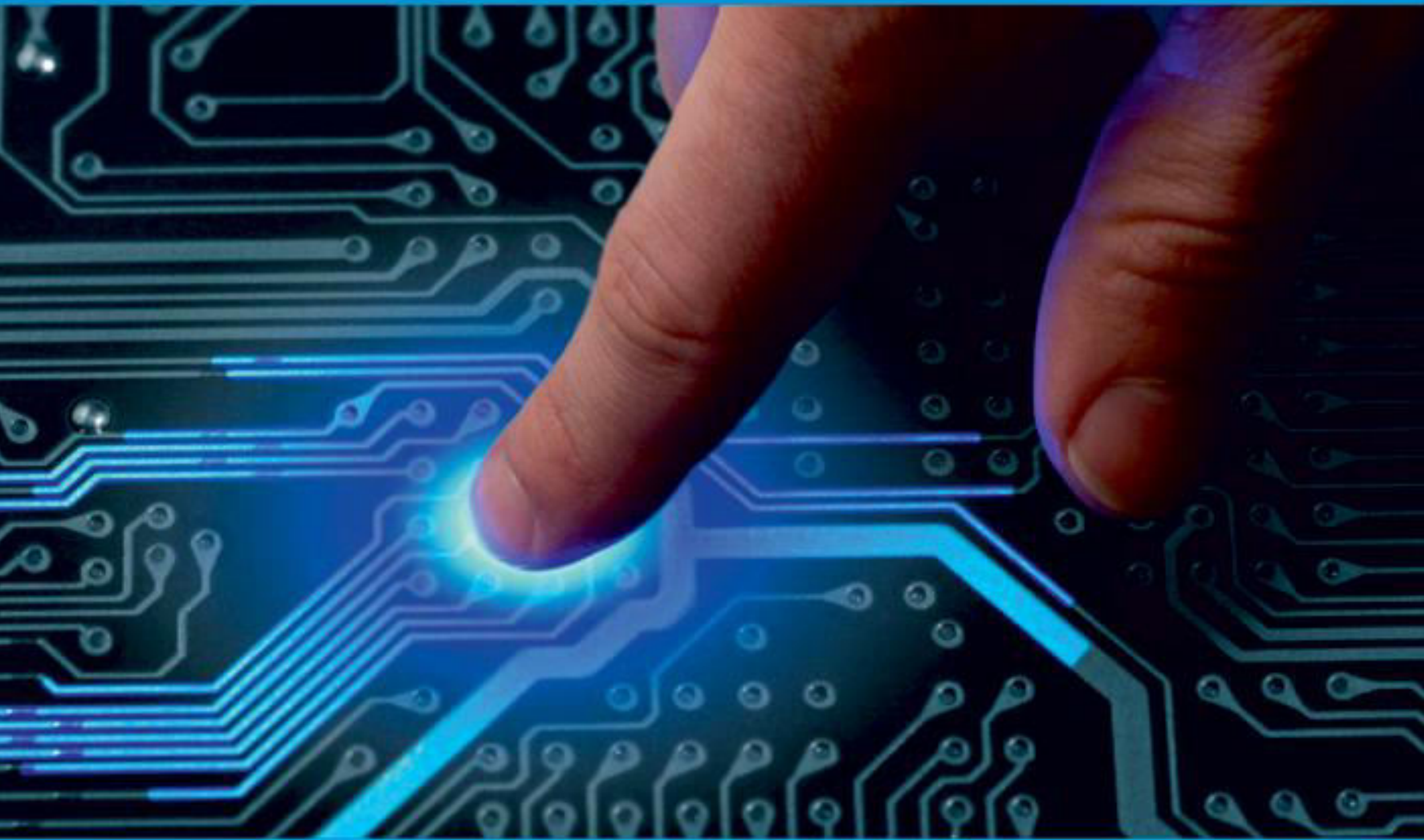
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 5, May 2024

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 | A Monthly Peer Reviewed & Referred Journal |

|| Volume 12, Issue 5, May 2024 ||

| DOI: 10.15680/IJIRCCE.2023.1205062 |

Deepfake Video Detection Using Deep Learning Techniques

K. Durga Bhavani[1], A. Bhavani[2], N. Dhilsha[3], K. Pavan[4], Ch. Thribhuvan Chowdary[5]

Assistant Professor, Department of Computer Science and Engineering, SRKIT, Vijayawada, India[1]

Student, Department of Computer Science and Engineering, SRKIT, Vijayawada, India[2-5]

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 pornography 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 Deepfake detection challenge. Numerous research has been undertaken in recent years to understand how deepfakes functions 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 databased in society will be discussed. A deepfake detection systems 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 concer to manipulation or now no longer and is also capable of hit upon the temporal inconsistencies among frames presented by DF introduction tools.

