Deepfakes Detection

Hala **Zayzafoun**, Tamás **Bukits**, Bolutife **Atoki**

Image Processing AND Computer Vision

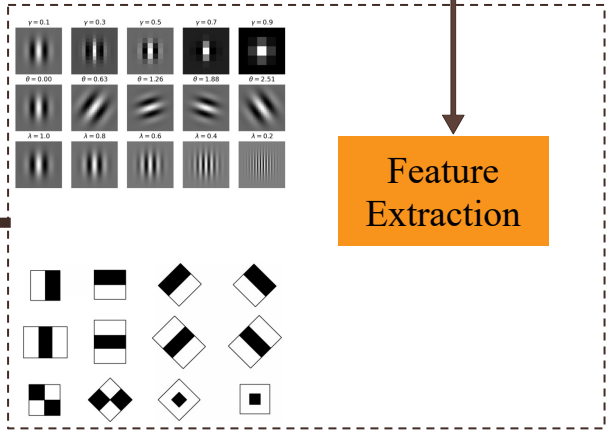
UAM Universidad Autónoma de Madrid

Pipeline

Face Detection

Processing

Reszing to 60 x 60



Feature Extraction

Model and training

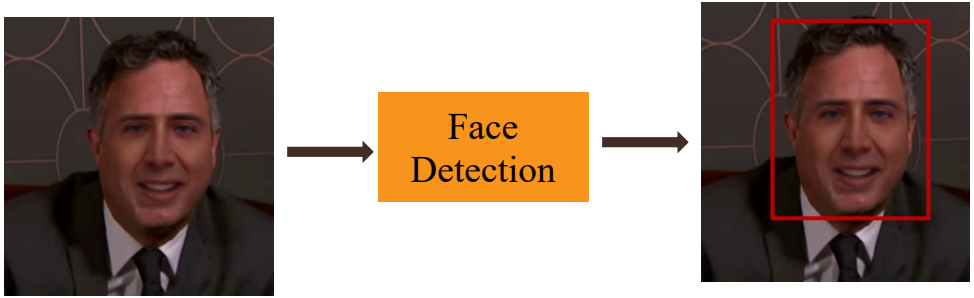
Predection

Face detection

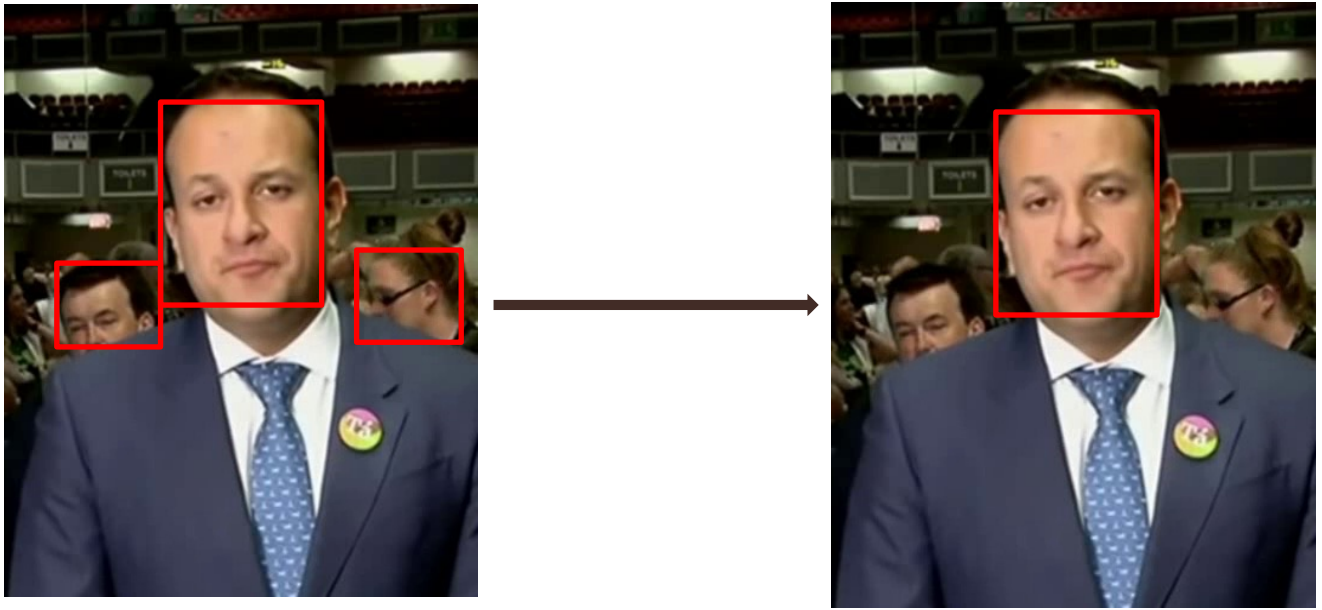
The face has several features that can be identified, like eyes, mouth, nose via Landmark Detection

