Deepfake Audio Detection

MENTORS: BRANDON WU, JERRY LIAO
MEMBERS: CHING-YUAN PAI, YIPIN PENG, ANDY CHEN, EVIAN CHEN

INTRODUCTION

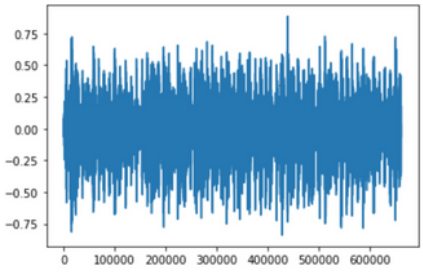
- As the application of the Internet continues to expand, the impact of synthetic audio cannot be underestimated. Though AI has been able to successfully synthesize voices, Synthetic audio could be like our relatives, friends or trusted experts, which are difficult for human ears to recognize.
- This project aims to build a system through CNN to detect the authenticity of any audio file.

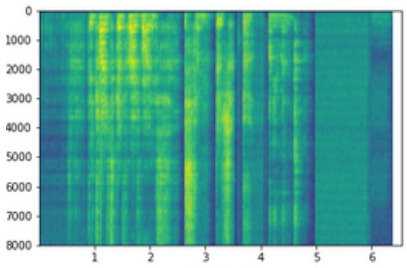
DATASET

- ASVspoof is a challenge hold in bi-annual which aim to promote the design of countermeasures to protect automatic speaker verification systems from manipulation. There are two types dataset and all is saved in FLAC.
- PA is made in a real physical space
- LA is generated using TTS and VC algorithms.

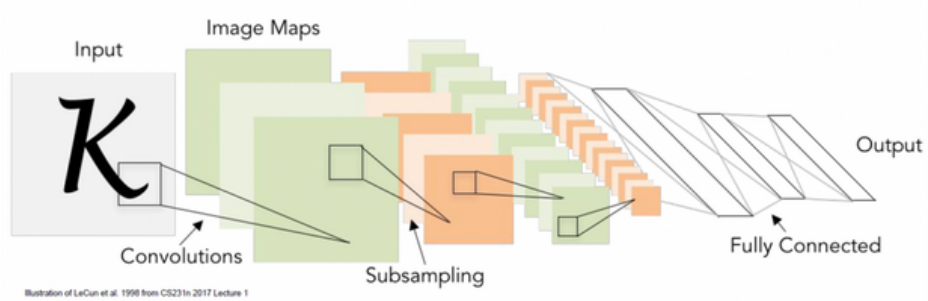
MRTHODOLOGY & RESULTS

- We convert FLAC to WAV, and use PyTorch module to generate waveform and spectrogram which characterizes in three features: frequency, time and intensity which shown by varying the color or brightness.





- Since our dataset contains 200,000 audio files, we found it extreamly time-consuming to load to G-drive, so we unzip the folder on colab to avoid run-time error.
- We construct two models for LA and PA dataset respectively.
- For LA dataset, we built a VGG-like model, which is 27 layers, to train and successfully achieve the accuracy of 91%.



- For PA dataset, we use ResNet50, which can be pre-trained by 1000 classifications dataset, and it achieves the accuracy of 95%.

DISSCUSSION & CONCLUSION

- Although spectrograms and waveform are two-dimensional information, the total number of datasets can't be increased by image flipping and other methods due to the time-ordered characteristics of spectrograms and waveform.
- According to the training results of the VGG-like model and the ResNet model, the validation accuracy is over 90%, and the individual accuracy of real and synthetic audio files is also similar.

FUTURE WORKS

- Establish an user interface where people can upload random audio files and obtain predections from our model.
- Test the effect of different compressed audio formats like MP3.

REFERENCE

[1] M. Todisco et al., "ASVspoof 2019: Future horizons in spoofed and fake audio detection", Proc. Interspeech, pp. 1008-1012, 2019.
[2] Y. Jia et al., "Transfer learning from speaker verification to multispeaker text-to-speech synthesis", arXiv:1806.04558, 2019.

OUR WORK

- poster session recording: link
- code files: link
- power point: link

MY GITHUB

- link

MENTORS: BRANDON WU, JERRY LIAO
MEMBERS: CHING-YUAN PAI, YIPIN PENG, ANDY CHEN, EVIAN CHEN