Efficacy of Deepfake Detection Methods

Saahil Sood

Background

- A deepfake is a specific kind of synthesis media where a person in an image or a video is swapped with another person's likeness.
    - Typically uses techniques from machine learning and artificial intelligence to generate the highly deceptive media
- There are three main categories of deepfakes as of today:
    - Faceswap
        - lightweight and heavyweight
    - Facial expression modification
    - Synthetic generation of faces
        - Uses generative adversial networks StyleGan to generate these images

The issue

- Deepfakes are becoming more sophisticated - it may be difficult for a person to distinguish between a real image and a deepfaked image
- Deepfakes are becoming faster and easier to make
- Debunking deepfakes is becoming harder as innovation outpaces counterefforts

Project

-My project aims to test the current efficacy of deepfake detectors with all forms of deepfakes and looking for areas for improvement

-Dataset: Faceforensics++ and StyleGAN-generated images

-Two reputable deepfake scanners: MesoNet and DeepwareAI

-StyleGAN will be used to generate realistic looking faces with particular facial features (long hair, short hair, etc.)