Research on CRNN-Based Piano Music Transcription Systems

Yin, Victor (School: Tsinghua University High School)

Transcribing audio files such as MP3s into digital symbolic formats such as Musical Instrument Digital Interface (MIDI) files has always been a complicated task, referred to as Automatic Music Transcription (AMT). AMT is very crucial since many musicians nowadays need single track music sheets, MIDI, to analyze and perform musical pieces. However, they cannot do so on an MP3 file. On the other hand, archived piano pieces without music sheets or jazz improvisations need transcription as well. So far, there has been a plethora of research in this field. Yet none of them have attained satisfying results. For this, we divide the AMT into 2 jobs: note and pedal transcription. After creating a model for each purpose, we combined them and generated a model capable of converting a musical piece to a MIDI file. Both models were trained on MAESTRO's dataset, reaching a note F1 (see "Keywords" for definition) of 0.87581, pedal F1 of 0.88206, and a pedal frame F1 of 0.92802. The average human ear can barely hear the difference between our results and the original audio. We also evaluated the model on MusicNet's dataset. Because MusicNet included all kinds of instruments and our model only focuses on piano transcription, we kept one track and converted it to piano. We attained a note F1 of 0.77430. We also established a system to convert an MP3 file to a PDF music sheet. This could also convert a MP3 in piano to another instrument such as a violin, oboe, etc. The system was later compared with the transcription skills of professional pianist Yan Zhou. Transcription tests were based on variables such as duration, complexity, speed, rhythm, note amount, and key difficulty. The system is shown to have an advantage over human transcription abilities in numerous aspects.