Meso Network for DeepFake Detection

CS3011 Introduction to Artificial Intelligence Project

Team Members:

Harshal Gadhe – 2019068

Kuldeep Singh Gahlot – 2019083

Under Supervision of:

Prof. Kusum Kumar Bharti

1. Introduction

Over the past few years, huge steps forward in the field of automatic video editing techniques have been made. According to several reports, almost two billion pictures are uploaded every day on the internet. This tremendous use of digital images has been followed by a rise of techniques to alter image contents, using editing software like Photoshop for instance. In particular, great interest has been shown towards methods for facial manipulation, face swapping, etc. thus becoming a great public concern recently. On the other hand, these technological advancements open the door to new artistic possibilities (e.g., movie making, visual effects, visual arts, etc.).

Traditional fake detection methods in media forensics have been commonly based on: i) in-camera fingerprints, the analysis of the intrinsic fingerprints introduced by the camera device, both hardware and software, such as the optical lens, colour filter array and interpolation, and compression, among others, and ii) out camera fingerprints, the analysis of the external fingerprints introduced by editing software, such as copy-paste or copy-move different elements of the image, reduce the frame rate in a video, etc. However, most of the features considered in traditional fake detection methods are highly dependent on the specific training scenario, being therefore not robust against unseen conditions. Recently, various researches have been carried out for building models to detect fakes using deep learning approaches.

This paper addresses the problem of detecting the forged videos created using the DeepFake technique.

1.1. DeepFake

DeepFake is a technique that aims to replace the face of a targeted person with the face of someone else in a video. It first appeared in autumn 2017 as a script used to generate face-swapped adult content.

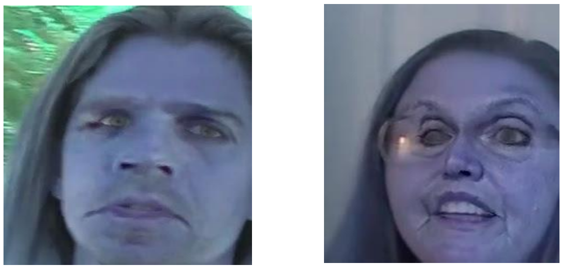

Figure 1. Some of the images extracted from the video forged using DeepFake technique. One can easily noticed that the images are forged but in the video there are many such images stack together to create a video and thus it becomes difficult to notice unless one plays the video frame by frame.

2. Proposed Method

We propose to detect forged videos of faces by placing our method at a mesoscopic level of analysis. As proposed in the research paper [1] MesoNet: a Compact Facial Video Forgery Detection Network where Meso-4 network as well MesoInception-4 was used for prediction, the proposed model was able to perform well either on videos generated using DeepFake or Face2Face method. It was found that these two problems can not be efficiently solved with a unique network.

While implementing we used the DeepFake detection challenge dataset and later on processing was done to get the desired input. The pre-processed data can be found at Dataset. In [1] while pre-processing specific frames at a defined gap were extracted from the video and fed to the deep neural network. During the prediction of the video certain number of frames from a specific interval were extracted and then were fed to the trained model one by one and their output was recorded. Then the average of the output was used to predict whether a video is fake or real.

Each class in the dataset had around 800 images, due to the unavailability of the dataset we weren’t able to train our model properly, and hence our model was barely able to reach accuracy close to 70%. During the training image augmentation technique was used to increase the size of dataset by adding techniques like rotation, rescaling ,etc.

