Face Corrupt Identification Using LSTM And ResNext

T. Devanshu Kumar[1], P. Ananditha [2], K. Sai Srikari[3], G. Anirvaj[4], V. Geethika Siri[5], G. Sateesh[6]

[1-5]B.TECH STUDENTS
[6]Associate Professor, LIET
[1,2,3,4,5] Computer Science and Information Technology, Lendi Institute of Engineering and Technology, Vizianagaram

---***---

Abstract - People will believe in what they see rather than hidden facts. In recent times as technology is improving and it became very easy to change a particular person in an existing image or video with someone else. So, spreading these types of modified videos causes spamming and peculating wrong information over social media and this will cause the great extent of misleading and threatening to the common people. To overcome these tedious situations, we are building a deep learning model that will identify the genuinity of the video or image in terms of the origin of its source. In our proposed system we are using Long Short-Term Memory and ResNext.

***Key Words*:** ResNext, Convolution neural network, Recurrent Neural Network (RNN), Long Short- Term Memory (LSTM), Computer vision, Deepfake Video Detection, Pre Processing, Deep Learning, Image Processing, Classification.

1.INTRODUCTION

Facial morphing is a technique that digitally blends two or more faces into a new image . It was initially proposed as a method to create a more accurate portrait of a suspect based on the verbal description of several witnesses . However, malicious actors have found ways to use facial morphing for fraudulent activities, such as bypassing security systems and accessing restricted areas or protected data . Face morphing is a type of attack that deceives face recognition systems by merging facial features of two or more individuals into one, creating a fraudulent identity . It is used to bypass facial recognition systems and counterfeit physical IDs, including passports, for illegal activities . This technique is also employed to cross state borders illegally and verify an illegal infiltrator as another legitimate person . Tools for morphing faces, such as Adobe Photoshop and Morph Thing, are easily available . In fact, Germany officially banned "photo morphing in passports" after a German activist obtained a government-issued passport containing his photo morphed with the face of an Italian politician . Detecting face morphing attacks can be challenging since artifacts left by image merging may not be visible to the naked eye . Signs of fraudulent face morphing attack include blurriness around the facial contour/hair, distortion, shadow anomalies, and misplaced facial landmarks . Researchers have proposed various detection methods based on high-frequency features and progressive enhancement learning to overcome these limitations . Despite these efforts, face morphing attacks continue to pose a significant threat for fraudulent activities.

To detect face morphing, researchers have utilized various deep learning models. The dataset developed by Makrushin et al. includes two methods of morphing – complete and splicing. The complete morph method involves the facial geometry of source images and offers 1,326 samples, while the splicing morph method clips out the "face pixels" out of the input image and provides 2,614 samples. Researchers have employed LSTM and ResNext models to detect face morphing using this dataset. LSTM models are capable of retaining information over a long period of time, making them suitable for detecting morphed faces, which often retains features from both source faces. ResNext models are deep convolutional neural networks that can handle complex data and are widely used for image recognition tasks. By utilizing these models, researchers can identify morphed faces with a high degree of accuracy and prevent security breaches caused by face morphing attacks.

2. Literature Survey

The exponential growth of deep fake video and its illegal utilization poses a significant danger to democracy, justice, and legitimacy. As a result, there is an increasing demand for the analysis, detection, and intervention of fake videos. Some of the terms associated with deep fake detection are provided below:

[1] Yuezun Li, Siwei Lyu conducted a study titled "Exposing Deepfake Videos by Detecting Face Warping Artifacts" where they employed a method to identify artifacts by comparing the generated face areas and their surrounding regions using a dedicated Convolutional Neural Network model. Their work focused on two types of face artifacts. Their approach is based on the observation that current Deepfake algorithms can only generate images of limited resolutions, which need to be further adjusted to match the faces in the original video.

