2023 6th International Conference on Advances in Science and Technology (ICAST) | 979-8-3503-5981-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICAST59062.2023.10454938

Review: DeepFake Detection Techniques using Deep Neural Networks (DNN)

Harsh Chotaliya[#1] , Mohammed Adil Khatri [#2], Shubham Kanojiya [#3], Mandar Bivalkar[#4]
Department of Computer Engineering
K J Somaiya Institute of Technology, Mumbai
Mumbai, India - 400022

harsh.gc@somaiya.edu[#1], mohammedadil.k@somaiya.edu [#2], shubham.kanojiya@somaiya.edu [#3], mbivalkar@somaiya.edu [#4]

Abstract – In the age of advanced digital manipulation, deepfake videos pose a significant threat to society by allowing the creation of highly convincing counterfeit footage. The application of deep learning has proven instrumental in tackling an extensive array of practical issues and real-world applications. However, alongside its substantial benefits, there exist certain disadvantages. Deepfake videos involve the substitution of one person's characteristics with those of another, achieved through the application of advanced Deep Learning techniques. This technology can be exploited with harmful intentions, leading to the dissemination of misinformation, manipulation, and persuasive content. This paper explores multiple deep learning techniques designed for detecting deep fake images and videos. It conducts a comparative analysis of these techniques, including CNN models like ResNet, VGG16, and Efficient Net, along with RNN models like LSTM, to assess their effectiveness in deepfake video detection.

Keywords: DeepFake detection, CNN, GAN, RNN, LSTM.

I. INTRODUCTION

In the era of false news, people and society as a whole fear that they will be unable to believe anything they see online[1]. It is dominated by digital technology and the widespread availability of powerful machine learning algorithms and the creation of deep fake videos has emerged as a significant concern.

The word "deepfake" refers to a specific subset of the production of synthetic media and is derived from the merger of "deep learning" and "fake." It entails the application of artificial intelligence (AI) and deep learning methodologies to substitute the appearance of an individual in an already-existing photo or video combined with someone else's. Synthetic media, known as deepfakes, have become increasingly convincing due to advancements in technologies like Deep Learning along with the utilization of GAN (Generative Adversarial Networks). These advancements make It easier to create totally fake videos that appear authentic to viewers. Deep learning algorithms are employed for these reasons because the fabricated videos they produce closely resemble the originals.[2] .

Detecting such videos has become a formidable challenge, even with the utilization of advanced computational algorithms. These misleading videos are frequently created with the intention of harming others, such as by altering forensic evidence, discrediting a well-known political figure and false photos of well-known actresses and actors. Since the fictitious videos produced by deep learning techniques resemble the genuine footage quite a bit, these approaches are employed for these purposes.[1]

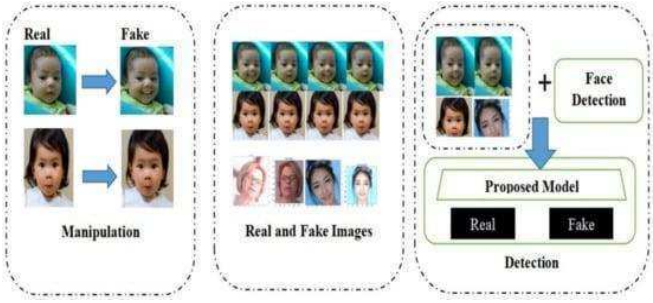

Fig. 1. A high-level view of the model for creating and detecting deep fakes[3].

Deepfake technology enables the generation of deceptive videos that mimic genuine ones as shown in figure 1. This can involve replacing one person's entire face with another's or simply altering eye movements, lip movements, and expressions to create a false impression of a different situation. Because of the proliferation of deepfake videos, which are

979-8-3503-5981-7/23/$31.00 ©2023 IEEE

being used more and more to create a variety of deceptive content, from fake news to fraudulent uses like celebrity pornography, the rapid advancement of deepfakes has drawn significant attention from both academia and the technology industry towards developing automated detection methods.[1].

II. LITERATURE SURVEY

Deepfake video detection is a challenging issue that has numerous suggested solutions, each of which has advantages and disadvantages. A thorough examination of the current status of Deepfake detection methods enumerates a few of these methods.

1. Convolution Neural Network(CNNs) with LSTM Architecture:

Combining CNNs with LSTM networks is considered to be a robust strategy for the videos that are deepfaked. This approach leverages the strengths of CNNs for analysis of image and LSTMs for handling temporal information in videos.

In [4] the authors introduced CLRNet, a Convolutional LSTM Residual Network that differs from others by accepting inputs of seq of parallel images from videos. This method successfully ca temporal information, making it possible to spot minute, artificial a in deepfake films.

[5] describes datasets which consist of video clips that include both genuine and deepfake data. Then, split these videos into individual frames and process them to make them suitable for analysis. To extract spatial features from these frames, CNN is used. We forward the characteristics to our LSTM layer after they have been discovered., where temporal sequences are used for face manipulation between frames.

