Machine Learning in Music Transcription

Andrew Zhuravchak, Yuriy Antentyk
https://github.com/antentyk/MusicTranscription

1 Motivation

Music Transcription is a process of extracting sheet music from audio recording (in a nutshell, mp3 to midi converter). There are several possible applications, some of which include:

- smuggling (some sheet music is not free of charge)
- improvisation capturing
- helping newbies to learng how to play musical instruments

2 Related Works

There are couple of similar works, such as Google Magenta project. While gathering information, I have seen the following approaches to solving this problem:

- LSTM/RNN
- DNN
- CNN

I decided to focus on a paper Luoqi Li, Isabella Ni and Liang Yang, Music Transcription Using Deep Learning, 2017 and try their DNN approach.

3 Dataset

As in majority of the papers I ran into, I was using MAPS database, which contains ~30GB of labeled piano recordings (isolated notes, chords and music pieces) in various conditions played on disklavier (piano, that can play midi without a man).

4 Data Preprocessing

All .wav files in the dataset were preprocessed using CQT (Constant Q Transform), which is a form of STFT (Short Time Fourier Transform). CQT tries to retain the same frequency resolution through logarithmic scale of frequencies in the piano keyboard by varying window size.

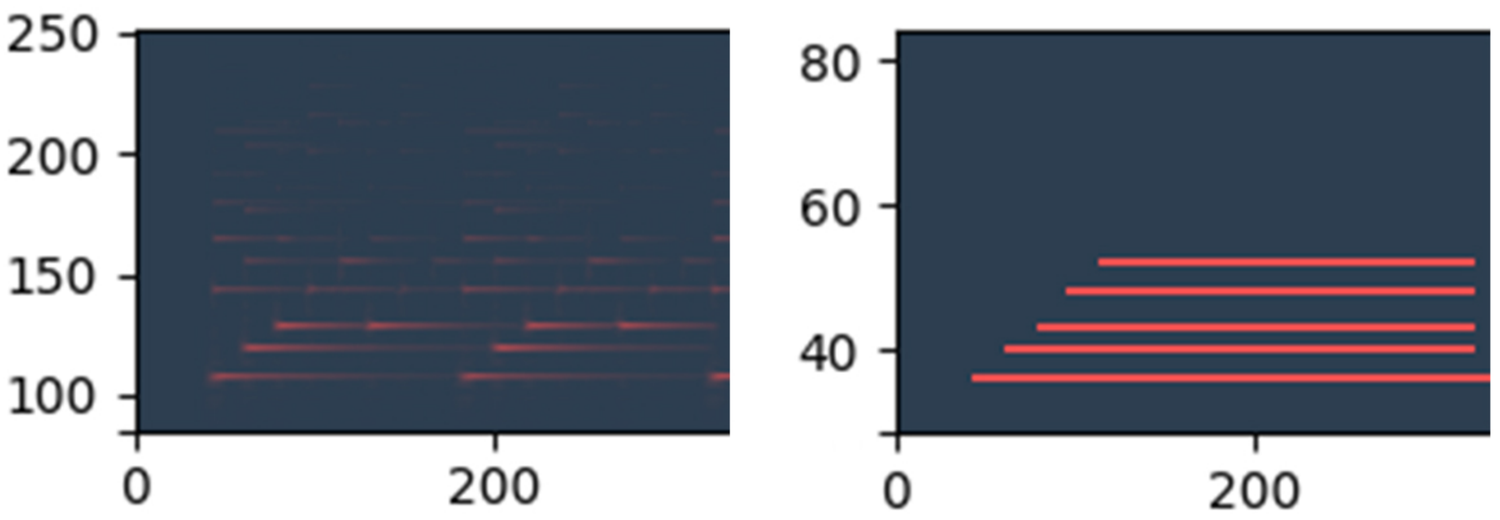

Figure 1. CQT applied to "Ave Maria Bach Gounod" (on the left) and MAPS ground truth (on the right).

5 Model

NN Model described in the article, consisted of 3 fully connected layers with relu activation function with Sigmoid at the end. Authors used BCE loss function. I slightly modified this arcitecture: removed Sigmoid and add 10% dropout and batchnorm.

6 Results

The model performed with 64% percent of accuracy, which is still not good enough to hear good music (after removing too short notes (0.1 second threshold), we can hear something remotely similar).

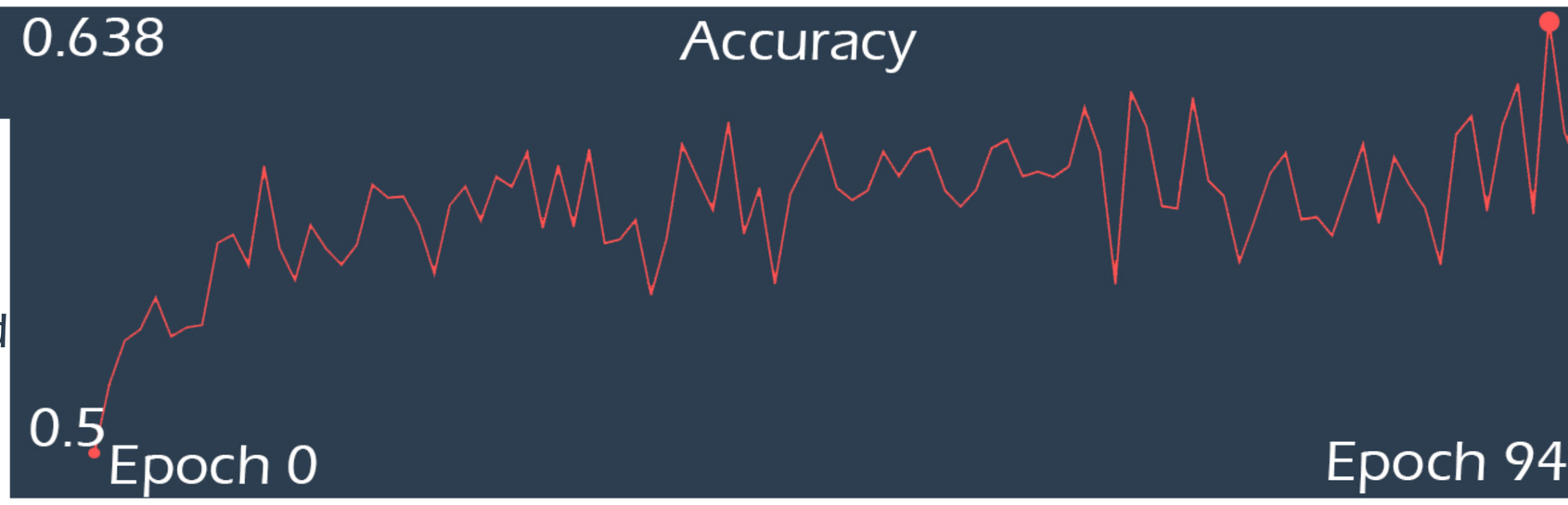

Figure 2. validation accuracy of 3-Layered DNN over time

However, there still are some problems with dataset interpretation, sustain pedal, difficult classical pieces and preprocessing (uncertainty principle and harmonics).

Applied Science Faculty, Ukrainian Catholic University, Lviv, Ukraine (antentyk@ucu.edu.ua)