Robust face detection was implemented with two Face Detectors:

- Landmark Detection
- Dlib's Face recognition



Face detection



In case multiple faces we detected the closest one

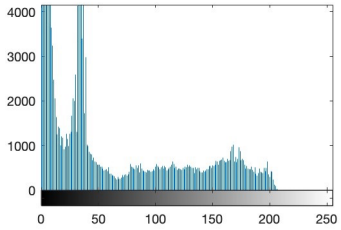
Processing

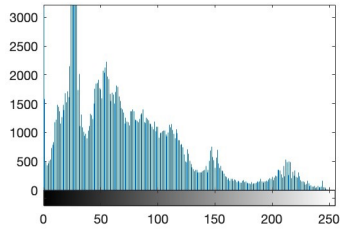
To pass faces to CNN, uniform face sizes were required.

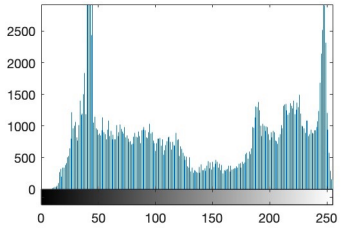
Batch resizing of face images to 60 x 60 was performed.

Initial Approaches

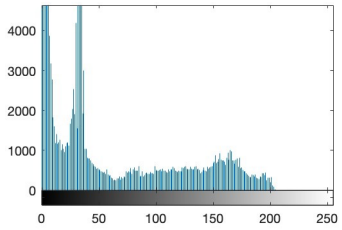
Database Analysis via Histogram

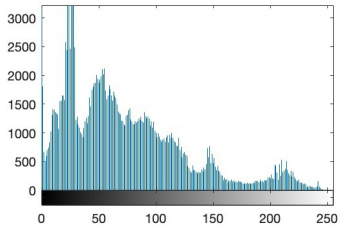


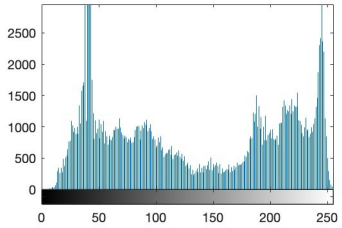




Histogram of Real Image





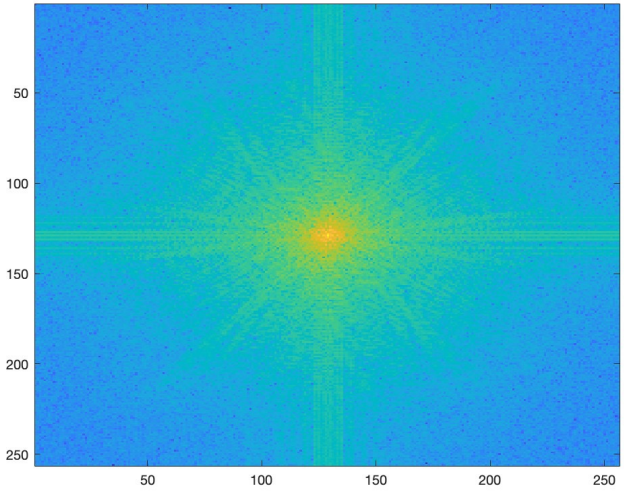


Histogram of Fake Image

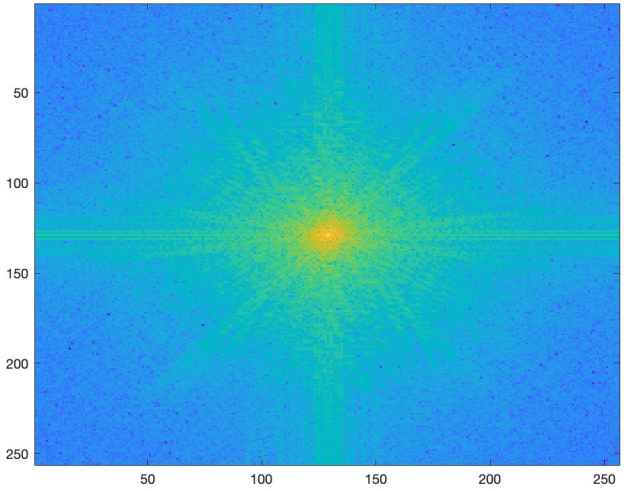
No differences between Real and deepfakes in all three colour channels!