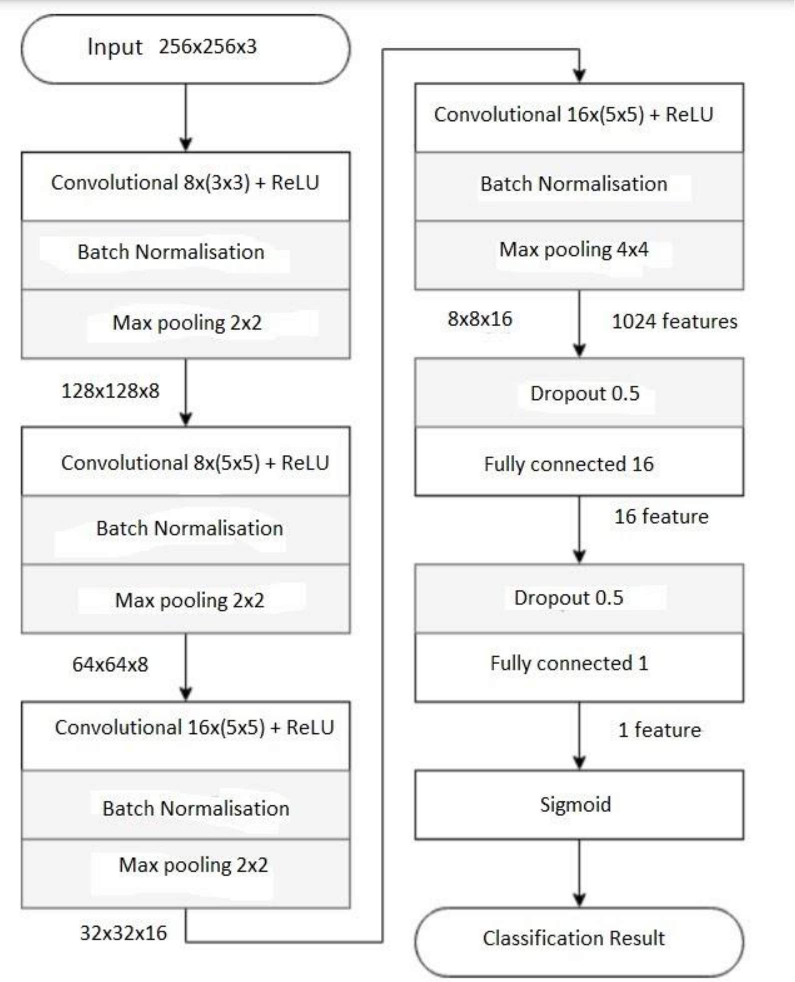

Figure 2. Meso-4 model architecture which was used to train the model

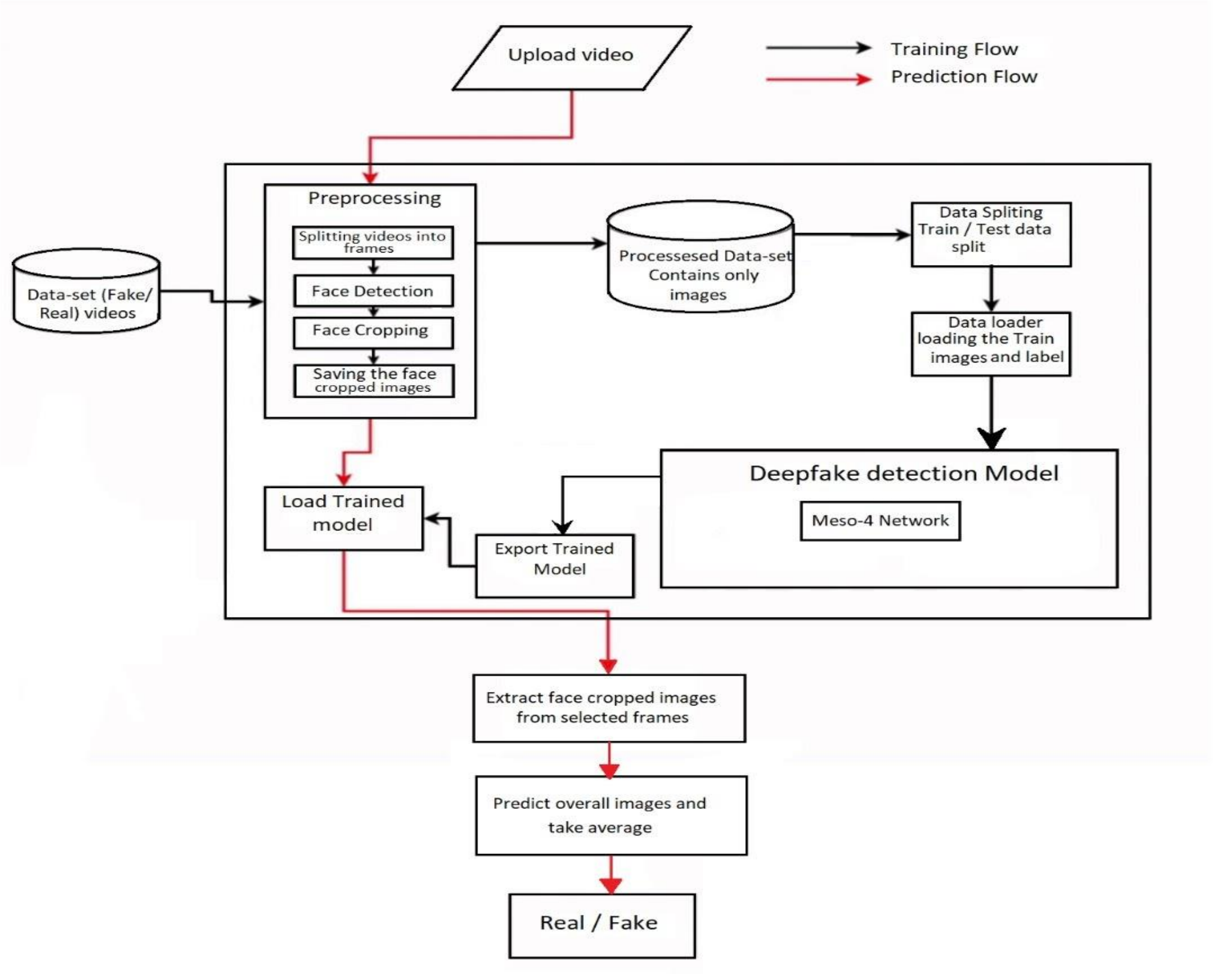
Upload video
Training Flow
Prediction Flow
Data-set (Fake/ Real) videos
Preprocessing
Splitting videos into frames
Face Detection
Face Cropping
Saving the face cropped images
Processesed Data-set Contains only images
Data Spliting Train / Test data split
Data loader loading the Train images and label
Deepfake detection Model
Meso-4 Network
Load Trained model
Export Trained Model
Extract face cropped images from selected frames
Predict overall images and take average
Real / Fake

Figure 3. Flowchart describing the flow of training and prediction pipeline

During prediction of unseen data we will pass the video to the pre-processing function where we have predefined the frames which we will extract from the video for the prediction. The frames to be selected are taken in order as 5,10,15….100 from the 20 frames extracted we will only keep 11 of them which have faces in them and using mtcnn face recognition package we will cropped the area containing the faces. Later these 11 frames will be pass to the model for prediction and the results of each frame is recorded in the array and then the average is taken for each class. The label is then predicted as the class having the maximum count.

3. Results

As you can see from the model that the accuracy of our model is increasing but due to insufficiency of dataset our model start to overfit after 6 epochs which can be seen from the graph.

