Deepfake Video Detection

Prof. Karthiga G[1], Ankit Kumar[2], Ankit Tiwari[3], Rishabh Pandey[4],Priyanshu[5]

[1] *Assistant Professor, CSE, AMCEC, Karnataka, India*
[2] *Student, CSE, AMCEC, Karnataka, India*
[3] *Student, CSE, AMCEC, Karnataka, India*
[4] *Student, CSE, AMCEC, Karnataka, India*
[5] *Student, CSE, AMCEC, Karnataka, India*

ABSTRACT

The growing computation power has made the deep learning algorithms so pow- erful that creating a indistinguishable human synthesized video popularly called as deep fakes have became very simple. Scenarios where these realistic face swapped deep fakes are used to create political distress, fake terrorism events, revenge porn, blackmail peoples are easily envisioned. In this work, we describe a new deep learning-based method that can effectively distinguish AI-generated fake videos from real videos.Our method is capable of automatically detecting the replacement and reenactment deep fakes. We are trying to use Artificial Intelligence(AI) to fight Artificial Intelligence(AI). Our system uses a Res-Next Convolution neural network to extract the frame-level features and these features and further used to train the Long Short Term Memory(LSTM) based Recurrent Neural Network(RNN) to clas- sify whether the video is subject to any kind of manipulation or not, i.e whether the video is deep fake or real video. To emulate the real time scenarios and make the model perform better on real time data, we evaluate our method on large amount of balanced and mixed data-set prepared by mixing the various available data-set like Face-Forensic++[1], Deepfake detection challenge[2], and Celeb-DF[3]. We also show how our system can achieve competitive result using very simple and robust approach.

1. INTRODUCTION

In the world of ever growing Social media platforms, Deepfakes are con- sidered as the major threat of the AI. There are many Scenarios where these realistic face swapped deepfakes are used to create political distress, fake ter- rorism events, revenge porn, blackmail peoples are easily envisioned.Some of the examples are Brad Pitt, Angelina Jolie nude videos.

It becomes very important to spot the difference between the deepfake and pris- tine video. We are using AI to fight AI.Deepfakes are created using tools like FaceApp[11] and Face Swap [12], which using pre-trained neural networks like GAN or Auto encoders for these deepfakes creation. Our method uses a LSTM based artificial neural network to process the sequential temporal analysis of the video frames and pre-trained Res-Next CNN to extract the frame level fea- tures. ResNext Convolution neural network extracts the frame-level features and these features are further used to train the Long Short Term Memory based artificial Recurrent Neural Network to classify the video as Deepfake or real. To emulate the real time scenarios and make the model perform better on real time data, we trained our method with large amount of balanced and combi- nation of various available dataset like FaceForensic++[1], Deepfake detection challenge[2], and Celeb-DF[3].

Further to make the ready to use for the customers, we have developed a front end application where the user the user will upload the video. The video willbe processed by the model and the output will be rendered back to the user with the classification of the video as deepfake or real and confidence of the model.

2. PROBLEM STATEMENT

Convincing manipulations of digital images and videos have been demonstrated for several decades through the use of visual effects, recent advances in deep learn- ing have led to a dramatic increase in the realism of fake content and the accessibilityin which it can be created. These so-called AI-synthesized media (popularly referred to as deep fakes).Creating the Deep Fakes using the Artificially intelligent tools are simple task. But, when it comes to detection of these Deep Fakes, it is major chal- lenge. Already in the history there are many examples where the deepfakes are used as powerful way to create political tension[14], fake terrorism events, revenge porn, blackmail peoples etc.So it becomes very important to detect these deepfake and avoid the percolation of deepfake through social media platforms. We have taken a step forward in detecting the deep fakes using LSTM based artificial Neural network.

3. BACKGROUND WORK

The field of deepfake detection involves developing techniques and algorithms to identify and differentiate between authentic and manipulated media content created using deep learning and artificial intelligence techniques. Here is an overview of the background work involved in a deepfake detection project:

1. Understanding Deepfakes: Researchers and developers need to gain a comprehensive understanding of deepfakes, which are digitally altered media, such as images, videos, or audio, created using deep learning algorithms. This includes studying the various techniques and tools used to generate deepfakes, such as generative adversarial networks (GANs).

2. Data Collection: A significant aspect of deepfake detection involves gathering a large and diverse dataset that consists of both authentic and deepfake content. This dataset is used to train and evaluate the deepfake detection models. It may include examples of manipulated media generated using different techniques, resolutions, and subjects.

