IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13359

Deepfake Face Detection System

Mr. H.M. Gaikwad[1], Aryan Sonawane[2], Manavaditya Rathawa[3], Ratnali Pawar[4], Uday Talele[5]

Head of AIML Dept, K.K. Wagh Polytechnic, Nashik[1]

Third Year Students of Artificial Intelligence and Machine Learning, K.K. Wagh Polytechnic, Nashik[2-5]

Abstract: The growing computation power has made the deep learning algorithms so powerful that creating an indistinguishable human synthesized image popularly called as deep fakes have become very simple. Scenarios where this realistic face swapped deepfake are used to create political distress, fake terrorism events, revenge porn, blackmail people are easily envisioned. In this project, System will detect fake images that have been generated using AI. Deep Fakes are created by using deep learning techniques and neural networks to manipulate or replace parts of an original image, such as the face of a person.

This project is an important application of deep learning technology, which is characterized by its strong capability of feature learning and feature representation compared with the traditional image detection methods. System will describe a new deep learning-based method that can effectively distinguish AI-generated fake images from real images. System is capable of automatically detecting the replacement, reenactment deep fakes and trying to use Artificial Intelligence (AI) to fight Artificial Intelligence (AI). Our system uses a Res-Next Convolution Neural Network to extract frame-level features and these features and further train a Convolutional Neural Network (CNN) based InceptionResnetV1 and InceptionResnetV2 to classify whether the image is subject to any Type manipulation or not, i.e. image is deep fake or real image. It will allow us to detect deep fake images and can further help in reducing fake news.

Keywords: Deep Learning, Deepfake, Neural Network, Artificial Intelligence, InceptionResnetV1, InceptionResnetV2

I. INTRODUCTION

Deep fake is a technique for human image synthesis based on neural network tools like GAN (Generative Adversarial Network) or Auto Encoders etc. These tools superimpose target images onto source videos using a deep learning technique and create a realistic looking deep fake video. These deep-fake videos are so real that it becomes impossible to spot differences with the naked eye. In this work, we describe a new deep learning-based method that can effectively distinguish AI-generated fake videos from real videos. We are using the limitation of the deep fake creation tools as a powerful way to distinguish between the pristine and deep fake videos. During the creation of the deep fake the current deep fake creation tools leave some distinguishable artifacts in the frames which may not be visible to the human being, but the trained neural networks can spot the changes.

Deepfake creation tools leave distinctive artifacts in the resulting Deep Fake videos, and we show that they can be effectively captured by Res-Next Convolution Neural Networks. Our system uses a Res-Next Convolution Neural Networks to extract frame-level features. These features are then used to train a Long Short-Term Memory (LSTM) based Recurrent Neural Network (RNN) to classify whether the video is subject to any kind of manipulation or not, i.e., whether the video is deep fake or real video.

We proposed to evaluate our method against a large set of deep fake videos collected from multiple video websites. We are trying to make the deep fake detection model perform better on real time data. To achieve this, we trained our model on a combination of available datasets. So that our model can learn the features from different kinds of images. We extracted an adequate number of videos from Face-Forensic++ , Deepfake detection challenge, and Celeb-DF datasets. We also evaluated our model against the large amount of real time data like YouTube dataset to achieve competitive results in the real time scenarios.

In the world of ever growing social media platforms, Deepfakes are considered as the major threat of AI. There are many Scenarios where these realistic face swapped deep fakes are used to create political distress, fake terrorism events, revenge pornography, blackmail people are easily envisioned. Some of the examples are Brad Pitt, Barack Obama videos. It becomes very important to spot the difference between the deep fake and real images. We used AI to fight AI. Deepfakes are created using tools like FaceApp and Face Swap, which is using pre-trained neural networks like GAN or Auto encoders for these deepfakes creation.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13359

Our method uses a CNN based artificial neural network to process the sequential temporal analysis of the images and pre-trained CNN to extract the frame level features. Convolution neural networks extracts the frame-level features and these features are further used to train. To emulate the real time scenarios and make the model perform better on real time data, we trained our method with a large amount of balance and combination of various available dataset.

