International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

# DEEP FAKE IMAGES AND VIDEOS DETECTION USING DEEP LEARNING

*GANDLA PRAVALLIKA, Ms. B. TEJASWINI*
*MCA Student, Assistant Professor*
*DEPT OF MCA*
*PVKK INSTITUTE OF TECHNOLOGY(AUTONOMOUS), Anantapuramu – 515001 (A.P)*
*pravallikagandla71@gmail.com, tejaswinib78@gmail.com*

**ABSTRACT**

Deep fakes are altered, high-quality, realistic videos/images that have lately gained popularity. Many incredible uses of this technology are being investigated. Malicious uses of fake videos, such as fake news, celebrity pornographic videos and financial scams are currently on the rise in the digital world. As a result, celebrities, politicians, and other well-known persons are particularly vulnerable to the Deep fake detection challenge. Numerous research has been undertaken in recent years to understand how deep fakes function and many deep learning-based algorithms to detect deep fake videos or pictures have been presented.

This study comprehensively evaluates deep fake production and detection technologies based on several deep learning algorithms. In addition, the limits of current approaches and the availability of databases in society will be discussed. A deep fake detection system that is both precise and automatic. Given the ease with which deep fake videos/images may be generated and shared, the lack of an effective deep fake detection system creates a serious problem for the world. However, there have been various attempts to address this issue, and deep learning-related solutions outperform traditional approaches. These capabilities are used to train a ResNext which learns to categorize if a video has been concern to manipulation or now no longer and is also capable of hit upon the temporal inconsistencies among frames presented by DF introduction tools.

**Index Terms**—-Deep Fakes, Deep Learning, Fake Generation, Fake Detection, Machine Learning.

## 1. INTRODUCTION

### 1.1 MOTIVATION

The deep fake generation and detection technologies based on several deep learning algorithms are thoroughly assessed in this project. Furthermore, the limitations of existing methodologies and the accessibility of databases across society will be examined. An automated technique for deepfake detection that is accurate. The absence of an efficient deep fake detection system poses a major threat to the global community, given the simplicity with which deepfake movies and pictures may be created and distributed. There have been many efforts to solve this problem, however, and deep learning-related solutions work better than conventional methods.

### 1.2 PROBLEM DEFINITION

Due to the huge loss of frame content during video compression, existing deep learning algorithms for image identification cannot effectively detect bogus videos. The severe deterioration of the frame data following video compression prevents the majority of image recognition techniques from being employed for videos. Additionally, videos provide a problem for techniques intended to identify only still fake images since their temporal features vary across sets of frames.

### 1.3 OBJECTIVE OF PROJECT

A framework on which low-level face manipulation defects are expected to further appear as temporal distortions with irregularities between the frames. However, deep learning algorithms frequently employ face photos from the internet that typically display people with wide eyes; fewer pictures of persons with closed eyes may be seen online. As a result, deep fake algorithms are unable to generate fake faces that blink often in the absence of photographs of actual people doing so. Deep fakes, in other words, have far lower blink rates than regular videos.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

## 1.4 SCOPE OF PROJECT

Detecting deep fake images and videos using deep learning techniques is an important and evolving area of research and development. The scope of this field is broad, encompassing both technological advancements and the societal implications of deep fake technology. Here are some key aspects to consider within the scope of deep fake detection using deep learning techniques

## 2. LITERATURE SURVEY

### 2.1 Deepfake video detection using recurrent neural networks

**AUTHORS:** D. Guera and E. J. Delp,

**ABSTRACT:** In recent months a machine learning based free software tool has made it easy to create believable face swaps in videos that leaves few traces of manipulation, in what are known as "deepfake" videos. Scenarios where these realistic fake videos are used to create political distress, blackmail someone or fake terrorism events are easily envisioned. This project proposes a temporal-aware pipeline to automatically detect deepfake videos. Our system uses a convolutional neural network (CNN) to extract frame-level features. These features are then used to train a recurrent neural network (RNN) that learns to classify if a video has been subject to manipulation or not. We evaluate our method against a large set of deepfake videos collected from multiple video websites. We show how our system can achieve competitive results in this task while using a simple architecture.

### 2.2 Face x-ray for more general face forgery detection

**AUTHORS:** L. Li, J. Bao, T. Zhang, H. Yang, D. Chen, F. Wen, and B. Guo

**ABSTRACT:** In this project we propose a novel image representation called face X-ray for detecting forgery in face images. The face X-ray of an input face image is a greyscale image that reveals whether the input image can be decomposed into the blending of two images from different sources. It does so by showing the blending boundary for a forged image and the absence of blending for a real image. We observe that most existing face manipulation methods share a common step: blending the altered face into an existing background image. For this reason, face X-ray provides an effective way for detecting forgery generated by most existing face manipulation algorithms. Face X-ray is general in the sense that it only assumes the existence of a blending step and does not rely on any knowledge of the artifacts associated with a specific face manipulation technique. Indeed, the algorithm for computing face X-ray can be trained without fake images generated by any of the state-of-the-art face manipulation methods. Extensive experiments show that face X-ray remains effective when applied to forgery generated by unseen face manipulation techniques, while most existing face forgery detection or deepfake detection algorithms experience a significant performance drop.

