Introduction

In recent months, AI-synthesized face swapping videos referred to as deepfake have become an emerging problem. False video is becoming more and more difficult to distinguish, which brings a series of challenges to social security.

The continuous advancement of video tampering technology and the improvement of video quality have also brought great challenges to deepfake detection.

By this project, we will try to detect if a video is deepfake or not using Deep Learning Techniques.

Methodology

- Converting the input video to frames.
- Locating face within the frames generated.
- Data exploration and clustering of images using PCA, T-SNE and DBSCAN for better train test split so that the obtained model learns to resist adversarial examples and optimized input for training.
- Performing inter-frame analysis on video to detect deepfake directly (in case when only a part of video is morphed).
- Using MTCNN and Inception Resnet models for feature extraction and to make classifications.
- Choosing the best model on the basis of accuracy scores obtained and making predictions and classifying videos as real/fake.

Inter Frame Analysis

- Extract frames from video using cv module.
- Locating faces in frames using face recognition module.
- Drawing facial landmarks using face recognitions’s facial landmarks functions.
- Frame by Frame detection of facial landmarks.
- Comparing the adjacent frames to detect morphing.

Test Train Split

- From the frames in each cluster, videos corresponding to those frames are mapped and clustered together.
- From each cluster 70% of videos goes to train set while the other 30% to the test set.

Dataset

Celeb-DF dataset includes 70 original videos collected from YouTube with subjects of different ages, ethic groups and genders, and 330 corresponding DeepFake videos

Screenshot from original video vs Screenshot from deepfaked video

Structure of dataset:

Celeb-DF

Celeb-real # 45 Celebrity videos downloaded from YouTube

YouTube-real # 25 Additional videos downloaded from YouTube

Celeb-synthesis 330 Synthesized videos from Celeb-real (260 from celeb-real) and Youtube real (70 from youtube real)

Work Done

1. Clustering of frames using following techniques:
    a. PCA
    b. t-DSNE
    c. DBSCAN

2. Inter-frame analysis to detect morphing in certain portion of videos
    a. MSE
    b. SSIM

3. Classify video as real or fake

Novelty

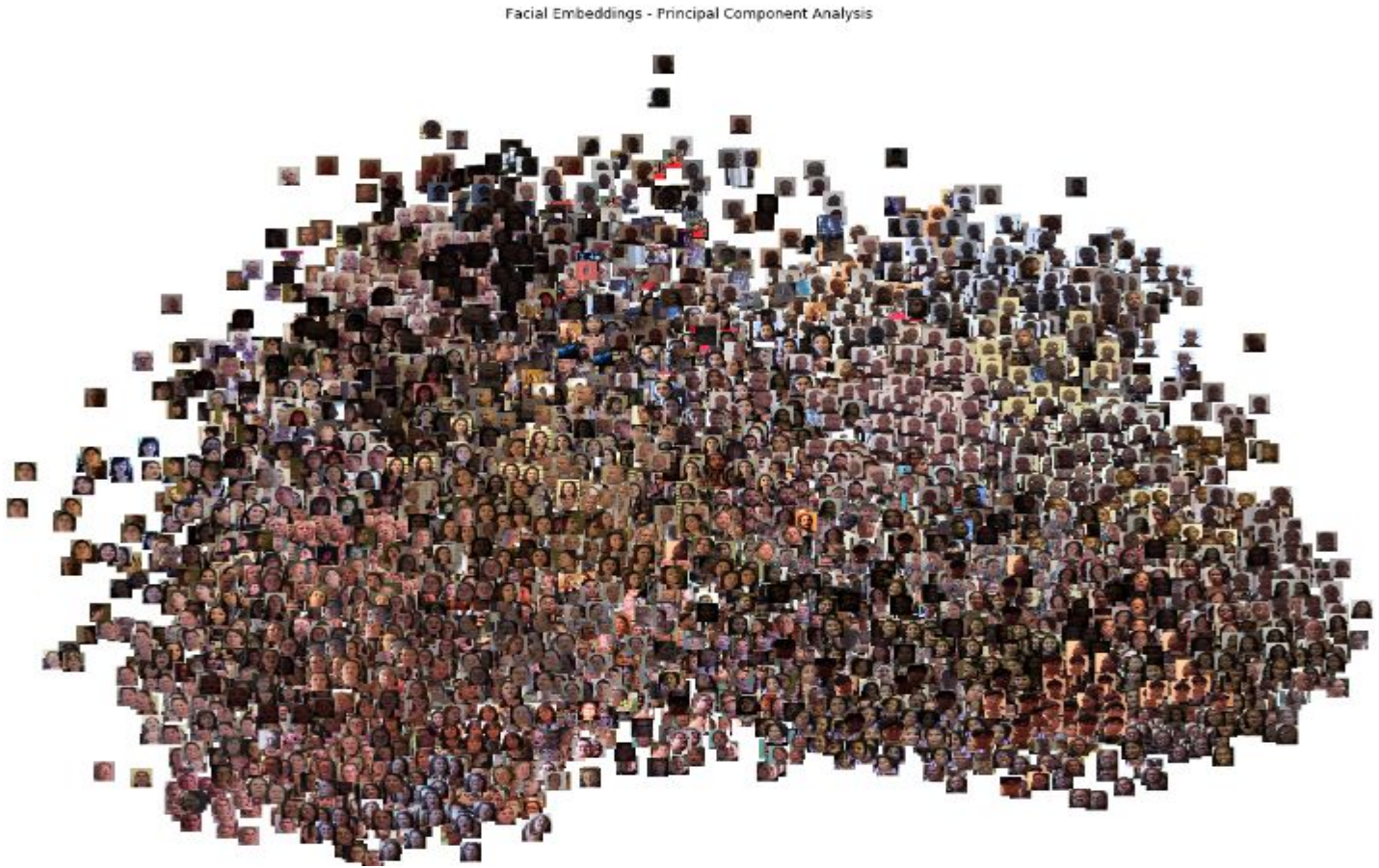
- In order to improve the train-test split strategy, we are looking forward to cluster the images following which the train and test data can be picked from each of the resulting clusters formed.
- In the base paper, if the video clip has only certain areas of deepfaking involved , the video is not classified as fake. So we look to overcome this drawback by interframe change analysis.
- The base paper makes use of RNN in addition to which we look to explore MTCNN and Inception Resnet models to make classifications.

