An Exploration of the Usage of Convolutional Neural Networks in Automatic Music Transcription

By Nakul Iyer, with Dr. Nicholas Zufelt
Concluded February 28, 2019

Repository

Available for Free Download at
https://github.com/nakuliyer/music-transcription-engine

Research Summary

This research paper supplements the application developed as a part of my Computer Science 600 (Research and Development) class. The research covered the length of one trimester and a two-week-long school break, totaling about 12 weeks. I was responsible for working an average of about an hour per day during this time frame. My main submission in this class comprised a structured, multi-file python program and an accompanying paper detailing my experimentation and results. My project and I were featured in my school's newspaper, the "Phillipian".

Abstract

Computers today possess great capability to aid musicians in the creation of music and help them preserve their work. In order to create music of their own, algorithms can exploit the fact that most songs follow similar "rules"; these consist of common scales, arpeggios, or chords. The purpose of this study was to train an algorithm to understand these "rules" and transcribe provided audio files into printable sheet music. This task was primarily complicated by musical overtones and similarity between adjacent semitones. Additionally, this project involved polyphonic music, meaning that more than one note could be playing simultaneously. In this study, a convolutional neural network (CNN) architecture was trained on ParisTech's MAPS (Midi-Aligned Piano Sounds) dataset, which consisted of scales, chords, and songs. CNNs are a relatively lesser-used model in music transcription. They were chosen as they could detect patterns over time, frequency (note pitch), and intensity (note volume). Two separate models with slight differences in convolution window sizes were trained over 100 epochs. These models yielded accuracy measurements, in F1 scores, of 0.967 and 0.962, out of a maximum of 1.

Upon analysis, the neural networks were partially successful in transcription. Transcriptions of single notes were generally correct, while the quantized estimates of note onset times were slightly less accurate. This led to some chords being miscategorized as single notes. However, scales and songs with separate single-note left and right-hand components were often interpreted with little errors. This showed that, in the future, CNNs may become more reliable with testing of different architectures and longer training times.