Initial Approaches

Database Analysis via Frequency Component



Real Image in frequency domain



Fake Image in frequency domain

Slight differnces but not enough for discrimination

Initial Approaches

Database Analysis via Gradient Images

Real Image with less gradients

Deepfake Image with much more gradients

Feature extraction

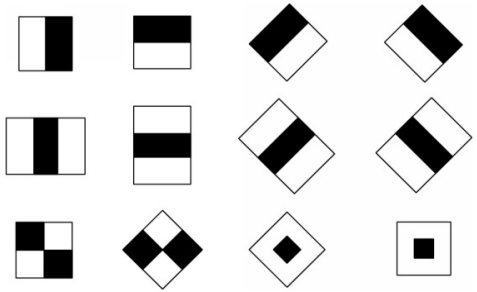
Feature Extraction

Fake Image

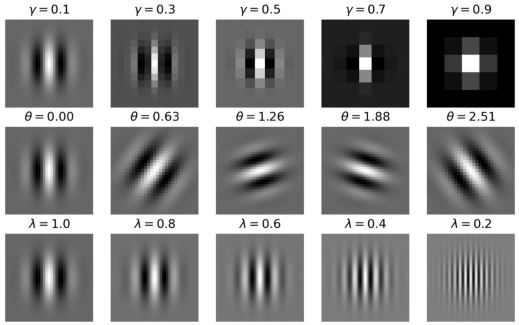
We can see that in the face image there are a lot of artifacts around the nose, eyes and mouth

Real Image

Feature extraction



Haar filter

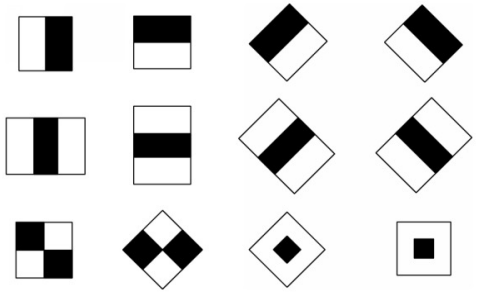


Gabor filter

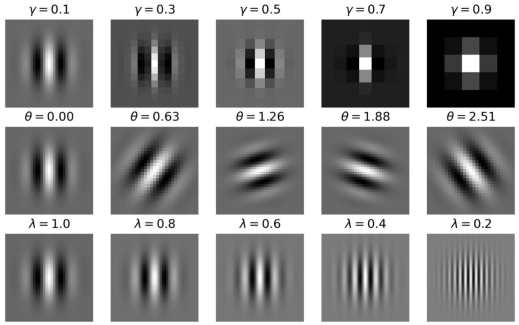
Image Processing AND Computer Vision

UAM Universidad Autónoma de Madrid

Feature extraction



Haar filter



Gabor filter

Image Processing AND Computer Vision

UAM Universidad Autónoma de Madrid

Initial Model Architecture

Training models on Handcrafted features alone

Statistical Classifier	Accuracy
KNN Classifier	0.84
SVC	0.828
Logistic Regression	0.747
Decision Trees	0.803
Naive Bayes	0.823