### 2.3 Deepfakestack: A deep ensemblebased learning technique for deepfake detection

**AUTHORS:** M. S. Rana and A. H. Sung.

**ABSTRACT:** Recent advances in technology have made the deep learning (DL) models available for use in a wide variety of novel applications; for example, generative adversarial network (GAN) models are capable of producing hyperrealistic images, speech, and even videos, such as the so-called “Deepfake” produced by GANs with manipulated audio and/or video clips, which are so realistic as to be indistinguishable from the real ones in human perception. Aside from innovative and legitimate applications, there are numerous nefarious or unlawful ways to use such counterfeit contents in propaganda, political campaigns, cybercrimes, extortion, etc. To meet the challenges posed by Deepfake multimedia, we propose a deep ensemble learning technique called DeepfakeStack for detecting such manipulated videos. The proposed technique combines a series of DL based state-of-art classification models and creates an improved composite classifier. Based on our experiments, it is shown that

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

DeepfakeStack outperforms other classifiers by achieving an accuracy of 99.65% and AUROC of 1.0 score in detecting Deepfake. Therefore, our method provides a solid basis for building a Realtime Deepfake detector.

## 3. SYSTEM ANALYSIS

### 3.1 EXISTING SYSTEM:

Zhao et al. recently introduced a methodology for deep fake detection utilizing the self-consistency of local source features, which are spatially-local, content-independent details of pictures. A CNN model employs a unique representation learning approach to extract these source features, which are represented as down-sampled feature maps referred to as pairwise self-consistency learning. This aims to punish feature vector pairings that correspond to areas in the same picture with poor cosine similarity scores. When dealing with false pictures created by technologies that output the entire image directly and whose source features are constant throughout each point inside each image, it could have a disadvantage.

In past months, free deep learning-based software tools have made the creation of credible face exchanges in videos that leave few traces of manipulation, in what are known as "DeepFake"(DF) videos.

Manipulations of digital videos has been demonstrated for many years through the good use of visual effects, recent advances in deep learning have led to a drastic increase in the making real looking of fake content and the accessibility in which it can be created.

### DISADVANTAGES OF EXISTING SYSTEM:

- Since, fake image-based methods use error functions for real or fake image detection. For video, it needs lots of computational power and is hence time-consuming by using such methods.
- Some poorly created deep fake videos keep some visual artifacts behind, which can be used for deepfake detection. Thus we can group methods used for classification based on classifiers used i.e either deep or shallow.

### 3.2 PROPOSED SYSTEM:

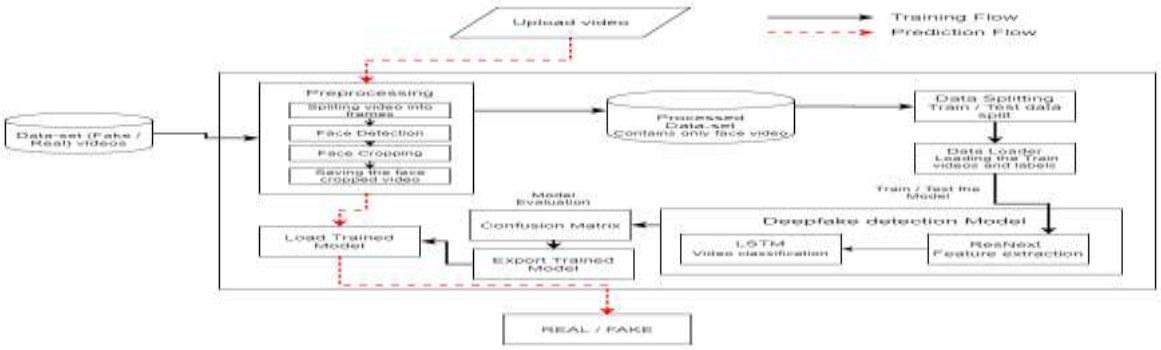
There are many tools available for creating the DeepFakes, but for DeepFakes detection there is hardly any tool available. Our approach for detecting the DF will be a great contribution in avoiding the percolation of the DF over the world wide web. We will be providing a web-based platform for the user for uploading the video and detect if its fake or real. This project is often scaled up from developing a web based platform to a browser plugin for automatic DF detections. Even big applications like WhatsApp, Facebook can integrate this project with their application for easy pre-detection of DF before sending it to another user. One of the important objectives is to evaluate its performance and acceptability in terms of security, user-friendliness, accuracy and reliability. Our method is focusing on detecting all types of DF like replacement DF, retrenchment DF and interpersonal DF.