KEYWORDS: Deep Fakes, Deep Learning, Fake Generation, Fake Detection, Machine Learning.

I. INTRODUCTION

Overview:

The project is designed in a way that the user can be able to use it as simple as using a Mobile. Here the user will be opening the website and he will be uploading the video at him and he will be checking whether the video uploaded by him is a deep fake or not.

About Project:

This website will be used to identify weather the video is real or not simply we can say it as deepfake identification. It is a step by step process first the user who will be using the website will be displayed an box where he should choose a file to be uploaded. Then he should be selecting the sequence length mentioned there and then he should be uploading the file in website by clicking on the upload button. Then the video will be uploaded in the website. After the video uploading the website will be using the CNN and RNN algorithms where this kind of algorithms will be helpful in splitting the video frame by frame. This would be helping in identifying the fake videos. Thus the frame by frame splitting will be displayed as a result and it will be giving the final result yes or no in the form of the thumb up for yes and thumb down for no.

II. RELATED WORK

1. **Deepfake video detection using recurrent neural networks**

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 paper 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. Face x-ray for more general face forgery detection

In this paper 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.

3. Deepfakestack: A deep ensemble based learning technique for deepfake detection

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 hyper realistic 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 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.

4. Detecting deepfake videos using attribution based confidence metric

Recent advances in generative adversarial networks have made detecting fake videos a challenging task. In this paper, we propose the application of the state-of-theart attribution based confidence (ABC) metric for detecting deepfake videos. The ABC metric does not require access to the training data or training the calibration model on the validation data. The ABC metric can be used to draw inferences even when only the trained model is available. Here, we utilize the ABC metric to characterize whether a video is original or fake. The deep learning model is trained only on original videos. The ABC metric uses the trained model to generate confidence values. For, original videos, the confidence values are greater than 0.94.

III. METHODOLOGY

A. *Problem-Solving*:

By this detecting the deepfake in video we will be able to control the fraud and scams such as impersonating individuals in video calls to deceive others into transferring money or revealing sensitive information. Deepfakes can be used to create fake identities for malicious purposes, such as accessing restricted areas, committing crimes or impersonating someone for social engineering attacks. Deepfakes pose a threat to the integrity of evidence in legal proceedings, as they can be used to fabricate incriminating or exculpatory evidence, undermining the justice system. Deepfakes can be used to create convincing fake videos or audios of public figures, leading to the spread of false information and manipulation of public opinion.

B. *Proposed Work*:

“Why should we use deepfake detection using deep learning?

- ✓ 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.
- ✓ In order to recognize fake videos & photos properly must be enhanced current.

The proposed approach is successful in detecting the video as a deep fake or actual primarily based totally on the listed parameters in the paper. We consider that it'll offer a very excessive accuracy on actual time data.

