Using Multimodal Biometric Bystem to Recognize Fake Videos CS282A Proposal

Daniel S. Rincon, Shruti Agarwal, Tanya Piplani

February 2019

1 Related Research

There has been research around Exposing DeepFake Videos By Detecting Face Warping Artifacts [LL18] that describes a new deep learning based method to distinguish AI-generated fake videos from real videos. The method is based on the observations that current DeepFake algorithm can only generate images of limited resolutions, which need to be further warped to match the original faces in the source video. Such transforms leave distinctive artifacts in the resulting DeepFake videos, and can be effectively captured by convolutional neural networks.

2 Research questions

Do individuals have specific facial movements that correlate uniquely with certain spoken words or phrases (speech biometrics)? (I'm not sure if this might be outside the scope) Do individuals have a specific speech biometric signature? Can we identify an individual by its specific speech biometric signature? Can we train a classifier that uses these speech biometrics to classify a person? Can we use this classifier to detect 'impersonations' that use face translation techniques?

3 Hypothesis

We believe that by extracting a set of over 100 facial expression features for each frame in a video, we can train a classifier to predict if the person speaking in that video is a given subject or not; even if the video is generated using face-to-face translation deepfake techniques.

4 Method

4.1 Fake Generation

There are two common techniques of producing fakes:

1. Audio-Mouth mapping (Audio-Video) : This technique trains a model to replicate the mouth movements that are associated with specific audio snippets. Therefore when the audio if faked that mouth movements are biometricaly accurate. This technique seems to be restricted to using original (or very similar) audio. Nonetheless given the biometric accuracy our proposed approach would not seem to work for these fakes.

2. Face-to-Face translations (Video to Video): This technique trains a model to map facial movements from one face (video) to another one. Given that the face movements in the output match those in the input, the output face can be faked into saying anything uttered by the input.

4.2 Detecting Forged Videos

We want to extract facial and biometric features like

5 Dataset

We will test this technique by building a binary classifier. Using data from the following sources: (Are we gonna do Obama vs. not Obama or multiple categories?)

Positive samples: Videos selected by [SSK17] which include over 14 hours of video from former President Barack Obama's weekly addresses.

Negative samples: Data from [**?**] (VoxCeleb2) data set that includes over 1 million utterances from 6112 celebrities, extracted from Youtube videos.

6 Tools

Compute/Storage: AWS S3+EC2 P3 Instance (GPU) Libraries: TensorFlow, Keras, (@Shruti, insert facial featuring and other video preprocessing library here)

7 Evaluation

References

[LL18] Yuezun Li and Siwei Lyu. Exposing deepfake videos by detecting face warping artifacts. *CoRR*, abs/1811.00656, 2018.

[SSK17] Supasorn Suwajanakorn, Steven M. Seitz, and Ira Kemelmacher-Shlizerman. Synthesizing obama: learning lip sync from audio. *ACM Trans. Graph.*, 36(4):95:1–95:13, 2017.