Stage 0: Adversarial attacks and anomaly detection in MIR

Vinod Subramanian, QMUL, ROLI Ltd.

Abstract

The abstract goes here.

Index Terms

MIR

I. Introduction

MUSIC companies have large collections of instrument sounds such as different drum hits, different violin playing techniques, piano sounds, environment sounds etc. Some of these sounds are live recordings of acoustic instruments and other sounds are created by sound designers using synthesizers. These large collections of sounds are used by musicians to compose and perform music. There are many challenges that come with these large sound collections.

For the creators of the sample packs it becomes challenging to curate the live recordings and manually listen for errors in the recordings. These errors can be in the form of environmental sounds, incorrect playing, audio corruption etc. Once all the sounds are created it is difficult to find an intuitive way to cluster the sounds. For acoustic instruments it might be slightly easier to classify a sound based on the source but it is much more challenging for synthetic sounds. In addition, for a sound designer it might be difficult to come up with ideas sometimes and it would be useful to have a tool that sounds to serve as inspiration.

On the musicians side given a sample pack of a 1000+ sounds it becomes cumbersome to navigate the sample pack to identify the sound they want. There isn't an intuitive search mechanism to search for sounds or identify similar sounds. This lack of navigability might be counterproductive to sample packs that boast large collections of sound.

Therefore, there is a need for research in the area.

II. Related Work

A. Instrument recognition

B. Playing technique detection

C. Audio clustering

Given a large collection of audio how do you classify it for efficient search and retrieval to make it user friendly

D. Fault detection

Given a large dataset of drum hits, how do you find the faulty recordings that contain too much noise

E. Multi-task learning

Learn multiple tasks with the same model to improve performance

F. Interpretable deep learning

Understanding what deep learning models are learning in a particular task

G. Different inputs for deep learning

Understand the strengths and weaknesses of different inputs for the deep learning tasks listed above

H. Adversarial attacks in audio

Related to fault detection, how do the classification models deal with data that doesn't belong to any of the categories

III. Proposed Methods

The research in the project is done in collaboration with ROLI ltd. ROLI is going to provide some of the datasets used in the research. Primarily there are three types of datasets:

1) Large corpus of drum strikes which need to be curated before releasing to the public
2) Public note level datasets
3) Rhythm pattern datasets

IV. Evaluation Methods

V. Timeline

VI. Conclusion

Acknowledgment

References

[1] H. Kopka and P. W. Daly, *A Guide to LaTeX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.