Common CNN architectures like VGG or ResNet are often employed. On top of these fused features, we add a classification layer, typically a fully connected layer, to perform binary classification—distinguishing between genuine and deepfake sequences. The model is then trained using labeled video data, enabling it to make decisions by considering both spatial and temporal cues.

2. Recurrent neural network (RNN) technology with a DenseNet structure:

An advanced technique for detecting deep fake videos combines Recurrent Neural Networks (RNNs) with DenseNet architecture. This approach effectively examines video data by leveraging the strengths of RNNs and DenseNets. The model's structure comprises two components. The RNN layer, which can be LSTM or GRU, specializes in identifying temporal patterns and relationships over time within sequences of frames. This plays a role in detecting deep fake content by capturing how features evolve over time. The DenseNet structure functions as the core framework for extracting features from video frames. DenseNets are renowned for their ability to capture information through dense connections between layers. By merging the features obtained from the DenseNet with the patterns learned by the RNN layer a comprehensive representation of each video sequence is achieved. To differentiate between deep fake and genuine sequences, a classification layer (a fully connected layer) is added on top of this combined representation to perform binary classification.

3. General Adversarial Network (GAN):

An architecture for deep learning known as a Generative Adversarial Network (GAN) is composed of two fundamental parts: the discriminator and the generator. In contrast to the discriminator, whose job it is to discern between produced and actual images, the generator's primary goal is to generate graphics from sounds.

In [2] 70 thousand original Facial dataset, which is a subpart of the Style General Adversarial Network dataset, is one of the three datasets used. The Deep fake Detection Challenge (DFDC) dataset was used, and a number of procedures were followed to turn videos into images for examination. In order to do this, the DFDC dataset had to be preprocessed, and three different models had to be trained before being merged using an ensemble method. A pre-trained discriminator was employed in the first model, and it was subsequently enhanced with the help of a GAN. The DFDC dataset served as the foundation for this pre-trained discriminator's initial learning, and it then underwent additional training using a mixture of the 70 thousand original Facial datasets and CelebA. Once a GAN was used for training, the second model was created using the DFDC dataset. And finally, there was a third model—a GANs classifier that had been trained using the Style General Adversarial Network dataset. Using an ensemble approach, these three models were carefully integrated to improve the accuracy. Table-I shows the comparison of existing technologies.

TABLE I- COMPARISON BETWEEN THE EXISTING TECHNOLOGIES

Sr. No	Reference of paper	Dataset in use	Methods used	Accuracy
1	[6]	HOHA dataset	1. CNN (Convolution Neural Networks). 2. LSTM (Long Short-Term Memory).	Conv-LSTM (with 20 frames) 96.7%, Conv-LSTM (with 40 frames) 97.1%
2	[7]	FaceForensics++	1.LRP and LIME 2.Xception net(CNN)	90.17%
3	[8]	Face2Face	1.CNN (Convolution Neural Networks).	VGG16 : 81.61%, ResNet50 : 75.46%
4	[9]	AFW, FDDB, CelebA	1.Convolution Neural Networks (CNN)	Discrete- 95% and for continuous 74%
5	[10]	Face2Face, Reddit user deepfakes	1.CNN 2.LSTM	95%
6	[11]	1.Celeb-DF 2. FaceForensics++ 3.DeepFake Detection Challenge	1. LSTM 2. CNN	84% With Transfer Learning 75% Without Transfer Learning
7	[12]	Face2Face, StarGAN, CycleGAN	Deep Neural Networks, LSTM, MesoNet	20 videos accuracy is 85%
8	[13]	High-quality 1080p HD video clips of 976 sequences	FlowNet-S CNN With SPMC layer	Method(F3) 36.71/0.96, M(F5) 36.62/0.96

III. PROPOSED SOLUTION

Datasets: Numerous datasets have been tested such as DeepFake, FaceSwap, Face2Face, and DeepFake Detection. In the FaceForensics++ dataset, All except DFD had more than 1,000 movies in the dataset; 750 videos out of 1000 videos were utilized for training, 125 out of 1000 for validation, and another 125 videos for the phase of testing. For DFD, a smaller subset of 300 videos (250 for training, and 50 for testing) is selected.

Preprocessing: Taking 16 samples from both real and fake videos, with each sample five consecutive frames will be extracted. To identify facial landmarks within these frames, using the Multi-task CNN model. Subsequently, cropped and aligned the faces to the center of each image using this landmark information. All frames were then uniformly resized to a resolution of 240 × 240 pixels.

Training: This stage involves employing training the model with transfer learning on several datasets with a limited number of samples once it has been trained on a well-known deep fake dataset. Specifically, the model was trained using 16 samples extracted from both the 750 real videos and 750 fake video datasets. This strategy leverages the knowledge gained from one dataset to enhance the model's performance on various other datasets, even when data is limited.