3. Feature Extraction: Extracting meaningful features from the media content plays a vital role in distinguishing deepfakes from authentic content. Various visual, audio, and temporal features are analyzed, such as facial expressions, eye movement, lip-syncing accuracy, and anomalies in pixel-level details. These features provide valuable cues for differentiating between real and manipulated content.

4. Machine Learning Algorithms: Researchers employ machine learning algorithms to analyze the extracted features and build predictive models for deepfake detection. Common techniques include supervised learning, unsupervised learning, and semi-supervised learning. These models are trained on the collected dataset, with labeled data indicating whether the content is authentic or manipulated.

5. Deepfake Generation Techniques: In order to effectively detect deepfakes, it is necessary to keep up with the latest advancements in deepfake generation techniques. Researchers actively monitor and analyze new methods employed by creators of deepfakes, which helps them stay one step ahead in developing detection methods.

6. Benchmarking and Evaluation: To assess the performance of deepfake detection models, benchmark datasets are used for testing and evaluation. These datasets are typically made publicly available and contain a variety of deepfake and real content. Metrics such as accuracy, precision, recall, and F1-score are calculated to measure the effectiveness of the detection models.

7. Enhancing Detection Techniques: Deepfake creators constantly evolve their methods to make their content more convincing and harder to detect. Consequently, researchers must continuously refine and enhance their detection techniques to keep pace with the evolving nature of deepfakes. This involves exploring new features, algorithms, and approaches to improve the accuracy and robustness of the detection models.

8. Collaboration and Community Efforts: Deepfake detection is a rapidly evolving field, and

collaboration among researchers, industry experts, and organizations is essential. Sharing knowledge, methodologies, and datasets helps foster advancements in deepfake detection and ensures a collective effort towards mitigating the risks associated with deepfake technology.

Overall, a deepfake detection project involves a multidisciplinary approach, combining expertise in computer vision, machine learning, data analysis, and domain knowledge. Continuous research, innovation, and collaboration are crucial to staying ahead in the ongoing battle against deepfake manipulation.

4. OBJECTIVE

The main objective of our project is to develop an detecting system that recognizes tools and ingredients in a fake videoes using Lstm and restnest algorithms and sorts them based on their frequency of use. Specifically, our objectives include:

- Our project aims at discovering the distorted truth of the deep fakes.
- Our project will reduce the Abuses' and misleading of the common people onthe world wide web.
- Our project will distinguish and classify the video as deepfake or pristine.
- Provide a easy to use system for used to upload the video and distinguishwhether the video is real or fake.

5. LITERATURE SURVEY

Face Warping Artifacts [15] used the approach to detect artifacts by com- paring the generated face areas and their surrounding regions with a dedicated Convolutional Neural Network model. In this work there were two-fold of Face Artifacts.
Their method is based on the observations that current deepfake algorithm can only generate images of limited resolutions, which are then needed to be fur- ther transformed to match the faces to be replaced in the source video. Their method has not considered the temporal analysis of the frames.

Detection by Eye Blinking [16] describes a new method for detecting the deep- fakes by the eye blinking as a crucial parameter leading to classification of the videos as deepfake or pristine. The Long-term Recurrent Convolution 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 can not be the only clue for detection of the deepfakes. There must be certain other parameters must be considered for the detection of deep- fakes like teeth enchantment, wrinkles on faces, wrong placement of eyebrows etc.

Capsule networks to detect forged images and videos [17] 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 [18] (RNN) for deepfake detection used the ap- proach of using RNN for sequential processing of the frames along with Ima- geNet pre-trained model. Their process used the HOHO [19] dataset consisting of just 600 videos.
Their dataset consists small number of videos and same type of videos, which may not perform very well on the real time data. We will be training out model on large number of Realtime data.

Synthetic Portrait Videos using Biological Signals [20] approach extract bio- logical signals from facial regions on pristine and deepfake portrait video pairs. Applied transformations to compute the spatial coherence and temporal consis- tency, capture the signal characteristics in feature vector and photoplethysmog- raphy (PPG) maps, and further train a probabilistic Support Vector Machine

(SVM) and a Convolutional Neural Network (CNN). Then, the average of au- thenticity probabilities is used to classify whether the video is a deepfake or a pristine.

6. METHODOLOGY

- **Solution Requirement**

  We analysed the problem statement and found the feasibility of the solutionof the problem. We read different research paper as mentioned in 3.3. After checking the feasibility of the problem statement. The next step is the data- set gathering and analysis. We analysed the data set in different approach of training like negatively or positively trained i.e training the model with onlyfake or real video's but found that it may lead to addition of extra bias in the model leading to inaccurate predictions. So after doing lot of research we found that the balanced training of the algorithm is the best way to avoid the bias and variance in the algorithm and get a good accuracy.