C. *Architecture:*

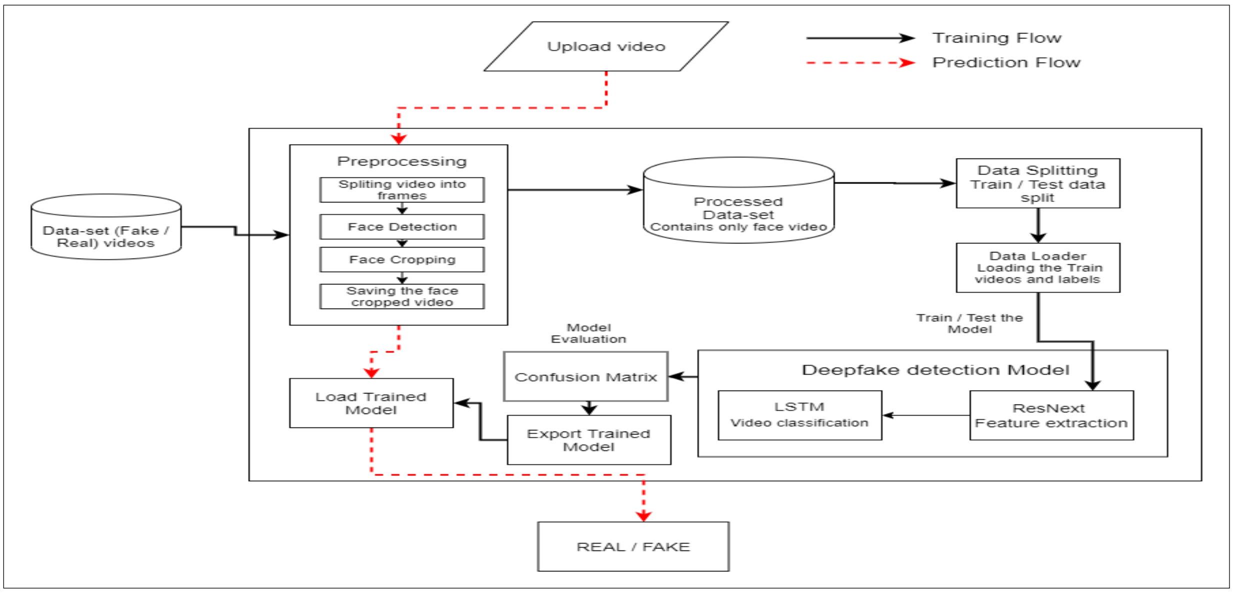

Upload video
Training Flow
Prediction Flow
Preprocessing
Spliting video into frames
Face Detection
Face Cropping
Saving the face cropped video
Data-set (Fake / Real) videos
Processed Data-set Contains only face video
Data Splitting Train / Test data split
Data Loader Loading the Train videos and labels
Train / Test the Model
Model Evaluation
Confusion Matrix
Deepfake detection Model
LSTM Video classification
ResNext Feature extraction
Load Trained Model
Export Trained Model
REAL / FAKE

Fig1: System Architecture

D. *Modules*

- **Dataset:** To built any machine learning and deep learning model we require a real-world data. First we collected data from different platform like Kaggle's Deepfake Detection challenge, Celeb-DF[8], Face Forensic. Kaggle's DeepFake detection challenge contains 3000 videos in which 50% data is real and 50% is manipulated data. Celeb-DF contains the videos of some famous celebrities and there are a total of 1000 videos in which 500 are real and 500 are manipulated videos Face Forensic ++ dataset contains a total of 2000 videos of which 1000 are real and the remaining are manipulated. Further all this three datasets are merged together and passed to the pre-processing of data.

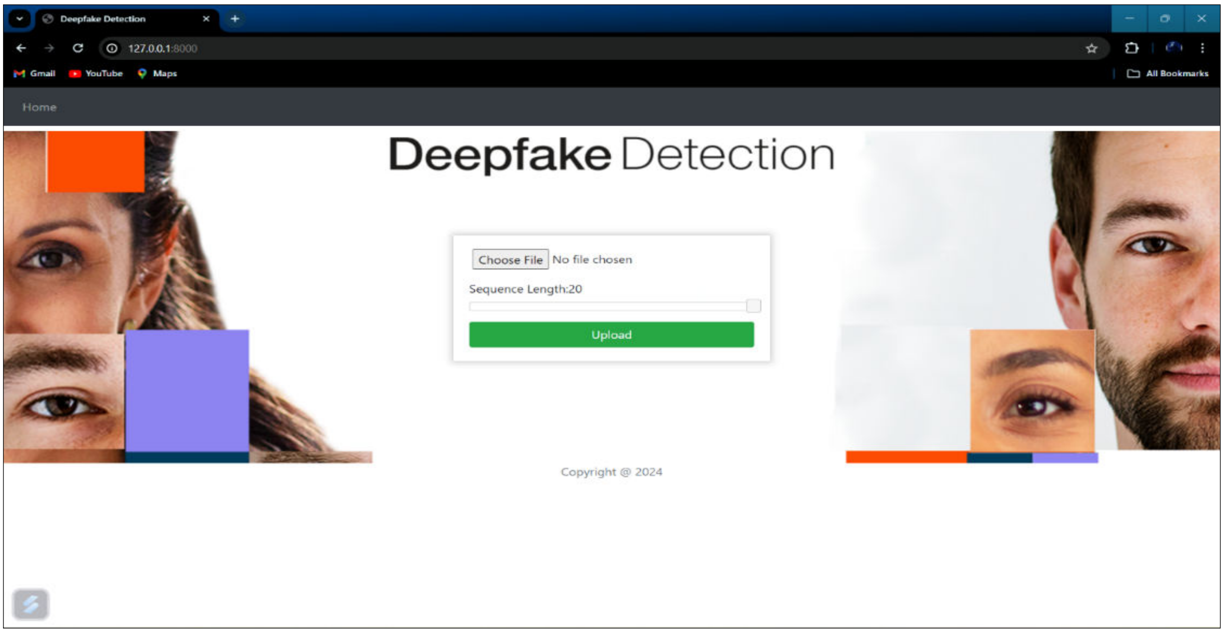
- **Data Pre-processing:** Pre-processing of data is a very important part as by doing pre-processing we actually try to get some important information from the data. We eliminate unnecessary data from original data. Splitting the movie into frames is part of the dataset pre-processing. Face detection is then performed, and the frame with the detected face is cropped. To preserve consistency in the number of frames, the mean of the video dataset is determined, and a new processed face cropped dataset containing the frames equal to the mean is constructed. During pre-processing, frames that do not include faces are ignored. Processing a 10-second movie at 30 frames per second, or 300 frames in total, will necessitate a significant amount of CPU power. So, for the sake of experimentation, we propose using only the first 100 frames to train the model.