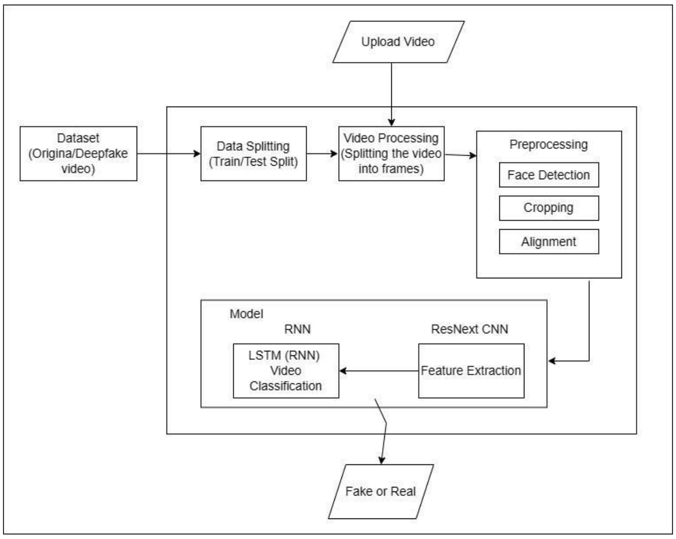
[2] Yuezun Li, Ming-Ching Chang, and Siwei Lyu presented a system called "Exposing AI Created Fake Videos by Detecting Eye Blinking" to uncover fake face videos generated using deep neural network models. They rely on the identification of eye blinking in the videos, which is a physiological signal that is not well

represented in synthesized fake videos. Their system uses the absence of blinking as a detection indicator. However, the discovery of deep fakes requires considering other parameters such as teeth enhancement and facial wrinkles. Our approach takes all these parameters into account.

[3] Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen developed a system that utilizes a capsule network to detect forged and manipulated images and videos in various scenarios, including iteration attack detection and computer-generated video detection. In their methodology, they introduced arbitrary noise during the training phase, which may not be an optimal choice. Although their model performed well on their dataset, it may fail when applied to real-time data due to the presence of noise in the training process. Our proposed methodology aims to be trained on noiseless and real-time datasets.

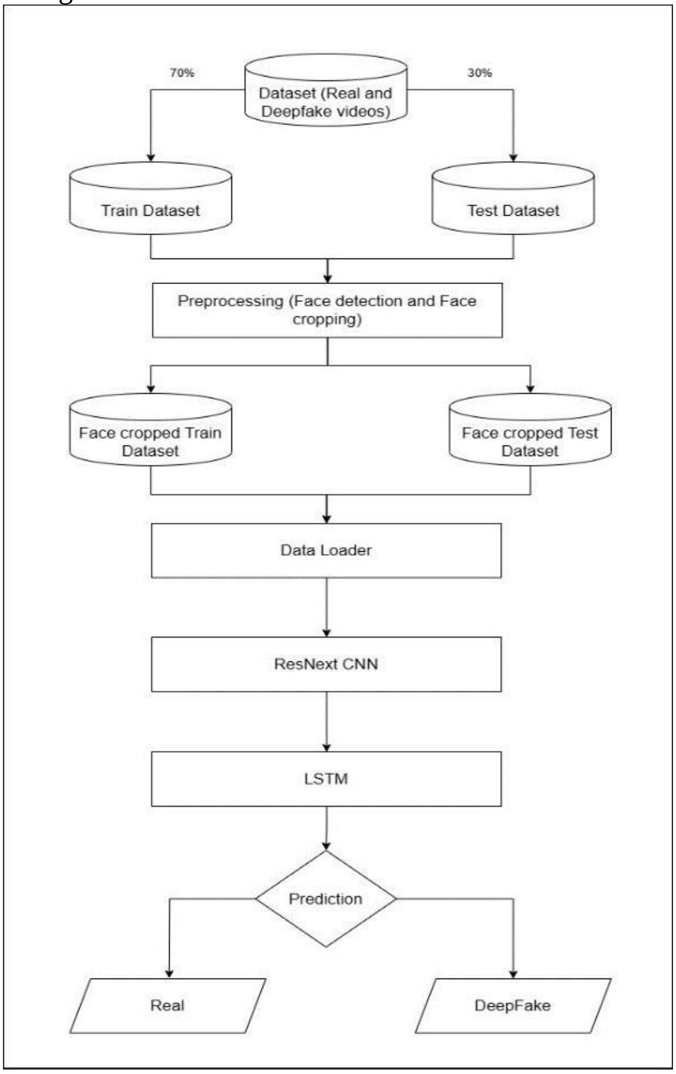
[4] Umur Aybars Ciftci, ˙Ilke Demir, Lijun Yin proposed an approach called "Discovery of Synthetic Portrayal Videos using Biological Signals" which extracts biological signals from facial regions in authentic and fake portrayal video pairs. They apply transformations to analyze the spatial coherence and temporal consistency, capture the characteristic signals in property sets and PPG maps, and train a probabilistic SVM and a CNN. Then, they aggregate the authenticity probabilities to determine whether a video is fake or authentic. Fake Catcher, on the other hand, can detect fake content with high sensitivity regardless of the generator, content, resolution, and quality of the video. However, their findings suffer from a lack of a discriminator to preserve biological signals, and formulating a differentiable loss function that aligns with the proposed signal processing approach is not a straightforward task.