II. LITERATURE REVIEW

Code with AJ[5] used LSTM and Resnext CNN to detect deepfakes. Detection by Eye Blinking [6] describes a new method for detecting the deepfakes by the eye blinking as a crucial parameter leading to classification of the videos as deepfake or pristine. The Long-term Recurrent Convolutional Network (LRCN) was used for temporal analysis of the cropped frames of eye blinking. As today the deepfake generation algorithms have become so powerful that lack of eye blinking cannot be the only clue for detection of the deepfakes. There must be certain other parameters that must be considered for the detection of deepfakes like teeth enchantment, wrinkles on faces, wrong placement of eyebrows etc. Capsule networks to detect forged images and videos [7] uses a method that uses a capsule network to detect forged, manipulated images and videos in different scenarios, like replay attack detection and computer-generated video detection. In their method, they have used random noise in the training phase which is not a good option. Still the model performed beneficial in their dataset but may fail on real time data due to noise in training. Our method is proposed to be trained on noiseless and real time datasets. Recurrent Neural Network [8] (RNN) for deepfake detection used the approach of using RNN for sequential processing of the frames along with the ImageNet pre-trained model. Their process used the HOHO [9] dataset consisting of just 600 videos. Their dataset consists of a small number of videos and the same type of videos, which may not perform very well on the real time data. We will be training our model on a large number of Realtime data. Synthetic Portrait Videos using Biological Signals [10] approach extract biological signals from facial regions on pristine and deepfake portrait video pairs. Applied transformations to compute the spatial coherence and temporal consistency, capture the signal characteristics in feature vector and photoplethysmography (PPG) maps, and further train a probabilistic Support Vector Machine (SVM) and a Convolutional Neural Network (CNN). Then, the average of authenticity probabilities is used to classify whether the video is a deepfake or pristine. Fake Catcher detects fake content with high accuracy, independent of the generator, content, resolution, and quality of the video. Due to lack of discriminator leading to the loss in their findings to preserve biological signals, formulating a differentiable loss function that follows the proposed signal processing steps is not a straightforward process.

III. SYSTEM ARCHITECTURE

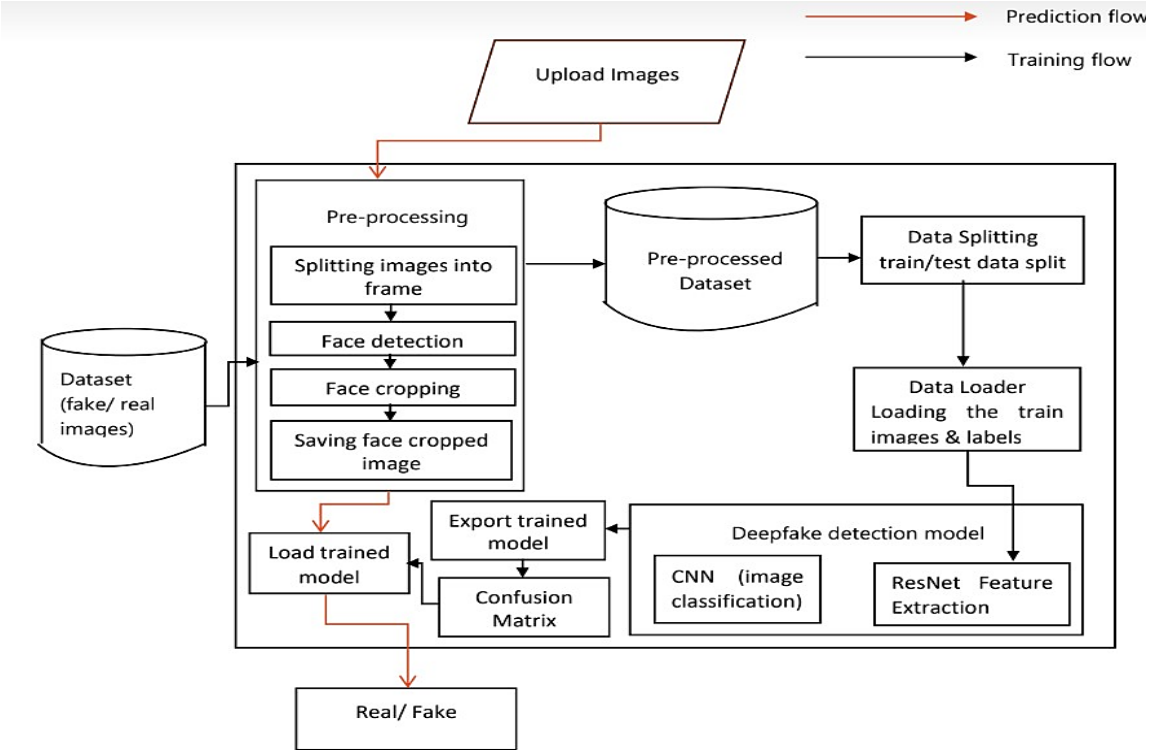

Fig.1 system architecture

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13359

In the first step, images are uploaded to the system. This dataset can include both authentic and manipulated images. The uploaded images are then pre-processed. This might involve resizing them to a standard size, converting them to a specific format, or normalizing the pixel values. If the uploaded content is video, it is split into individual frames at this stage. Next, faces are detected within the images or frames. Once faces are identified, they are cropped from the images or frames. The cropped faces are then saved as separate images. The processed images (which can be full images or cropped faces) are then split into two datasets: training data and testing data. The training data is used to train the deep fake detection model, while the testing data is used to evaluate the model's performance.