- **Model:**The model is made up of resnext50 32x4d and one LSTM layer. The Data Loader loads the pre-processed face cropped films and divides them into two groups: train and test. In addition, the frames from the processed videos are supplied to the model in tiny batches for training and testing.

- **ResNextCNN for Feature Extraction:** We propose using the ResNext CNN classifier for extracting features and reliably recognizing frame-level characteristics instead of rewriting the classifier. Following that, we'll fine-tune the network by adding extra layers as needed and setting a correct learning rate to ensure that the gradient descent of the model is properly converged. LSTM for Sequence Processing: Assume a 2-node neural network with the probabilities of the sequence being part of a deep fake video or an untampered video as input and a sequence of ResNext CNN feature vectors of input frames as output. The main problem that we must solve is the design of a model that can recursively process a sequence in a meaningful way. For this task, we propose using a 2048 LSTM unit with a 0.4 likelihood of dropping out, which is capable of achieving our goal. The LSTM is used to analyze the frames sequentially in order to do a temporal analysis of the video by comparing the frame at't' second with the frame at't' second.

- **Predict:** The trained model is given a new video to forecast. A fresh video is also pre-processed to incorporate the trained model's format. The video is divided into frames, then face cropped, and instead of keeping the video locally, the cropped frames are sent immediately to the trained model for identification.