Embedding images

Using a pre-trained Facenet model to embed images into 512 dimensional embeddings.

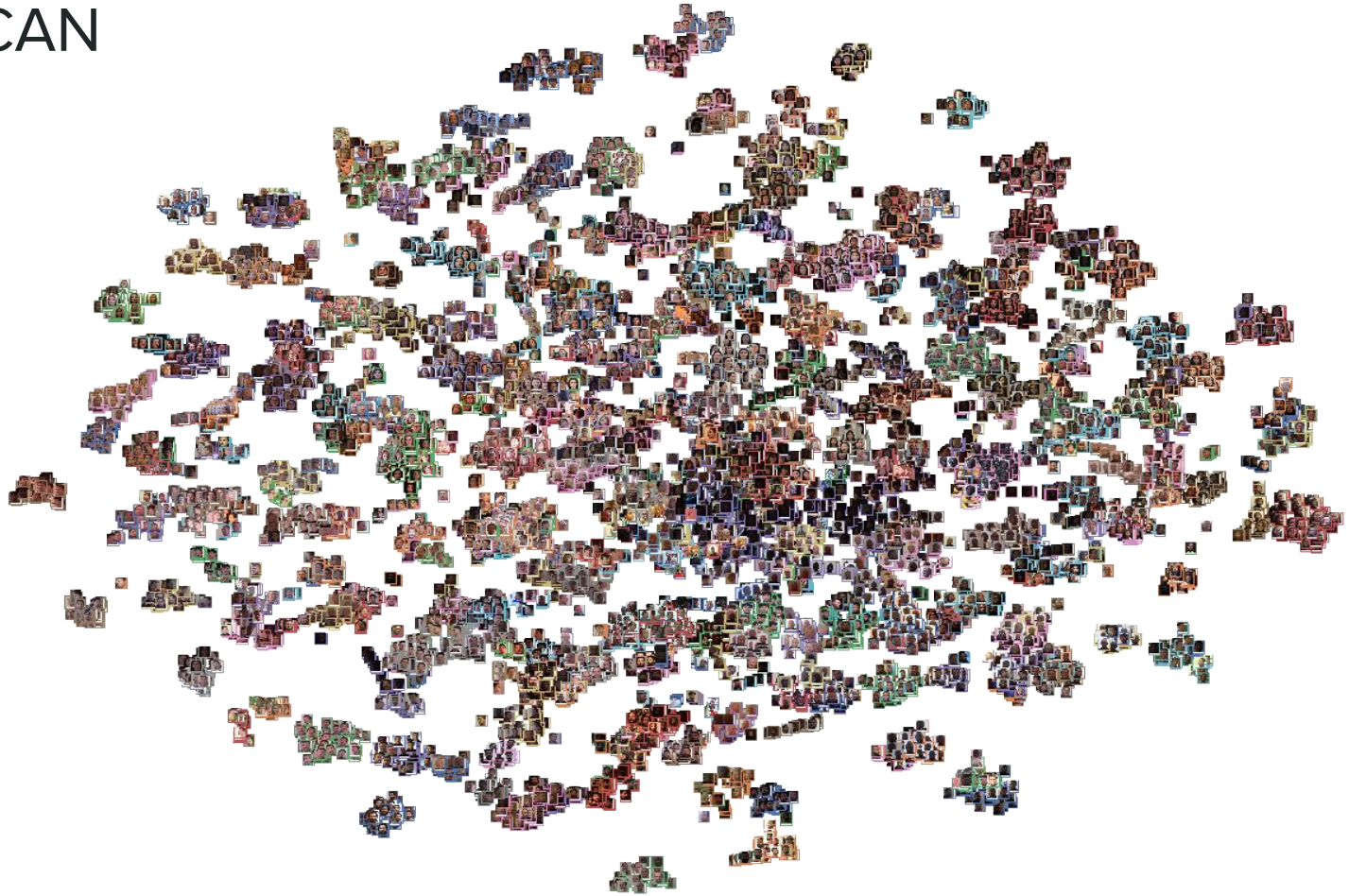
	video	face	chunk	embedding
0	jwhdfrrrsf.mp4	/kaggle/working/faces/jwhdfrrrsf.jpg	24	[-0.05965087562799454, 0.028812751173973083, -...
1	xoxnjeqvbx.mp4	/kaggle/working/faces/xoxnjeqvbx.jpg	24	[-0.06630755960941315, 0.08379176259040833, -0...
2	lowluqbxcz.mp4	/kaggle/working/faces/lowluqbxcz.jpg	24	[-0.0553269200026989, 0.06995047628879547, -0....
3	tyzozcjoxq.mp4	/kaggle/working/faces/tyzozcjoxq.jpg	24	[-0.10966362804174423, -0.004628702532500029, ...
4	odwhmezcnz.mp4	/kaggle/working/faces/odwhmezcnz.jpg	24	[-0.0967181995511055, 0.027173113077878952, -0...