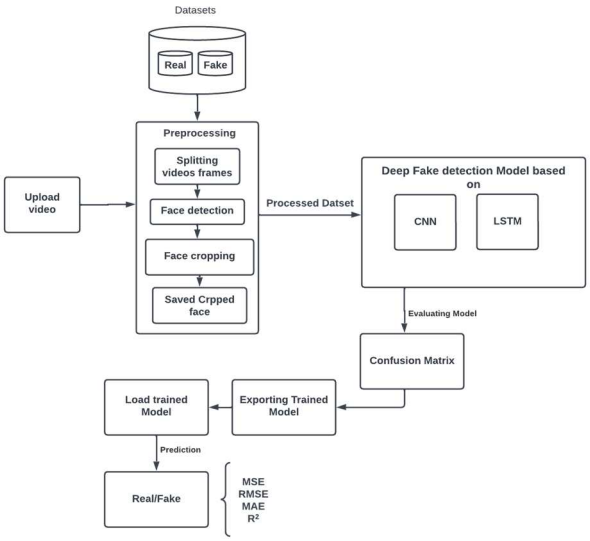

Fig. 2. Workflow for detection of Deep Fake videos

Once the original model has been trained using the main dataset, the transfer learning approach is used to adapt other datasets using a small sample size of just 10 videos from each target dataset. Specifically, for the first 120 frozen layers of the CLRNet model, and then made adjustments to the model using the same number of clips (10 films from each source and target dataset).

Evaluation Metrics: For the evaluation of CLRNet, Recall, Precision, F1-Score metrics are considered. The model utilized five consecutive frames as input. The pair of training and testing datasets, we kept the identical proportion of genuine and fictitious photos to provide a fair assessment, aiming to mitigate any potential bias caused by data imbalances or any less accuracy of models.

IV. CONCLUSION

The rapid advancement in the quality of deep fake videos necessitates a corresponding acceleration in the development of detection methods. Deepfake producers and catchers are facing more and more competition every day. The idea that deep learning approaches can address the issues brought by those similar deep learning techniques that are responsible for creating these misleading movies is what motivated the development of the suggested deepfake detection model.

A range of techniques rooted in Machine Learning and Deep Learning, employing diverse feature sets, are deployed for the categorization of videos as genuine or manipulated. Among the various methods, Higher accuracy in video categorization has continuously been shown when CNN and LSTM networks are integrated. To achieve this, researchers have employed numerous datasets containing both authentic and counterfeit videos for training and evaluation. Extensive research findings confirm that combining CNN and LSTM models results in superior accuracy and effectiveness in discerning fake videos.

REFERENCES

[1] D. Pan, L. Sun, R. Wang, X. Zhang and R. O. Sinnott, "Deepfake Detection through Deep Learning," 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT), Leicester, UK, 2020, pp. 134-143, doi: 10.1109/BDCAT50828.2020.00001.

[2] S. A. Aduwala, M. Arigala, S. Desai, H. J. Quan and M. Eirinaki, "Deepfake Detection using GAN Discriminators," 2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService), Oxford, United Kingdom, 2021, pp. 69-77, doi: 10.1109/BigDataService52369.2021.00014.

[3] Bansal, N.; Aljrees, T.; Yadav, D.P.; Singh, K.U.; Kumar, A.; Verma, G.K.; Singh, T. Real-Time Advanced Computational Intelligence for Deep Fake Video Detection. Appl. Sci. 2023, 13, 3095. https://doi.org/10.3390/app13053095

[4] Tariq, S., Lee, S. and Woo, S., A convolutional lstm based residual network for deepfake video detection. arXiv 2020. arXiv preprint arXiv:2009.07480.

[5] V. Jolly, M. Telrandhe, A. Kasat, A. Shitole and K. Gawande, "CNN based Deep Learning model for Deepfake Detection," 2022 2nd Asian Conference on Innovation in Technology (ASIANCON), Ravet, India, 2022, pp. 1-5, doi: 10.1109/ASIANCON55314.2022.9908862.

[6] David Guera, Edward J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks", 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS).

[7] Badhrinarayan Malolan, Ankit Parekh, Faruk Kazi ,"Explainable Deep-Fake Detection Using Visual Interpretability Methods",2020 3rd International conference on Information and Computer Technologies(ICICT).

[8] Irene Amerini, Leonardo Galteri, Roberto Caldelli, Alberto Del Bimbo, "Deepfake Video Detection through Optical Flow based CNN", 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW).

[9] Mingzhu Luo, Yewei Xiao, Yan Zhou," Multi-scale face detection based on convolutional neural network",IEEE 2018

[10] Digvijay Yadav, Sakina Salmani, "Deepfake: A Survey on Facial Forgery Technique Using Generative Adversarial Network", Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2019).IEEE Xplore Part Number: CFP19K34-ART; ISBN: 978-1-5386-8113-8.

[11] Ranjan, Sarvesh Patil, Faruk Kazi,"Improved Generalizability of Deep-Fakes Detection Using Transfer Learning Based CNN Framework",(IEEE 2020).

[12] L. Rebello, L. Tuscano, Y. Shah, A. Solomon and V. Shrivastava, "Detection of Deepfake Video using Deep Learning and MesoNet," 2023 8th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2023, pp. 1022-1026, doi: 10.1109/ICCES57224.2023.10192854.

[13] Xin Tao, Hongyun Gao, Renjie Liao, Jue Wang, Jiaya Jia, “Detail-revealing Deep Video Super-resolution”, 2017 IEEE International Conference on Computer Vision (ICCV).