Accuracy results on evaluation set

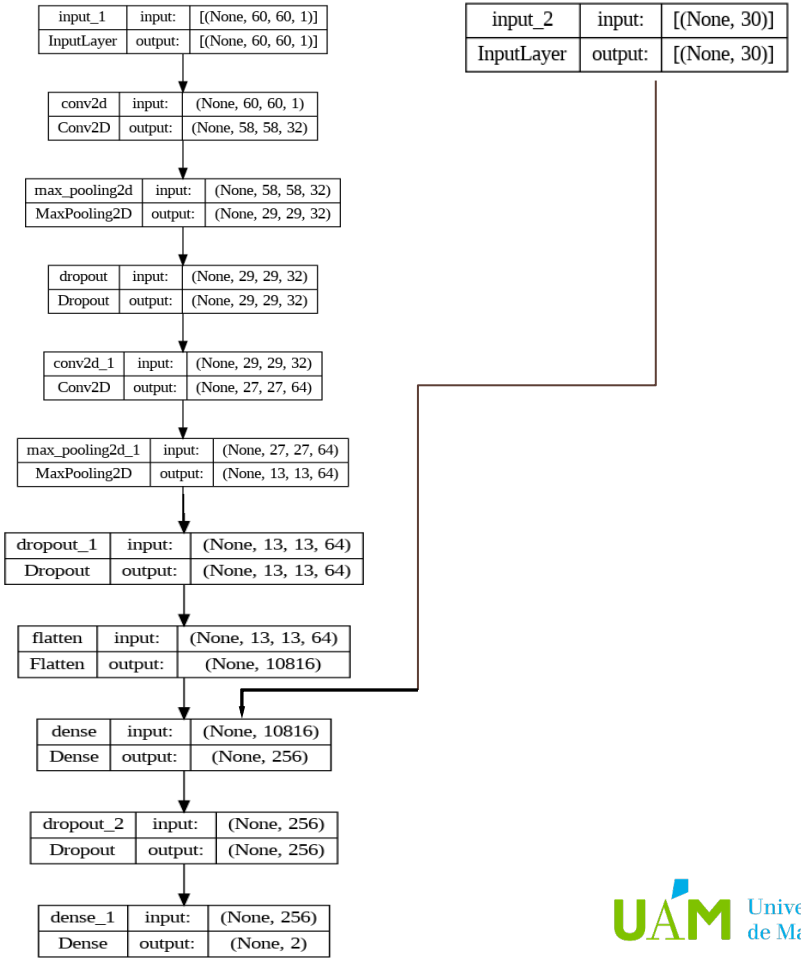
Final Model Architecture

Two input layers

- Input Layer 1: Resized 60x60 face image for CNN feature extraction
- Input Layer 2: 30 Extracted Handcrafted features using HAAR and Gabor filters