Training flow:

The training data is loaded into the system. A Convolutional Neural Network (CNN) is used to classify the images. The CNN is a type of deep learning model that is particularly well-suited for image recognition tasks. During training, the CNN learns to identify features in the images that are indicative of deep fakes. ResNet is a specific type of CNN architecture that is used in the model for feature extraction. In essence, it extracts a set of features from the images that will be used for classification. After That There is Prediction Flow, which illustrates what happens when a new image is uploaded for processing: After the prediction flow, The trained deepfake detection model is loaded. The CNN classifies the new image using the features it learned during training. The image is then classified as either real or fake. Once trained, the model can be exported and used to detect deepfakes in new images or videos. After everything, Confusion Matrix is used. The confusion matrix is a tool used to evaluate the performance of the model on the testing data. It helps to visualize how many images were correctly classified as real or fake, and how many were misclassified.

IV. METHODOLOGY

Deepfake detection often relies on a two-step process: pinpointing faces of interest and then analyzing them for signs of manipulation. Inception-ResNet and MTCNN are deep learning models that can be combined for this purpose. MTCNN, short for Multi-task Cascaded Convolutional Networks, tackles the first step: face detection. It utilizes a series of increasingly complex convolutional neural networks, acting like filters that scan the image and progressively identify potential faces with higher accuracy.

Once a face is located, Inception-ResNet steps in. This model excels at feature extraction, a crucial step in image recognition tasks. It analyzes the face in detail, extracting a unique mathematical representation called an "embedding." This embedding encodes information like facial landmarks, skin texture, and other subtle details. By comparing this embedding to a database of genuine faces, or by analyzing inconsistencies within the embedding itself (e.g., unrealistic smoothness or unnatural movements), Inception-ResNet can help determine if the face is real or a deepfake.

The power of this approach lies in the synergy between the two models. MTCNN efficiently locates potential faces, while Inception-ResNet provides a robust analysis of the extracted facial features. Additionally, both models are often pretrained on large datasets, allowing researchers to leverage their capabilities without the intensive computational cost of training from scratch. This makes Inception-ResNet and MTCNN valuable tools for streamlining deepfake detection research.

V. RESULT AND DISCUSSION

Fig 5.1 Output by uploading Image of Obama (Fake Image)

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024

DOI: 10.17148/IJARCCE.2024.13359

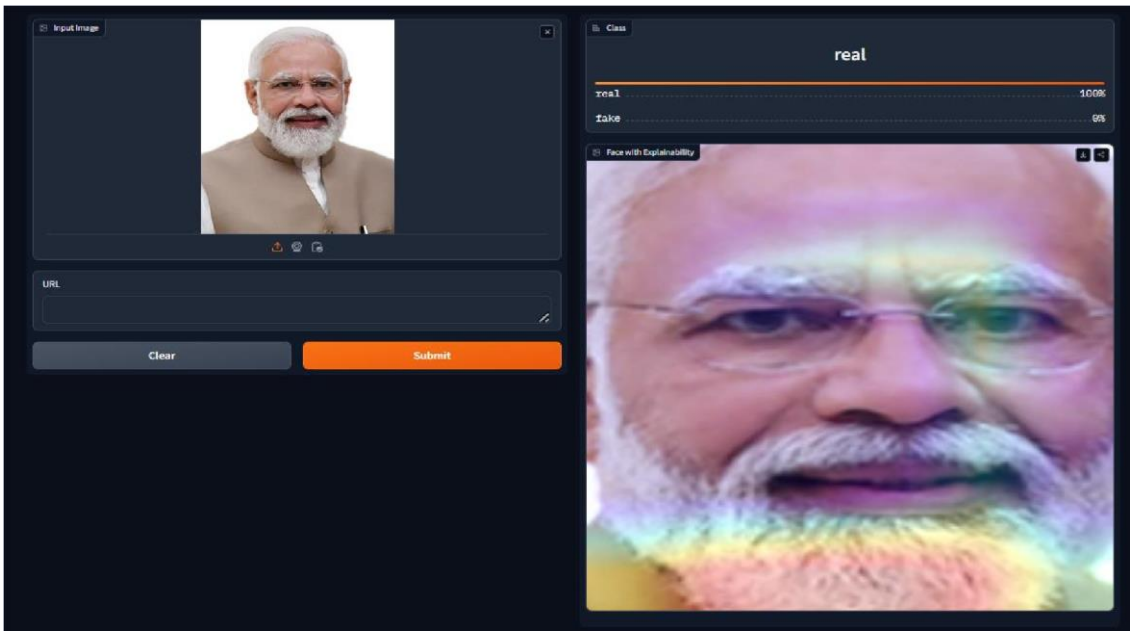

Fig 5.2 Output by uploading Image of PM. Narendra Modi (Real Image)

The Front end is build with grade Which is an open-source web framework for machine learning. There is a text box where you can paste a URL to an image, or upload an image file, that you want the tool to analyze for signs of being a deep fake. “Drop Image Here” is the drop zone for uploading an image file. You can drag and drop an image file here, or click "Click to Upload" to select a file from your computer. There is a progress bar just besides the ‘Drop image Here’ section where the tool will display the results of its analysis.