- **Solution Constraints**

  We analysed the solution in terms of cost,speed of processing,requirements,level of expertise, availability of equipment's.

- **Parameter Identified**

1. Blinking of eyes
2. Teeth enchantment
3. Bigger distance for eyes
4. Moustaches
5. Double edges, eyes, ears, nose
6. Iris segmentation
7. Wrinkles on face
8. Inconsistent head pose
9. Face angle
10. Skin tone
11. Facial Expressions
12. Lighting
13. Different Pose
14. Double chins
15. Hairstyle
16. Higher cheek bones

7. ARCHITECTURE

For making the model efficient for real time prediction. We have gathered the data from different available data-sets like FaceForensic++(FF)[1], Deepfake detection challenge(DFDC)[2], and Celeb-DF[3]. Futher we have mixed the dataset the col- lected datasets and created our own new dataset, to accurate and real time detection on different kind of videos. To avoid the training bias of the model we have consid- ered 50% Real and 50% fake videos.

Deep fake detection challenge (DFDC) dataset [3] consist of certain audio alerted video, as audio deepfake are out of scope for this paper. We preprocessed the DFDC dataset and removed the audio altered videos from the dataset by runninga python script.After preprocessing of the DFDC dataset, we have taken 1500 Real and 1500 Fake videos from the DFDC dataset. 1000 Real and 1000 Fake videos from the FaceForensic++(FF)[1] dataset and 500 Real and 500 Fake videos from the Celeb- DF[3] dataset. Which makes our total dataset consisting 3000 Real, 3000 fake videos and 6000 videos in total. Figure 2 depicts the distribution of the data-sets