2.1 PROPOSED SYSTEM


Upload Video
Dataset (Origina/Deepfake video)
Data Splitting (Train/Test Split)
Video Processing (Splitting the video into frames)
Preprocessing
Face Detection
Cropping
Alignment
Model
RNN
ResNext CNN
LSTM (RNN) Video Classification
Feature Extraction
Fake or Real

System Architecture-Figure 1

While there exist a plethora of tools that are accessible for the creation of Deepfakes, the number of tools available for their detection is quite limited. In order to impede the dissemination of Deepfakes across the Internet, our method for identification and recognition will prove to be immensely advantageous. We shall provide a web-based platform wherein users can upload videos and categorize them as either counterfeit or genuine. The scope of this project is vast, extending from the development of a web-based platform to the creation of a browser plugin for automatic detection of Deepfakes. Even large-scale applications like Facebook's WhatsApp can integrate this project into their software, enabling the simple pre-detection of Deepfakes before sharing with other users. Assessing its performance and acceptance with regards to security, user-friendliness, accuracy, and reliability constitutes one of the key objectives. An illustration of the basic system architecture of the proposed system is depicted in Figure 1.


70%
Dataset (Real and Deepfake videos)
30%
Train Dataset
Test Dataset
Preprocessing (Face detection and Face cropping)
Face cropped Train Dataset
Face cropped Test Dataset
Data Loader
ResNext CNN
LSTM
Prediction
Real
DeepFake

Training Flow

A.Dataset: We are employing a heterogeneous dataset consisting of an equal number of videos sourced from various datasets such as YouTube, FaceForensics++,and the Deep fake detection challenge dataset. Our recently created dataset includes 50% original videos and 50%

altered deepfake videos. As We Don't want the model to be biased and to recheck it with some of the human touch the dataset has been divided into a test set of 30% and a train set of 70%.

B.Preprocessing: As part of the dataset preprocessing procedure, the video is segmented into frames. Subsequently, face detection is performed, and the frames are cropped to encompass the identified faces. In order to maintain a consistent number of frames, we calculate the mean of the video dataset and create a new processed face-cropped dataset using the frames that comprise the mean. Frames that do not contain any faces are disregarded during the preprocessing stage.Given that processing 300 frames in a 10-second video with a frame rate of 30 frames per second requires substantial computing power,so to ease the process we propose training the model using only the first 150 frames for experimental purposes.

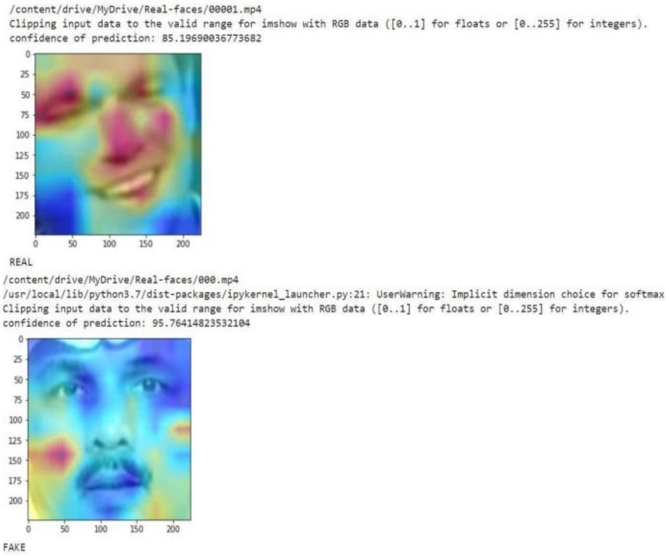
C.Model: The model consists of a single LSTM layer followed by the resnext50_32x4d architecture. The preprocessed face-cropped videos are loaded into the data loader, which partitions them into a train set and a test set. Furthermore, the model receives frames from the edited videos in small batches for both training and testing.

