Deepfake Detection Challenge

Kaamraan Khan [MT19064], Vedant Desai [MT19074]

Motivation:

Deepfake techniques, which present realistic AI-generated videos of people doing and saying fictional things, have the potential to have a significant impact on how people determine the legitimacy of information presented online. These content generation and modification technologies may affect the quality of public discourse and the safeguarding of human rights—especially given that deepfakes may be used maliciously as a source of misinformation, manipulation, harassment, and persuasion. The generation of these malicious techniques will not only increase in magnitude but also advance technically. Thus, we aim to devise innovative strategies that will suffice the need to counter such deepfake techniques.

Data Acquisition:

AWS, Facebook, Microsoft, the Partnership on AI's Media Integrity Steering Committee, and academics have come together to build the deepfake detection challenge and its corresponding dataset. The dataset is divided into four sets.

1. Training Set: - The dataset is further divide into 50 files consisting of videos. And a corresponding label is given with each video, REAL or FAKE.
2. Public Validation Set: - It is a small set consisting of 400 videos which can be used for validation.
3. Public Test Set
4. Private Test Set: - It contains videos with a similar format and nature as the Training and Public Validation/Test Sets, but are real, organic videos with and without deepfakes.

Data Pre-processing:

We have observed that there a few missing samples. We would work on it. The data is skewed with about 80% belonging to FAKE label. We aim to perform stratified sampling as a counter measure. Our focus would be more devoted in understanding the facial features like eyes, lips, nose and more. Thus using object detection, we would crop out the important parts.

Evaluation Metric:

For our problem, we will use log loss error.
The use of the logarithm provides extreme punishments for being both confident and wrong. In the worst possible case, a prediction that something is true when it is actually false will add infinite to your error score. In order to prevent this, predictions are bounded away from the extremes by a small value. Thus, out of all the evaluation matrices log-loss seems the most suitable.

Deliverables:

Preprocessing part 1: Kaamraan Khan
Preprocessing part 2: Vedant Desai
Merging preprocessing parts: Kaamraan Khan
Strategy Plan and framework: Vedant Desai
Data transformation: Kaamraan Khan
Data segregation and image analysis: Vedant Desai
Modelling part 1: Kaamraan Khan
Modelling part 2: Vedant Desai
Model analysis and selection: Kaamraan khan and Vedant Desai