faces," 2018, *arXiv:1803.09179*.

[12] H. Kim, P. Garrido, A. Tewari, W. Xu, J. Thies, M. Niessner, P. Pérez, C. Richardt, M. Zollhöfer, and C. Theobalt, ``Deep video portraits," *ACM Trans. Graph.*, vol. 37, no. 4, pp. 1_14, Aug. 2018, doi: 10.1145/3197517.3201283.

[13] C. Chan, S. Ginosar, T. Zhou, and A. A. Efros, ``Everybody dance now," 2018, *arXiv:1808.07371*.