Stage 0: Adversarial attacks and anomaly detection in MIR

Vinod Subramanian, QMUL, ROLI Ltd.

Abstract

The abstract goes here.

Index Terms

MIR

I. Introduction

1) Introduce the increasing importance of data and how there is an abundance of large collections of data in the world
2) Give examples of movie collections, music streaming, personal information, instagram, youtube etc.
3) This data has facilitated research in deep learning. So a combination of an abundance of data and better computing has brought about a new age of deep learning
4) Challenge with larger datasets is mislabelling, data corruption, noise which are currently being done by mostly manual curation there is a need to work on automatic methods to curate. Practically this manifests as when you have live data coming in quickly identifying an error.
5) Challenge with deep learning is that while deploying them in the real world you need to be sure they are working well and always perform as expected, in effect they must be robust to noise
6) ROLI is interested in problems such as acoustic instrument dataset creation where high quality recordings are key, so an automatic method to identify faulty recordings that don't meet the standard
7) ROLI is interested in high accuracy and highly robust deep learning models that they can confidently release to their customers so there is a need to do research in investigating robustness

II. Related Work

I'd like to see if I can create a summary of some of the research techniques used in some of the literature I am reviewing in the form of a table to compare techniques and performance.

A. Adversarial attacks

1) Discovery of adversarial examples where small perturbations in data causes high accuracy deep learning models to misclassify the input
2) Further, these adversarial examples were transferable between different models which is alarming
3) More and more efficient and cheap ways to generate adversarial attacks were discovered in an attempt to understand the underlying problem of deep learning models
4) Simultaneous work on solving these attacks became popular with distilled learning methods
5) More state of the art approaches towards defending from attacks happened
6) Most research was in computer vision or natural language processing, Carlini found that some of these adversarial attacks work in audio as well
7) Bob Sturm showed that these attacks exist in music too

B. Anomaly detection

1) Summarize the survey paper on anomaly detection from 2007. This survey paper talks a lot about older approaches to anomaly detection that are still relevant
2) For mechanical failures there are deep learning approaches for anomaly detection, mostly inspired by the fact that it is harder to define the anomaly.
3) Work done on anomaly detection in medicine to identify diseases from measurements automatically
4) Work done on anomaly detection on GTZAN dataset using classical machine learning and on mammal sound recognition

III. Research questions and methods

In this section I'll list all of my ideas for this project so far which I will reference again in the timeline section. I am going to try and include block diagrams and images to make my ideas clearer.

A. Recreating adversarial attack experiments in audio

1) Transferring adversarial examples in audio between models with identical inputs: In computer vision they found that adversarial examples generated in one model served as an adversarial example in another model. This indicates that there seems to be some fundamental way in which we train deep learning models that make them susceptible to errors
2) Can adversarial examples be played over the air: In computer vision a photo of an adversarial example taken from a phone camera still works as an adversarial example
3) Can you create a universal perturbation that has a high chance of making an audio an adversarial example: In computer vision they came up with a universal method to create an adversarial example

B. Adversarial attacks for audio

1) Unlike computer vision audio domain uses different types of input features for deep learning. It would be interesting to see if it is possible to design an adversarial attack in the time domain that is robust to the different input representations
2) Design attacks for a specific task such as singing voice transcription. Singing voice transcription has a few different high accuracy models to work with so it is a good candidate to perform adversarial attacks on
3) Some research is done showing that mp3 compression preserves the adversarial examples but it would be interesting to try different bitrates and different compression methods

C. Defenses against attacks

1) Apply existing defenses to the audio problem to see if they work
2) Design defenses that are tailored to the specific tasks that I choose to work on
3) Explore the idea of using adversarial examples as a data augumentation technique to create higher accuracy models

D. Anomaly detection in note level datasets

1) Ensemble of methods to identify poor quality, corrupted audio and noise in datasets
2) Replace ensemble of methods with a single deep learning model and explore the use of multi-task learning

IV. Timeline

Table listing different projects, their timeline and targeted conferences

V. Conclusion

Summarize the whole proposal

Acknowledgment