Output layer

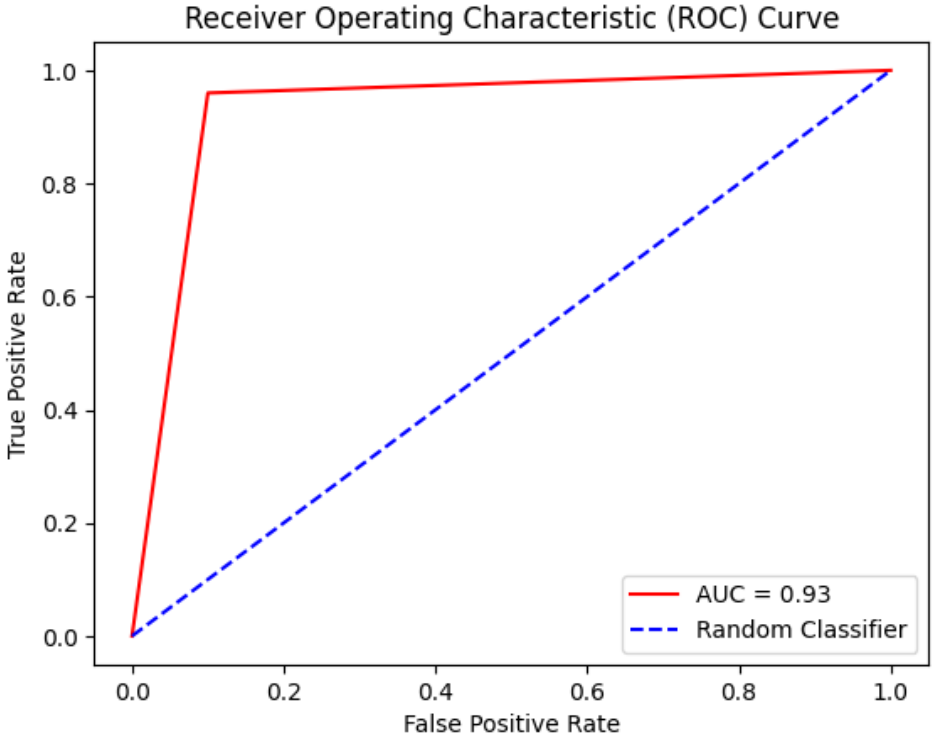
- Results of prediction per frame



Results

Task 1 Evaluation

Evaluating UADFV (evaluation) dataset using Model trained on UADFV dataset

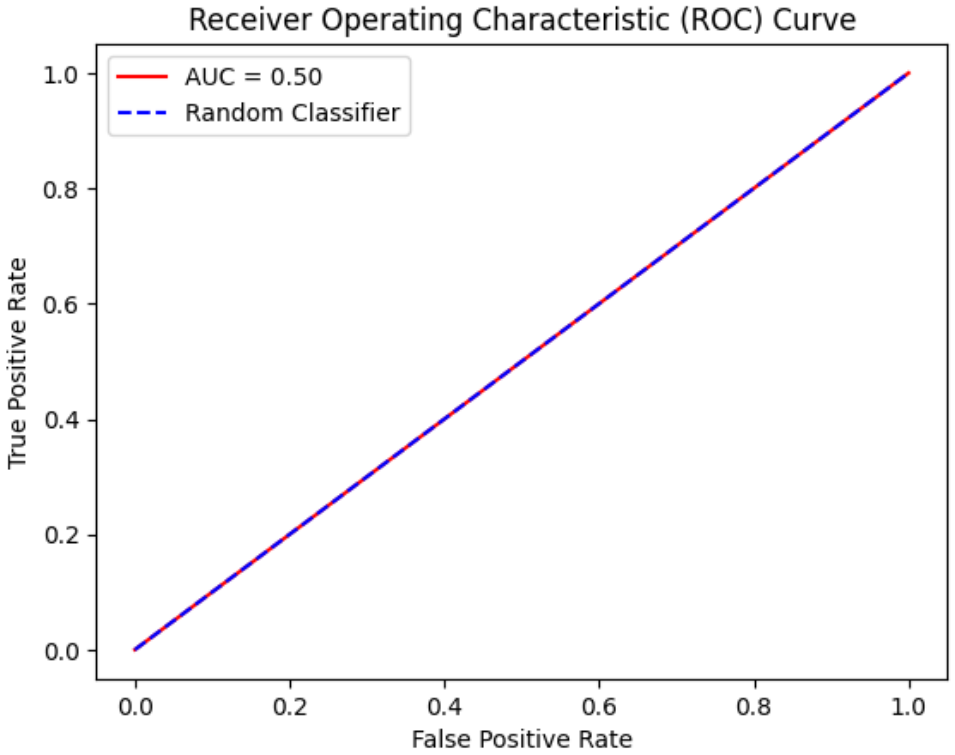


ROC Curve

Results

Task 2 Evaluation

Evaluating CELEBDF dataset using Model trained on UADFV dataset

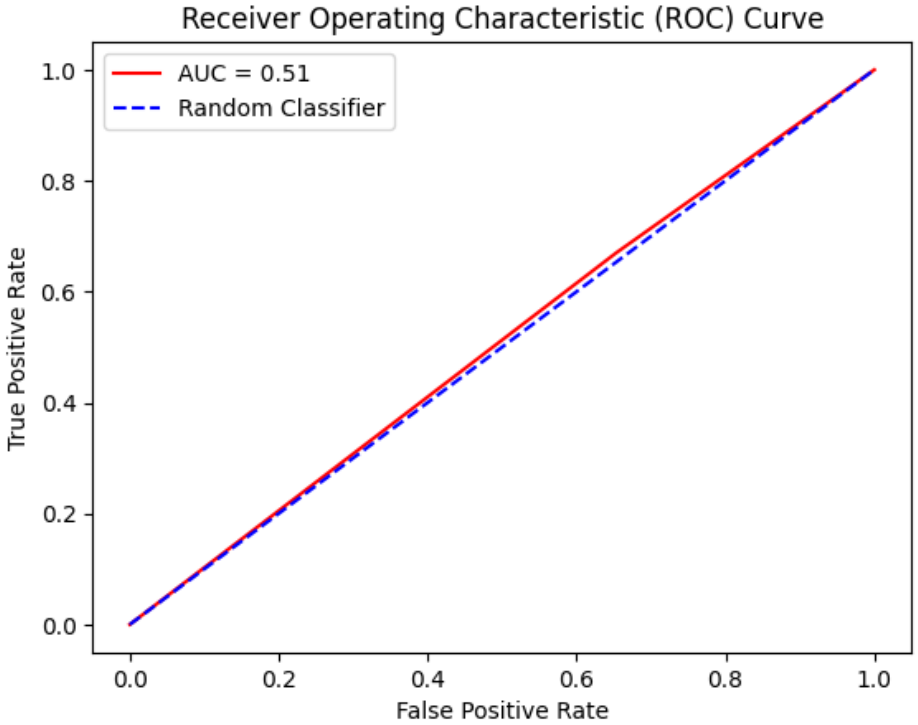


ROC Curve

Results

Task 3 Evaluation

Evaluating CELEBDF dataset by testing on pre-trained MesoNet Model.



ROC Curve

Conclusion

- The first task proved that a combination of Handcrafted and Deep learning features provide better results as the AUC of statistical classifiers trained only on Handcrafted features was 0.84, compared to 0.93 obtained via the fusion

- Performance of Deepfake detection models depends significantly on the similarity between the training and the evaluation dataset

- Training Deepfake detection models on frames from a sequence of files would aid model performance as prediction is not done on just a single frame, but on a couple of frames in the sequence.

Thank you!