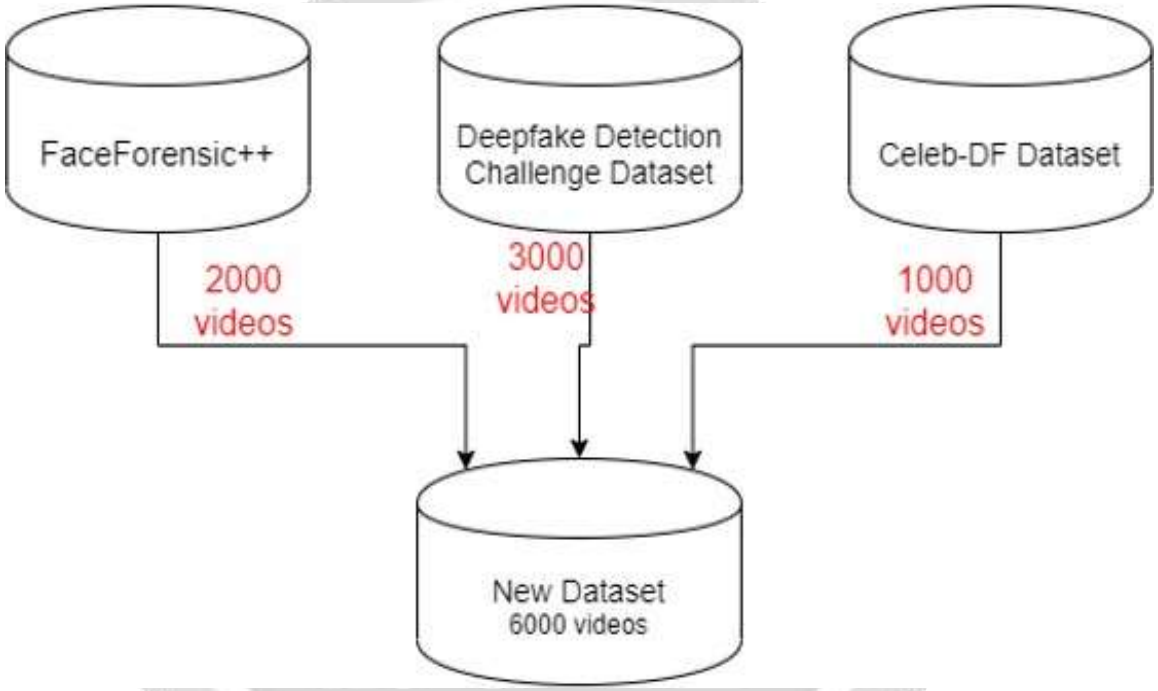
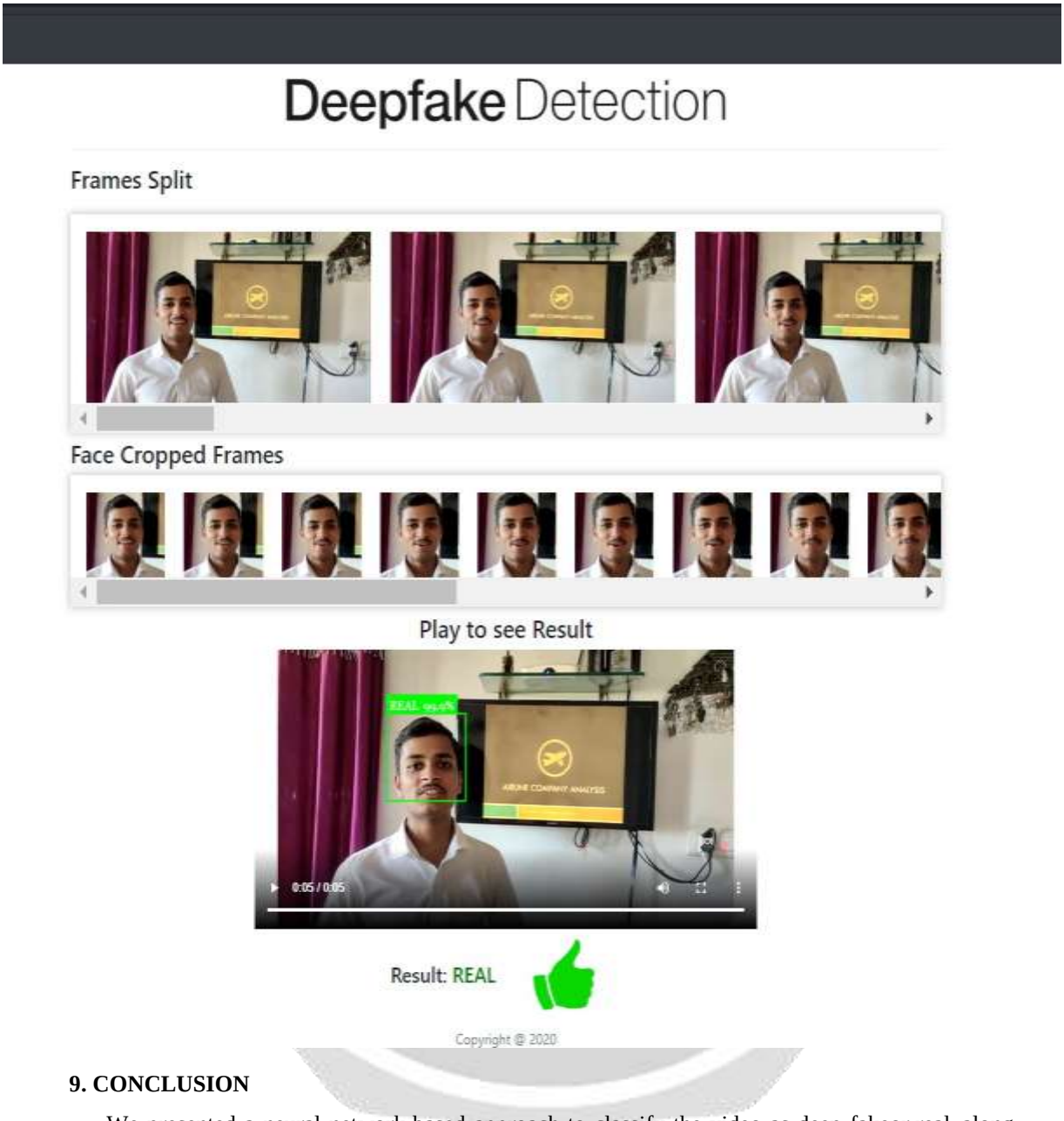


Figure 7.4: Dataset

8. RESULTS



9. CONCLUSION

We presented a neural network-based approach to classify the video as deep fakeor real, along with the confidence of proposed model. Our method is capable of predicting the output by processing 1 second of video (10 frames per second) witha good accuracy. We implemented the model by using pre-trained ResNext CNN model to extract the frame level features and LSTM for temporal sequence process- ing to spot the changes between the t and t-1 frame. Our model can process the video in the frame sequence of 10,20,40,60,80,100.

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/deepfake-detection- 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] Deepfake Video of Mark Zuckerberg Goes Viral on Eve of House A.I. Hearing : https://fortune.com/2019/06/12/deepfake-mark-zuckerberg/ Accessed on 26 March, 2020
[5] 10 deepfake examples that terrified and amused the internet : https://www.creativebloq.com/features/deepfake-examples Accessed on 26 March,2020