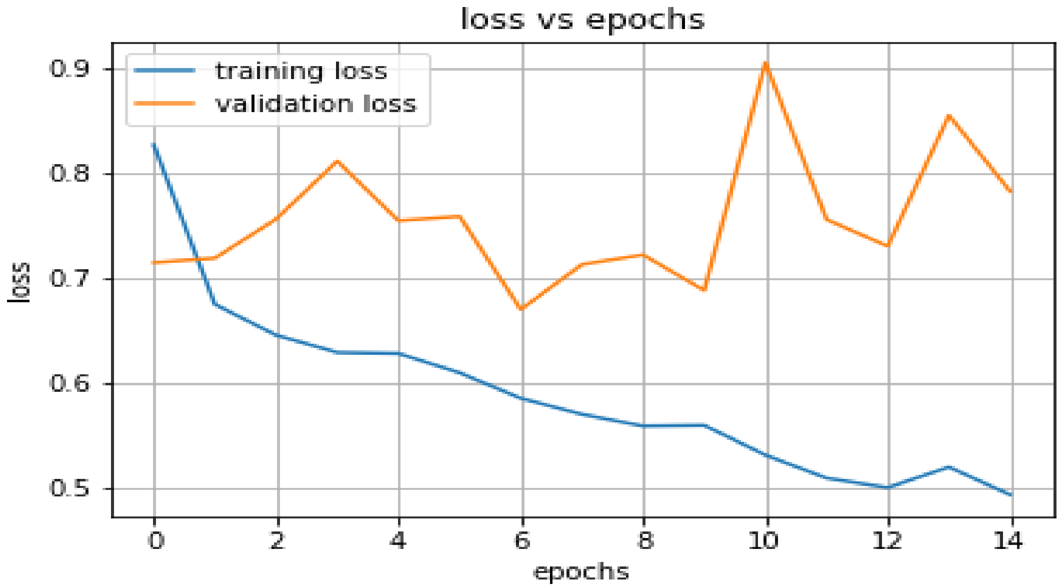

Figure 4. Plot of loss vs epochs

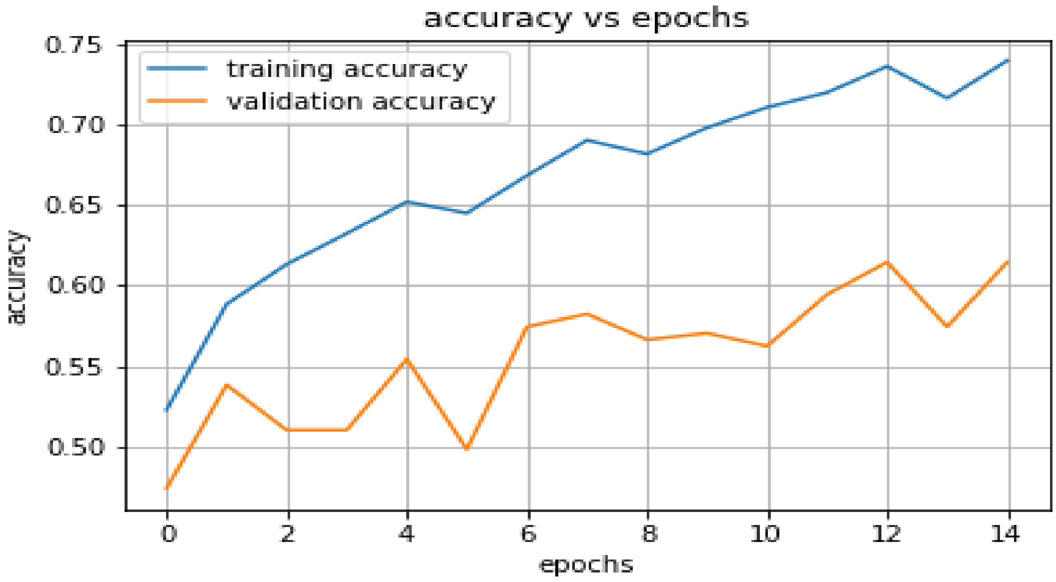

Figure 5. Plot of accuracy vs epochs

4. Conclusion

During training overfitting was seen after certain epochs which can be avoided by using dropout layer on the cost of training accuracy. Model proposed in these paper uses images to predict the output by selection of the frames but getting the right frame could be hard or rather impossible. In further scope we propose use of image sequencing instead of images.

5. References :

1. Deep Learning for DeepFakes Creation and Detection: A Survey Thanh Thi Nguyen, Quoc Viet Hung Nguyen, Cuong M. Nguyen, Dung Nguyen, Duc Thanh Nguyen, Saeid Nahavandi, Fellow, IEEE(April 2021)

2. DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection, International Research Journal of Engineering and Technology (2020)

3. MesoNet: a Compact Facial Video Forgery Detection Network (2018)

4. Video Face Manipulation Detection Through Ensemble of CNNs, International Research Journal of Engineering and Technology (April 2020)

5. Unmasking DeepFakes with simple Features (March 2020)

6. Reverse Engineering of Generative Models: Inferring Model Hyperparameters from Generated Images (June 2021)