### ADVANTAGES OF PROPOSED SYSTEM:

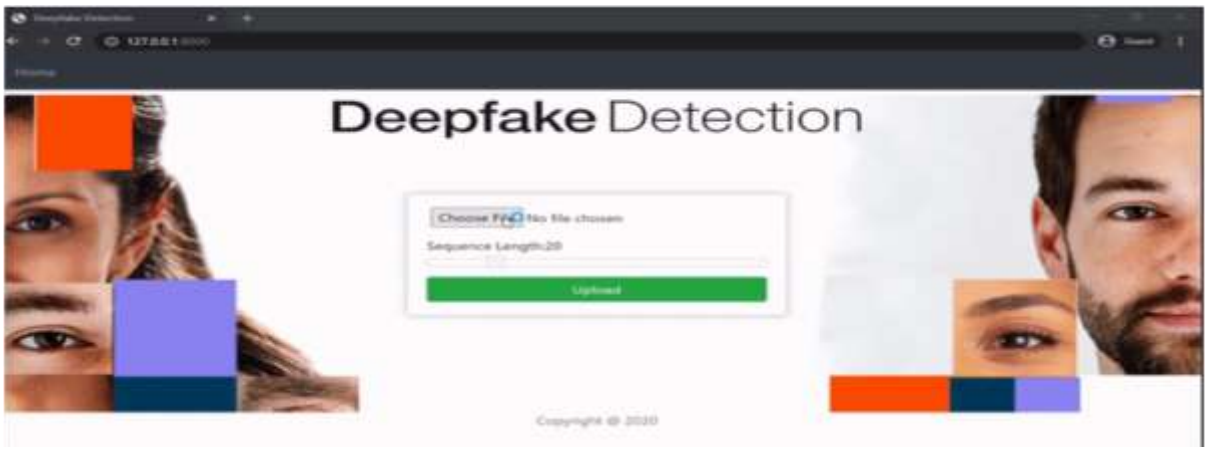
- Deep learning has shown considerable achievement in the identification of deep fakes.
- In order to recognize fake videos & photos properly must be enhanced current deep learning approaches.
- It primarily covers classic detection methods as well as deep Learning based methods such as CNN, RNN, and LSTM.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

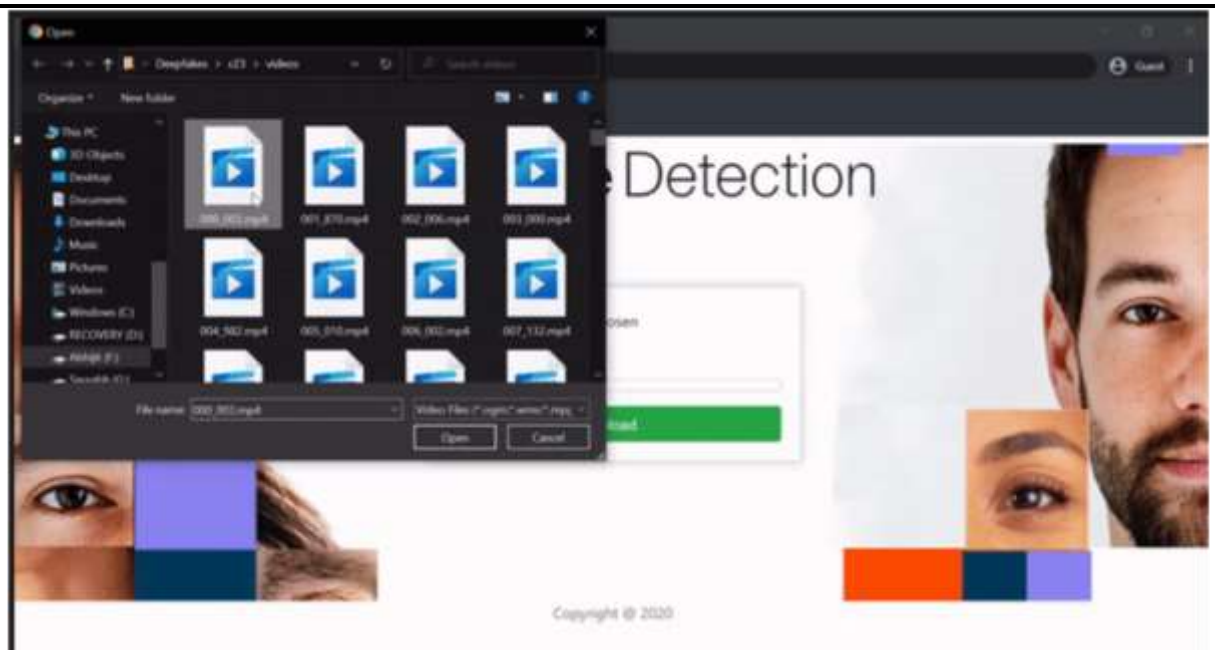
### 3.3 SYSTEM ARCHITECTURE:




