

Midi Shark: For Piano Transcription

Jonah Chen, QiLin Xue, Joe Hattori, Khanatat Thangwatthanarat

University of Toronto

{jonah.chen,qilin.xue,joe.hattori,k.thangwatthanarat}@mail.utoronto.ca

December 8, 2021

1 INTRODUCTION

Transcription of music is the process of determining the pitches and timing of notes from recorded audio files. Transcription has always been a specialized task that requires years of musical training. Transcription is even more challenging for polyphonic music, such as piano, which features the simultaneous production of two or more tones. The majority of traditional transcription models focus on extracting all of the notes from the recording using the note onset. This, however, is not the way a trained musician approaches the problem[1].

We developed a model that is more accurate at transcribing piano recordings by analyzing the recording with a neural network and focusing on both the onsets and offsets of the note. Moreover, since images and audios both have common two-dimensional time-frequency input representations, the fact that CNN performs well in image classification problems suggests that CNN could potentially be used for music transcription [2].

2 ILLUSTRATION

3 BACKGROUND

4 DATA PROCESSING

5 ARCHITECTURE

6 BASELINE

7 QUANTITATIVE RESULTS

8 TESTING

9 DISCUSSION

10 ETHICAL CONSIDERATIONS

11 REFERENCES

REFERENCES

- [1] S. Hainsworth and M. Macleod, *The automated music transcription problem*, Department of Engineering, University of Cambridge, 2004.