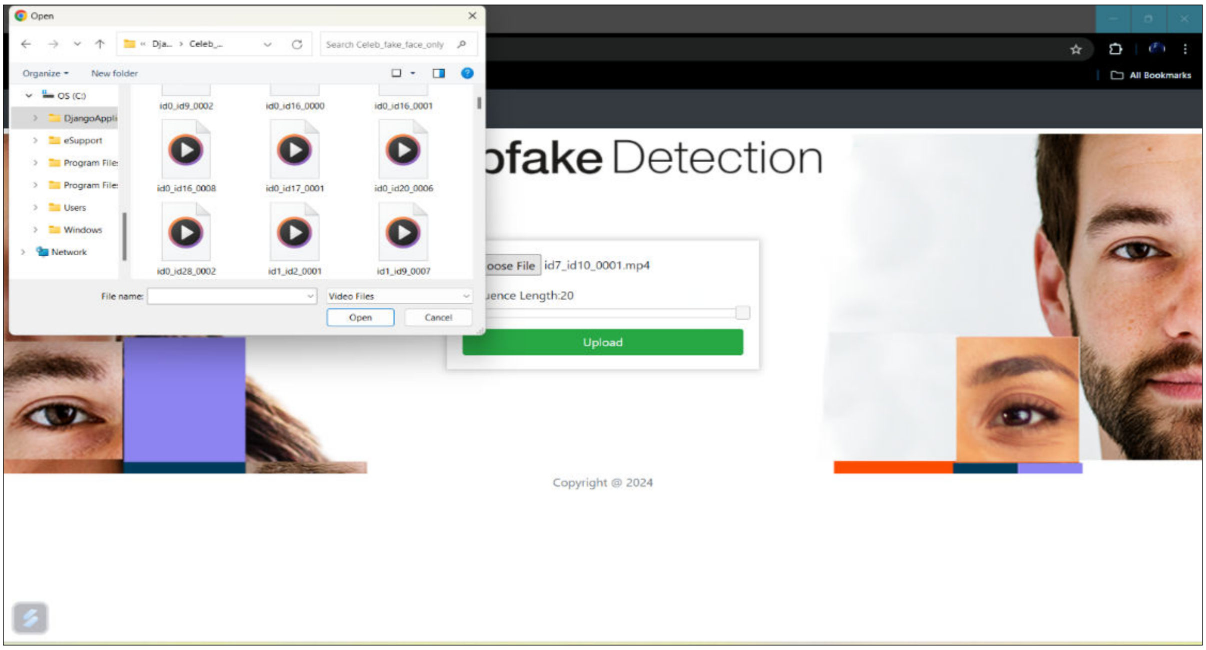
IV. Experimental Result



No file chosen

Sequence Length:20

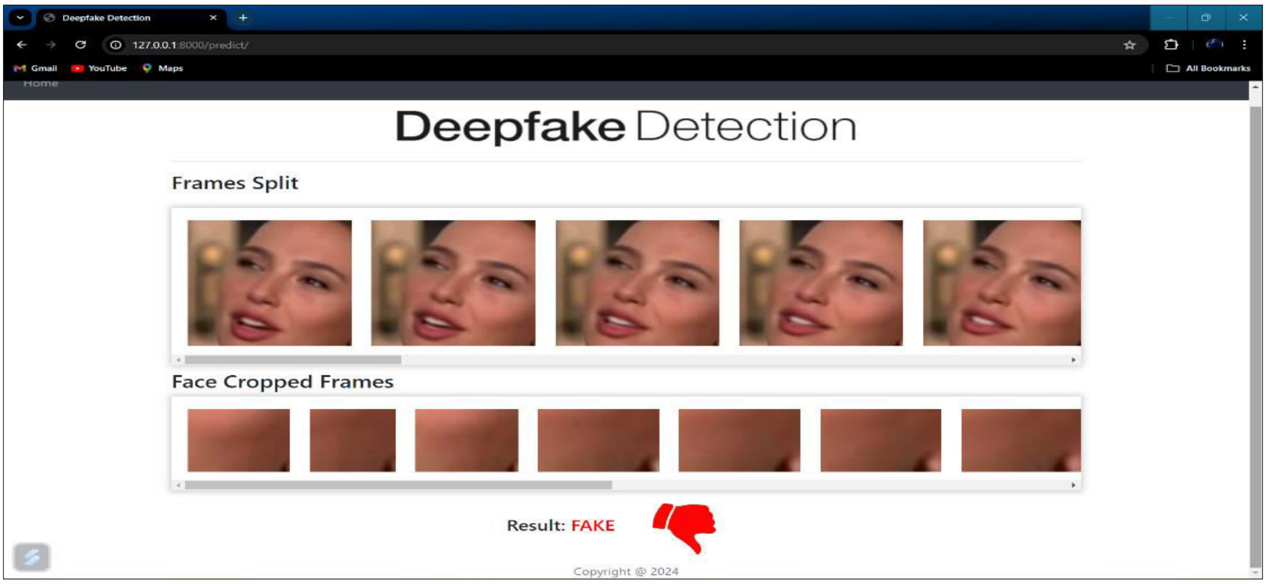
Fig 4.1 The above figure will be showing the video uploading webpage



id7_id10_0001.mp4

ence Length:20

Fig 4.2 In this above figure it will be showing the uploading process.



Deepfake Detection

Frames Split

Face Cropped Frames

Result: FAKE

Copyright @ 2024

Fig 4.3 In the above figure will be showing the uploaded video image splitting and the result.

V. Conclusion and Future Work

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 paper. We consider that it'll offer a very excessive accuracy on actual time data.

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. Good fellow, 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.
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.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

Scan to save the contact details

www.ijircce.com