DEEPFAKE DETECTION USING DEEP LEARNING

Guide : Dr. Sairamesh L

Name : Gurunathan M

Roll No : 2019202015

MCA – 3rd year(Regular)

Abstract :

The increasing sophistication of smartphone cameras and the availability of good internet connection all over the world has increased the ever-growing reach of social media and media sharing portals have made the creation and transmission of digital videos more easy than ever before. The growing computational power technology, "DeepFake" are produced by deep generative adversarial models that can manipulate video. Spreading of the DF over the social media platforms have become very common leading to spamming and peculating wrong information over the platform. These types of the DF will be terrible, and lead to threating, misleading of common people. So our method detects such artifacts by comparing the generated face areas and their surrounding regions by splitting the video into frames and extracting the features with a ResNext Convolutional Neural Network (CNN) and using Long Short Term Memory(LSTM) capture the temporal inconsistencies between frames introduced by GAN during the reconstruction of the DF. By these way, we can predict the video is DeepFake or Real.

Introduction:

Deep learning is a branch of machine learning which is completely based on artificial neural network, as neural network is going to mimic the human brain so deep learning is also a kind of mimic of human brain. In deep learning, we don’t need to explicitly program everything. Machine learning allows a system to learn and improve from experience

automatically. Deep learning is an application of machine learning that uses complex algorithms and deep neural nets to train a model.

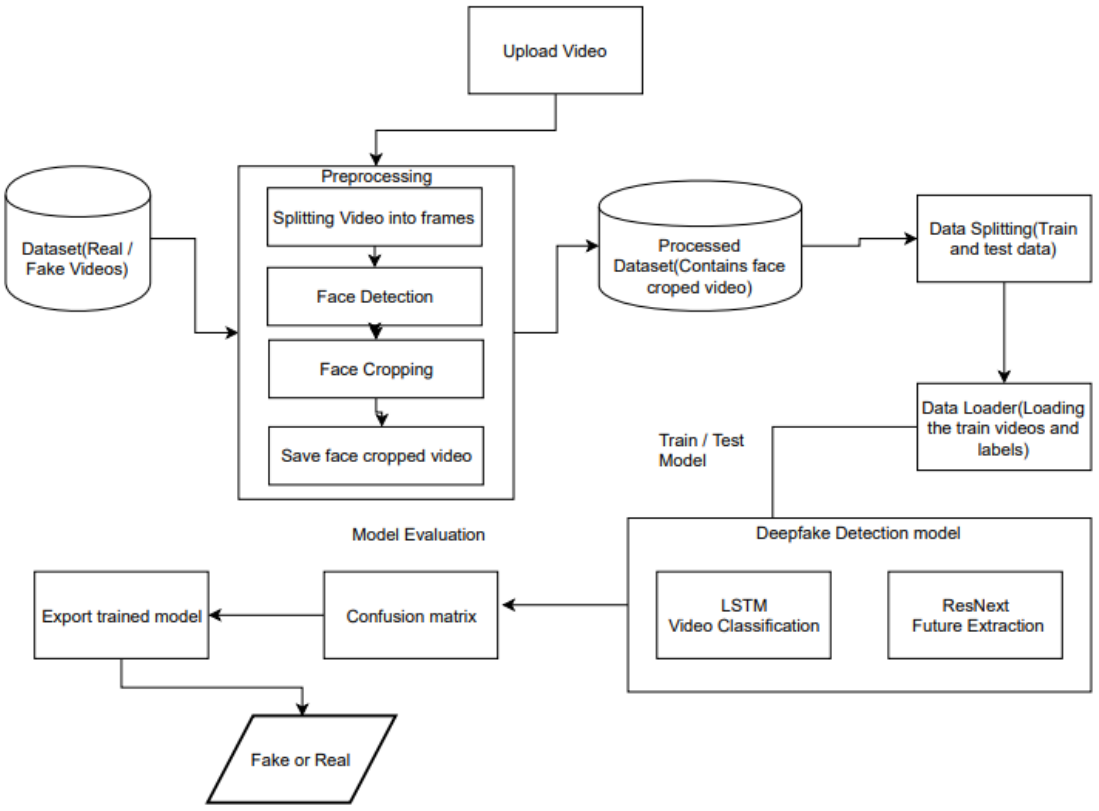
Problem Statement:

The growing computational power technology, "DeepFake" are produced by deep generative adversarial models that can manipulate video. Spreading of the DF over the social media platforms have become very common leading to spamming and peculating wrong information over the platform. These types of the DF will be terrible, and lead to threating, misleading of common people.