D. ResNext CNN for Feature Extraction: To accurately capture frame-level features, we recommend utilizing the ResNext CNN classifier instead of constructing a classifier from scratch. To extract the features, the network is fine-tuned in such a way that it would merely focus for such parts of the face through which it can easily identify the morphing. The sequential LSTM input is then derived from the 2048-dimensional feature vectors obtained after the last pooling layers.

E. LSTM for Sequence Processing: We propose a 2-node neural network that takes a sequence of ResNext CNN feature vectors of input frames as input, along with the probability of the sequence being a deepfake video or an unaltered video. To address this, we suggest employing a 2048 LSTM unit with a 0.4 dropout likelihood to achieve our objective. By comparing the frame at second 't' with the frame at second 't-n', LSTM is utilized to sequentially process the frames and perform temporal analysis on the video, where 'n' represents the number of frames before 't'. it works as if it remembers a short story

F. Predict: The trained model is capable of predicting new videos. Additionally, a fresh video undergoes preprocessing to align with the format of the trained model. The video is segmented into frames, followed by face cropping, and the cropped frames are promptly fed into the trained model for detection, rather than being stored locally.

3. RESULTS

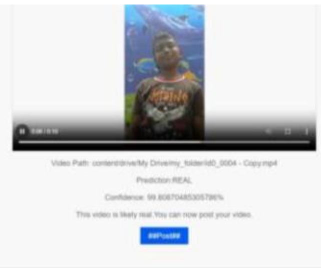


```
/content/drive/MyDrive/Real-faces/00001.mp4
Clipping input data to the valid range for imshow with RGB data ([0..1] for floats or [0..255] for integers).
confidence of prediction: 85.19690036773682
```

```
REAL
/content/drive/MyDrive/Real-faces/000.mp4
/usr/local/lib/python3.7/dist-packages/ipykernel_launcher.py:21: UserWarning: Implicit dimension choice for softmax
Clipping input data to the valid range for imshow with RGB data ([0..1] for floats or [0..255] for integers).
confidence of prediction: 95.76414823532104
```

```
FAKE
```

The model's output will include the model's confidence level and a determination of whether the video is authentic or a deep fake.



Video Path: content/drive/My Drive/my_folder/id0_0004 - Copy.mp4

Prediction:REAL

Confidence: 99.80870485305796%

This video is likely real.You can now post your video.

4.Conclusions

We provided a neural network-based method for determining if a video is a deep fake or the real, along with the model's level of confidence. The deep fakes produced by GANs with the aid of Autoencoders serve as an inspiration for the suggested strategy. Our approach uses ResNext CNN for frame level detection and RNN and LSTM for video classification. Based on the factors stated in the study, the suggested method can determine if a video is a deep fake or real. We think it will deliver real-time data with extremely high accuracy We presented a neural network-based approach to classify the video as deep fake or real, along with the confidence of proposed model. The proposed method is inspired by the way the deep fakes are created by the GANs with the help of Autoencoders, And we are kind of just reversing the approach. Our method does the frame level detection using ResNext CNN and video classification using RNN along with LSTM. The proposed method is capable of detecting the video as a deep fake or real based on the listed parameters in paper. We believe that, it will provide a very high accuracy on real time data We presented a LSTM based approach to detect the video as deep fake or real, by processing 1 sec of video with good accuracy.

REFERENCES

1.Yuezun Li, Siwei Lyu, "ExposingDF Videos by Detecting Face Warping Artifacts," in arXiv:1811.00656v3.
2.Yuezun Li, Ming-Ching Chang and Siwei Lyu "Exposing AI Created Fake Videos by Detecting Eye Blinking" in arxiv.
3.Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen "Using capsule networks to detect forged images and videos".
4. Hyeongwoo Kim, Pablo Garrido, Ayush Tewari and Weipeng Xu "Deep Video Portraits" in arXiv:1901.02212v2.
5.Umur Aybars Ciftci, ˙Ilke Demir, Lijun Yin "Detection of Synthetic Portrait Videos using Biological Signals" in arXiv:1901.02212v2.