Deliverable 4

by Sarah Zhou

**1. Problem Statement & Background**

This is a current real-world problem where the objective is to accurately determine if a video contains deepfake content. Research papers have shown that human behaviours such as the natural blinking pattern of the eye is hardly mimicable by deepfake videos. Theoretical models have been discussed namely involving a mixture of LRCN and LSTM. However, the challenge remains to implement the detection of these faults in a machine learning model with over 50% accuracy.

**2. Data**

The large dataset of 400 videos of a total of 470GB is provided by the Kaggle challenge. The one label per video is either REAL or FAKE, used for binary classification and feedback. Two pretrained models were considered during this project. One was pretrained on the FaceForensics++ dataset and the other, on the Facenet dataset.

**3. Methodology 1**

The first pretrained facial detection model chosen is MTCNN and Inception Resnet models. Using Facenet by Pytorch, the face feature vector were calculated for all the face in each video. The distance from each face to the centroid for its video was then calculated and acted as means of discrimination [4].

**4. Results 1**

From figure 1.0, we see that the model performed poorly for a binary classification problem. The clusters are continuous and there is a lack of separation of the output. The small rage of 0.46 to 0.56 cannot be used to distinctly. Although a face detection pipeline was successfully created, it was not effective.

**5. Methodology 2**

The second pretrained model released by **FaceForensics++** used **dlib package** and had a significantly better jump off point then the previous model. It uses 3 types of image preprocessing and applied many models namely the Xception model, a 71-layer CNN pretrained on more than a million images from the ImageNet. The author has provided the training results of several models. The comparison on training set resulted that **the Face\_detection Xception model with all compressed 23 images performed best**.

**6. Results 2**

The predictions were run on 50 frames at a time and were averaged into a single value. 0.5 is the value returned when unable to predict. For every video, we return the maximum, minimum and average all "fake" prediction, which is all the frames compiled together [2].

In the **final training predictions**, the model has a somewhat clear binary partition, with almost no confusion on real videos.

In the **final test predictions**, the model was confused about 80 videos.

**7. Conclusion**

The rest of the videos follows a somewhat normal distribution. (Since Kaggle holds a private test set, we cannot compare thoroughly the results. e.g. See how the test set differs from the training set to then conduct an analysis.)

Certainly, the results are not satisfactory from the models we have seen in this project. Many combinations of CNNs and LSTM should be tried out and further tuned to solve this important problem.

This challenge certainly holds high value in terms of application in real world products. A successful model can be commercialized to help the public detect fake news, help in cyber security, assist the police for criminal cases and serve as an educational tool.

**Bio:**

Sarah is a second year Statistics and Computer Science student. Her interests include using machine learning to explore the field of data science and bridging the gap between data analysis and data visualization in real-world settings.

**Headshot:**

A person posing for the camera

Description automatically generated

**Poster screenshot:**

A screenshot of a social media post

Description automatically generated

**Blurb:**

In this digital age, the diverse media channels facilitate the transmission of information. Without caution, fake news can be spread across the globe at lightning speed. It is thus extremely crucial to investigate the source and validity of information received. Sarah’s final project consist of exploring the current issue of deepfake content in videos which is arguably a form of identity theft and can harm the validity of information as well as an individual’s reputation. She has explored two models pretrained on face detection which implemented MTCNN and Xception.