It will show either "Real" or "Fake" to indicate whether the tool believes the image is a real photo or a deep fake. There are three buttons, clear, submit and Flag. Clear button will clear the contents of the input field and the image upload area, and Submit button will start the analysis of the image you provided, Flag button allows you to flag the image as inappropriate or incorrect.

VI. CONCLUSION

In conclusion, deepfake face image detection represents a critical and evolving field of research and technology with significant implications for society, security, and digital integrity. As the capabilities of deepfake technology continue to advance, so too must our efforts to detect and mitigate its potential misuse.

Through the development of sophisticated algorithms, multi-modal approaches, and human-in- the-loop systems, researchers and technologists are striving to enhance the accuracy and effectiveness of deepfake detection methods. However, challenges persist, including the arms race between creators of deepfake content and detection systems, the need for privacy- preserving solutions, and the importance of regulatory frameworks and educational initiatives to address the broader societal impact of deep fakes.

Ultimately, deepfake face image detection holds immense promise in safeguarding against the spread of deceptive and harmful content, protecting individual privacy and security, and promoting trust and integrity in digital media and communication. By fostering collaboration across disciplines, investing in research and development, and prioritizing transparency and accountability, we can continue to advance the field of deep face detection and mitigate the risks associated with this emerging technology.

ACKNOWLEDGMENT

With a deep sense of gratitude, we would like to thank all the people who have lit our path with their kind guidance for our Project Selection, Design and Development. We are very grateful to these intellectuals, experts, who did their best to help during our completion of project work.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 3, March 2024
DOI: 10.17148/IJARCCE.2024.13359

It is our proud privilege to express a deep sense of gratitude to, **Prof. P. T. Kadave-** Principal, K. K. Wagh Polytechnic, Nashik for his comments and kind permission to complete this project. We Remain indebted to **Prof. H. M. Gaikwad,** Head of Artificial Intelligence & Machine Learning Department for his timely suggestion and valuable guidance.

The special gratitude goes to our Internal Faculty Guide **Mr. H. M. Gaikwad**, staff members, technical staff members, of Artificial Intelligence & Machine Learning Department for his/her technical, timely, excellent and coercive guidance in completion of this project work. We thank all the class colleagues for their appreciable, encouraging help for completion of our project.

REFERENCES

[1]. Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, Matthias Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images" in arXiv:1901.08971.
[2]. Deepfake detection challenge dataset: https://www.kaggle.com/c/deepfakedetection challenge/data Accessed on 26 March 2020
[3]. Yuezun Li, Xin Yang, Pu Sun, Honggang Qi and Siwei Lyu "Celeb-DF: A Large Scale Challenging Dataset for DeepFake Forensics" in arXiv:1909.12962
[4]. The rise of the deepfake and the threat to democracy: https://www.theguardian.com/technology/nginteractive/2019/jun/22/the-rise-ofthedeepfake and-the-threat-to-democracy
[5]. Yuezun Li, Siwei Lyu, "ExposingDF Videos by Detecting Face Warping Artifacts," in arXiv:1811.00656v3.
[6]. Yuezun Li, Ming-Ching Chang and Siwei Lyu "Exposing AI Created Fake Videos by Detecting Eye Blinking" in arXiv:1806.02877v2.
[7]. Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen "Using capsule networks to detect forged images and videos" in arXiv:1810.11215.
[8]. D. Güera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 2018, pp. 1-6.
[9]. I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 1–8, June 2008. Anchorage, AK
[10]. Umur Aybars Ciftci, ˙Ilke Demir, Lijun Yin "Detection of Synthetic Portrait Videos using Biological Signals" in arXiv:1901.02212v2
[11]. Deepfake Video of Mark Zuckerberg Goes Viral on Eve of House A.I. Hearing: https://fortune.com/2019/06/12/deepfake-mark-zuckerberg/
[12]. https://www.geeksforgeeks.org/software-engineering-cocomo-model/

BIOGRAPHY

Name: Mr. H.M. Gaikwad
Qualification: **B.E. Computer Engineering**

Name: Aryan Kiran Sonawane
Qualification: Diploma, Artificial Intelligence and Machine Learning

Name: Ratnali Anil Pawar
Qualification: Diploma, Artificial Intelligence and Machine Learning

Name: Manavaditya Mangalprasad Rathawa
Qualification: Diploma, Artificial Intelligence and Machine Learning

Name: Uday Hemchandra Talele
Qualification: Diploma, Artificial Intelligence and Machine Learning