Objective:

To design and develop a system to classify whether the video is fake or real by using deep learning algorithms.

Architecture Diagram:



Architecture Explanation:

Initially if we upload a video, in preprocessing it will split the video into frames. Followed by the face detection and cropping the frame with detected face. To maintain the uniformity in the number of frames the mean of the dataset video is calculated and the new processed face cropped dataset is created containing the frames equal to the mean. The Data Loader loads the preprocessed face cropped videos and split the videos into train and test set. Further the frames from the processed videos are passed to the model for training and testing in mini batches. We are proposing to use the ResNext CNN classifier for extracting the features and accurately detecting the frame level features. Following, we will be fine-tuning the network by adding extra required layers and selecting a proper learning rate to properly converge the gradient descent of the model. LSTM for Sequence Processing Let us assume a sequence of ResNext CNN feature vectors of input frames as input and a 2-node neural network with the probabilities of the sequence being part of a deep fake video or an untampered video. LSTM is used to process the frames in a sequential manner so that the temporal analysis of the video can be made, by comparing the frame at ‘t’ second with the frame of ‘t-n’ seconds. Where n can be any number of frames before t.

List of Modules:

1. **Preprocessing**

   Our newly prepared dataset contains 50% of the original video and 50% of the manipulated deepfake videos. The dataset is split into 70% train and 30% test set. Dataset preprocessing includes the splitting the video into frames. Followed by the face detection and cropping the frame with detected face. To maintain the uniformity in the number of frames the mean of the dataset video is calculated and the new processed face cropped dataset is created containing the frames equal to the mean.

2. **Future Extraction**

   We are proposing to use the ResNext CNN classifier for extracting the features and accurately detecting the frame level features. Following, we will be fine-tuning the network by adding extra required layers and selecting a proper learning rate to properly converge the

gradient descent of the model. The 2048-dimensional feature vectors after the last pooling layers are then used as the sequential LSTM input.

3. Sequence Processing:

LSTM for Sequence Processing Let us assume a sequence of ResNext CNN feature vectors of input frames as input and a 2-node neural network with the probabilities of the sequence being part of a deep fake video or an untampered video. The key challenge that we need to address is the de- sign of a model to recursively process a sequence in a meaningful manner. For this problem, we are proposing to the use of a 2048 LSTM unit with 0.4 chance of dropout, which is capable to do achieve our objective. LSTM is used to process the frames in a sequential manner so that the temporal analysis of the video can be made, by comparing the frame at 't' second with the frame of 't-n' seconds. Where n can be any number of frames before t.

4. Prediction:

A new video is passed to the trained model for prediction. A new video is also preprocessed to bring in the format of the trained model. The video is split into frames followed by face cropping and instead of storing the video into local storage the cropped frames are directly passed to the trained model for detection.

References:

[1] Yuezun Li, Ming-Ching Chang and Siwei Lyu "Exposing AI Created Fake

Videos by Detecting Eye Blinking" in arxiv.

[2] Abhijit Jadhav, Abhishek Patange, Jay Patel, Hitendra Patil, Manjushri Mahajan

"DeepFake Videos using neural networks"in IEEE,2020.

[3] R. Raghavendra, Kiran B. Raja, Sushma Venkatesh, and Christoph Busch, "Transferable deep-CNN features for detecting digital and print-scanned morphed face images," in CVPRW. IEEE, 2017.

[4] https://www.kaggle.com/c/deepfake-detection-challenge/data