Transcribing Solo Piano Performances from Audio to MIDI using Neural Networks

Victoria Ebert

Abstract

Audio transcription is a complicated problem in the field of music information retrieval. Many attempts have been made in this field using different types of neural networks and different formats of the audio and MIDI. This attempt uses a BLSTM and a time-series MIDI.

Introduction

Transcription is the act of conversion to a symbolic format. For musical transcription, that format is MIDI.This project uses a neural network to transcribe audio recordings of solo piano performances into MIDI files.

Methods

Data was collected from the MAESTRO dataset [1]. The audio was then downsampled to 16kHz, and the MIDI was converted to a time-series. The specific type of neural network used for this project is a Bidirectional Long Short-Term Memory (BLSTM) neural network where the input is an audio file, and the output is a time-series representation of a MIDI file. Currently, the network is only being trained on a single song. The results of the network were evaluated using subjective measures, i.e., does this sound like what it is supposed to.

Results

When the velocity of the song was included in the training, the neural network produced a MIDI file that got quieter when the audio got quieter, and louder when the audio got louder. The MIDI also, in general, has higher notes in places where the audio has higher notes, and the same thing with the lower notes. The main difference between the true MIDI and the predicted MIDI is the specific location of the notes, and the duration.

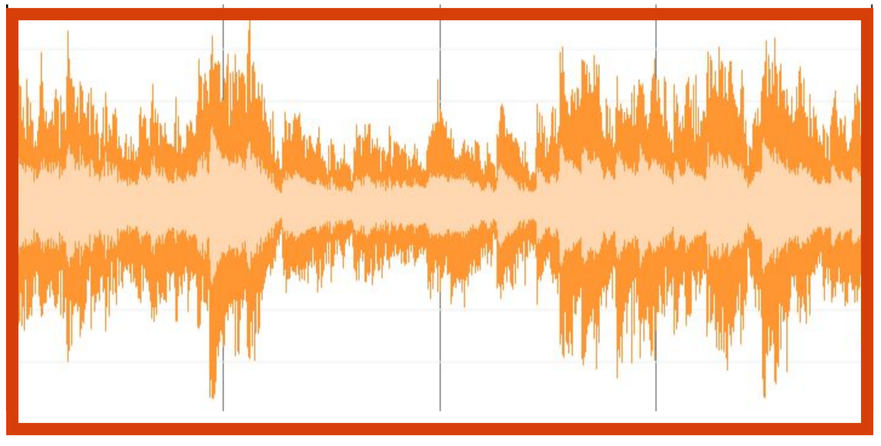

Figure 1: The audio input to the neural network

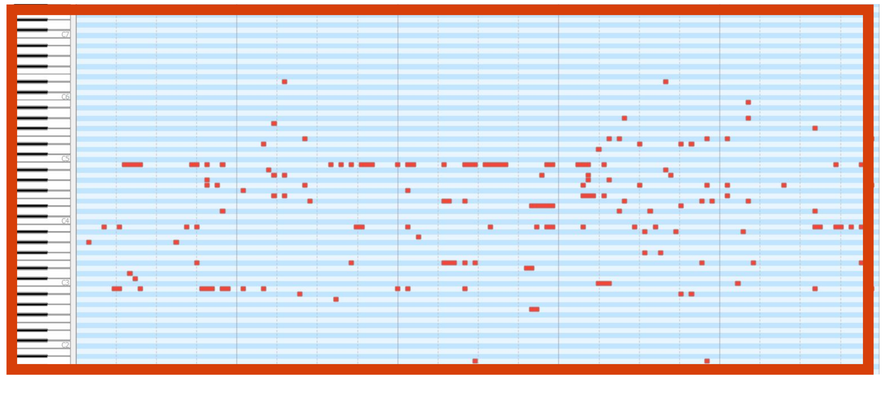

Figure 2: The MIDI output of the neural network

Discussion

Given that the neural network was only trained on one song, there was not a lot of information that it could learn. I am currently working on training the network on the entire MAESTRO dataset [1], which should allow the network to perform better when given a song that it has never seen before. In general however, these results are very promising, and point towards a future breakthrough.

Acknowledgments

Thank you to Patrick Donnelly, and the Soundbendor Lab

References

[1] Hawthorne, C. et al. 2019. Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset. arXiv:1810.12247 [cs, eess, stat]. (Jan. 2019).

Oregon State University