PCA

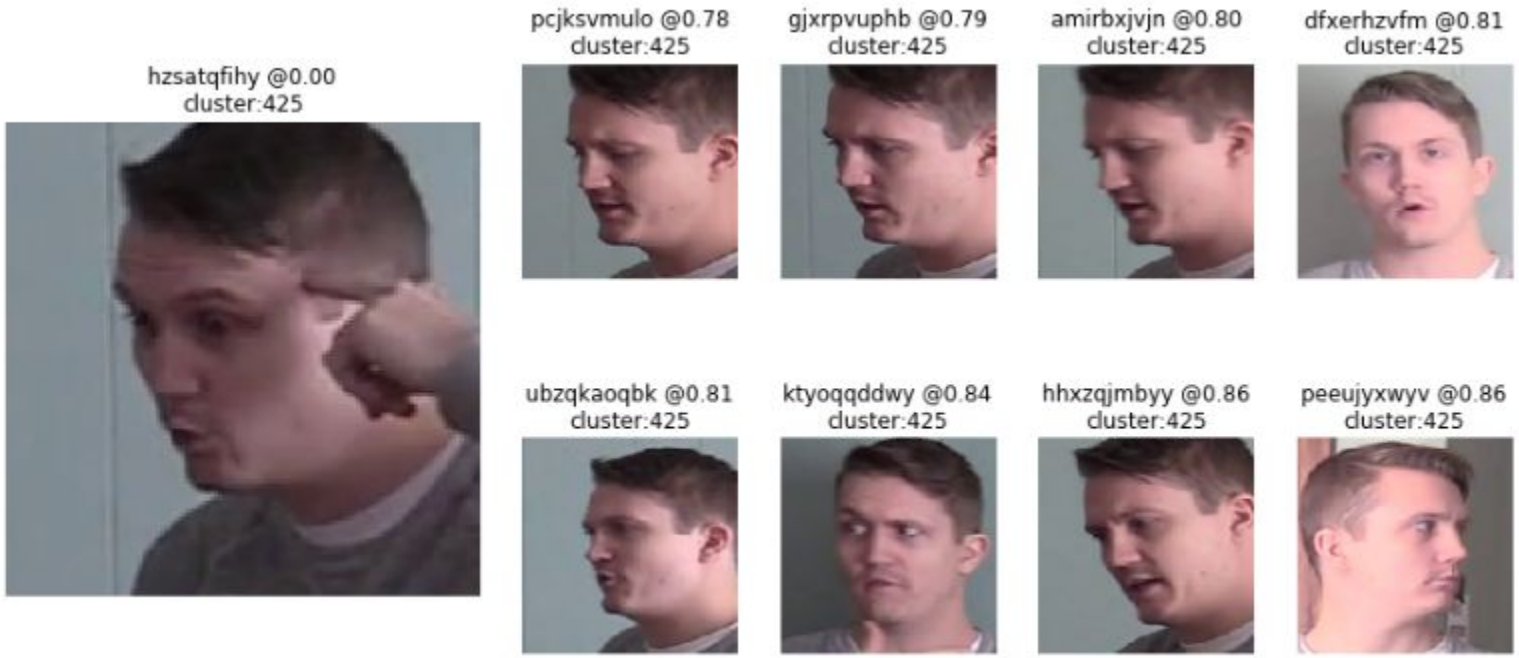


t-DSNE

DBSCAN



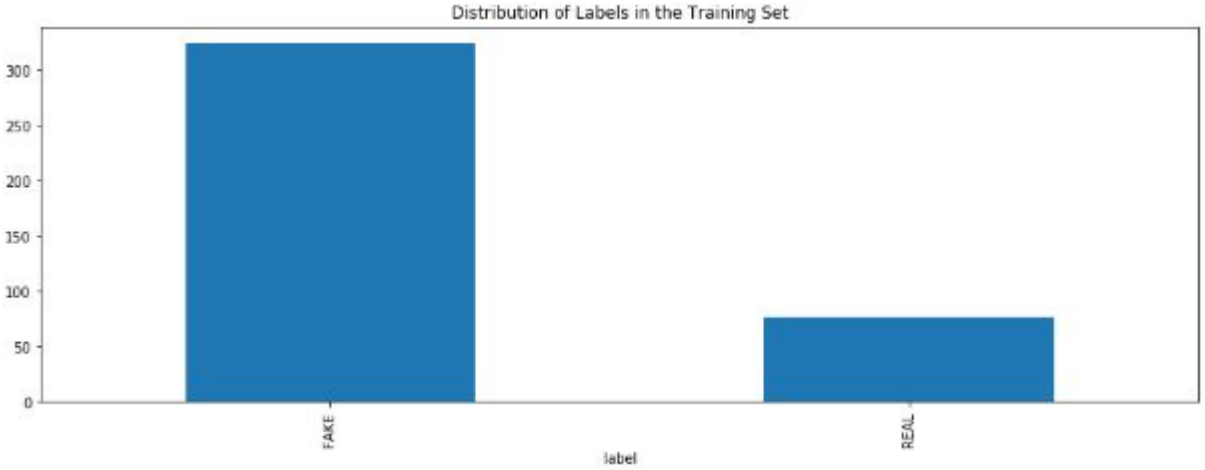
CLUSTER ANALYSIS



CLUSTERS

	video	face	chunk	embedding	cluster
0	jwhdfrrrsf.mp4	/kaggle/working/faces/jwhdfrrrsf.jpg	24	[-0.05965087562799454, 0.028812751173973083, -...	0
1	xoxnjeqvbx.mp4	/kaggle/working/faces/xoxnjeqvbx.jpg	24	[-0.06630755960941315, 0.08379176259040833, -0...	1
2	lowluqbxcz.mp4	/kaggle/working/faces/lowluqbxcz.jpg	24	[-0.0553269200026989, 0.06995047628879547, -0....	1
3	tyzozcjoxq.mp4	/kaggle/working/faces/tyzozcjoxq.jpg	24	[-0.10966362804174423, -0.004628702532500029, ...	2
4	odwhmezcnz.mp4	/kaggle/working/faces/odwhmezcnz.jpg	24	[-0.0967181995511055, 0.027173113077878952, -0...	2
...	...	...	...	...	...
18707	jygbqsigcr.mp4	/kaggle/working/faces/jygbqsigcr.jpg	45	[-0.05358447507023811, -0.040880199521780014, ...	771
18708	kxarjtvlly.mp4	/kaggle/working/faces/kxarjtvlly.jpg	45	[-0.05434044823050499, -0.002579746302217245, ...	771
18709	imsacxixmv.mp4	/kaggle/working/faces/imsacxixmv.jpg	45	[-0.0748559758067131, -0.05435269698500633, -0...	771
18710	bnriwtwfof.mp4	/kaggle/working/faces/bnriwtwfof.jpg	45	[-0.035284534096717834, -0.0566943995654583, -...	771
18711	vwigtbeivy.mp4	/kaggle/working/faces/vwigtbeivy.jpg	45	[-0.0642971470952034, -0.014180679805576801, -...	771

Dataset


Distribution of Labels in the Training Set
300
250
200
150
100
50
0
FAKE
REAL
label

Predictions