### 4.SCREENSHOTS



International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper






## 5.CONCLUSION

Various researchers have created a number of deep-learning approaches for deep fake images and videos. Due to the extensive availability of photographs and videos in social media material, deep fakes had grown in popularity. This is especially crucial in social networking sites that make it simple for users to spread and share such fake information. Numerous deep learning-based approaches have recently been put out to deal

with this problem and effectively identify fake images and videos. The first section discussed the existing programs and technologies that are extensively used to make fake photos and videos. And in the second section discuss the different type of techniques that are used for deep fake images and videos. Also, provide details of available datasets and evaluation metrics that are used for deep fake detection. Despite the fact that deep learning has done well in detecting deep fakes, the quality of deep fakes has been increasing. In order to recognize fake videos & photos properly must be enhanced current deep learning approaches.

We provided a neural network-primarily based totally method to classify the video as deep fake or actual, at the side of the self-assurance of the proposed model. Our approach does the frame stage detection the use of ResNext CNN and video class the use of LSTM. The proposed approach is successful in detecting the video as a deep fake or actual primarily based totally on the listed parameters in the project. We consider that it'll offer a very excessive accuracy on actual time data.

## FUTURE ENHANCEMENT

Furthermore, given present deep learning approaches, it is unknown how to identify the number of layers necessary and the appropriate architecture for deep fake detection. To improve their capacity to cope with the ubiquitous impacts of deep fakes and mitigate their consequences, social media companies are integrating deep fake detection tools.

## REFERENCES

[1] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” arXiv preprint arXiv:1411.1784, 2014.

[2] Y. Bengio, P. Simard, and P. Frasconi, “Long short-term memory,” IEEE Trans. Neural Netw, vol. 5, pp. 157–166, 1994.

[3] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press, 2016. [4] S. Hochreiter, “Ja1 4 rgen schmidhuber (1997).“long short-term memory”,” Neural Computation, vol. 9, no. 8.

[5] M. Schuster and K. Paliwal, “Networks bidirectional reccurent neural,” IEEE Trans Signal Proces, vol. 45, pp. 2673–2681, 1997.

[6] J. Hopfield et al., “Rigorous bounds on the storage capacity of the dilute hopfield model,” Proceedings of the National Academy of Sciences, vol. 79, pp. 2554–2558, 1982.

[7] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al., “Google’s neural machine translation system: Bridging the gap between human and machine translation,” arXiv preprint arXiv:1609.08144, 2016.

[8] L. Nataraj, T. M. Mohammed, B. Manjunath, S. Chandrasekaran, A. Flenner, J. H. Bappy, and A. K. Roy-Chowdhury, “Detecting gan generated fake images using co-occurrence matrices,” Electronic Imaging, vol. 2019, no. 5, pp. 532–1, 2019.

[9] B. Zi, M. Chang, J. Chen, X. Ma, and Y.-G. Jiang, “Wilddeepfake: A challenging real-world dataset for deepfake detection,” in Proceedings of the 28th ACM international conference on multimedia, 2020, pp. 2382– 2390.

[10] H. A. Khalil and S. A. Maged, “Deepfakes creation and detection using deep learning,” in 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC). IEEE, 2021, pp. 1–4.

[11] J. Luttrell, Z. Zhou, Y. Zhang, C. Zhang, P. Gong, B. Yang, and R. Li, “A deep transfer learning approach to fine-tuning facial recognition models,” in 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE, 2018, pp. 2671–2676.

International Journal of
DATA SCIENCE AND IOT MANAGEMENT SYSTEM
Peer Reviewed, Referred & Indexed Journal
ISSN: 3068-272X www.ijdim.com Original Research Paper

[12] S. Tariq, S. Lee, H. Kim, Y. Shin, and S. S. Woo, "Detecting both machine and human created fake face images in the wild," in Proceedings of the 2nd international workshop on multimedia privacy and security, 2018, pp. 81–87.

[13] N.-T. Do, I.-S. Na, and S.-H. Kim, "Forensics face detection from gans using convolutional neural network," ISITC, vol. 2018, pp. 376–379, 2018.

[14] X. Xuan, B. Peng, W. Wang, and J. Dong, "On the generalization of gan image forensics," in Chinese conference on biometric recognition. Springer, 2019, pp. 134–141.

[15] P. Yang, R. Ni, and Y. Zhao, "Recapture image forensics based on laplacian convolutional neural networks," in International Workshop on Digital Watermarking. Springer, 2016, pp. 119–128.