Carleton University
COMP4107A–Fall18

Music Genre Identification

Using Convolutional Neural Networks on the FMA dataset

David N. Zilio	100997259
Aidan Crowther	#########

16th December 2018

Abstract

Music can be a very simple thing for a person to categorize and sort based on the sound of it but, could a neural network be trained to do the same kind of categorization? The convolutional neural network (CNN) which was built was designed to answer this question. Thirty second samples of music with additional metadata, from the Free Music Archive, were converted into heatmaps of total power throughput in frequency bands vs time at some bitrate. The heatmaps were then fed to the CNN with a genre label to train it in the identification of patterns which relate to each of the specified genres. Although computationally intensive and difficult to tune, the interpretation of generated heatmaps into genre groupings with a CNN was found to be possible.

Background & Introduction

- Expand this into paragraphs
- add references to for the reasoning through it
- who came up with this shit?

Music, the art form that it is, can be categorized by people in a very subjective manor. Attempts could be made to compare what makes pieces of music similar. We have many metrics by which music is written, which leaves the possibility to gauge things based on music theory. Solutions based on the decomposition of these identified characteristics, although intuitive fall into line with what would be defined as an expert system. Expert systems have been shown however to be ineffective; one needs look no further than the Cyc AI project which ultimately can't provide the required accuracy. The modern truth of the matter is that for a machine to be able to effectively and automatically analyze data, it needs to be shown whole examples.

The translation of something as complex as the waveforms of music into something that could be consumed all at once for analysis is complex. To overcome the issue a solution was derived to make music into a digital image. An image is an ideal form because there has been extensive research into the analysis and identifications of images using neural networks for the purposes of computer vision. Out of the neural networks used to achieve computer vision, convolutional neural networks (CNN) have been found to be extremely adaptable to the image recognition problem. They have even been found to be capable of near human accuracy with identification of objects given a large enough training set.

With a neural network identified as a capable system for image analysis, identifying an encoded piece of digital music doesn't translate well. To remedy the issue a preprocessing step was required to make music the ideal form, an image. The preprocessing was designed to take the audio clips in mp3 encoding and separate the files by varying ranges of pitch with a 256-value integer to represent the power output of the range. The resulting ranges were plotted as pixels on a vertical vector representing a slot of time which was then iterated over the length of the clip. The preprocessing step, by design, reduces the quantity of information by grouping the sets of ranges together.

Problem statement

Why would a machine need to be able to identify the genre, or some other feature, of music?

Results and Discussion

- Preprocessing
  - Loss?
  - What makes a good ‘image’?
  - Dataset corruption??? (honourable mention maybe, not necessarily relevant)
  - Reference to talk about resolution?
- CNN?
  - How did the CNN apply? – they’re good for images but music isn’t an image????
  - Layers?
  - Optimizers
  - Accuracy, train vs test
  - Overtraining?
  - References for days here man
    - Is our as good as the referred? Why?
    - How does our network differ from the way they designed theirs?
    - Adv/disadv of how we did ours vs how they did theirs
    - Justify why ours is the way that it is
- Application
  - Could this extract more information than just a genre?
  - Does it work as expected? Why?
  - Is it currently good enough to be used for more than just academic purposes?
  - Reference current applications of this technology

asdf

Conclusions

- Does the thing work?
- Well?
- Did it achieve the objective of creating it?
- Can it be used for more than this?

Asdf

References

Defferrard, Michaël, Kirell Benzi, Pierre Vandergheynst, and Xavier Bresson. 2016. "FMA: A Dataset For Music Analysis." *arXiv.org*. https://arxiv.org/abs/1612.01840.