

# Modeling Jazz Piano: Symbolic Music Generation via Large-scale Automatic Transcription

Andrew Edwards

2021

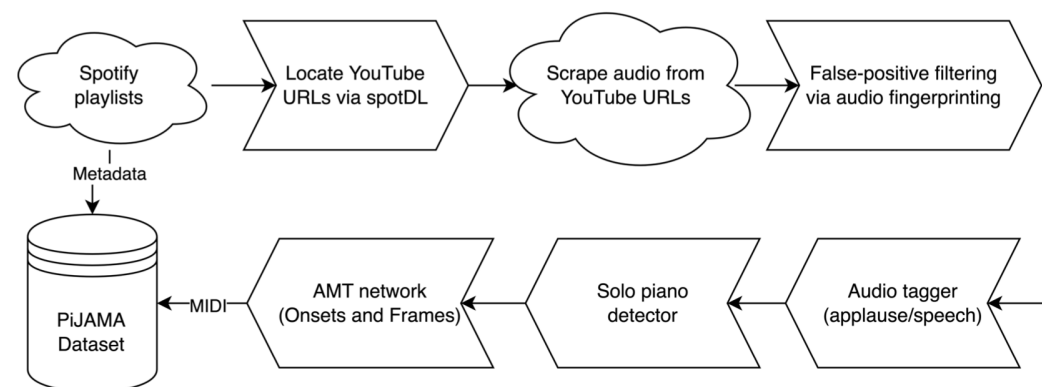


Figure 1: Diagram of the data collection process for the PiJAMA dataset.

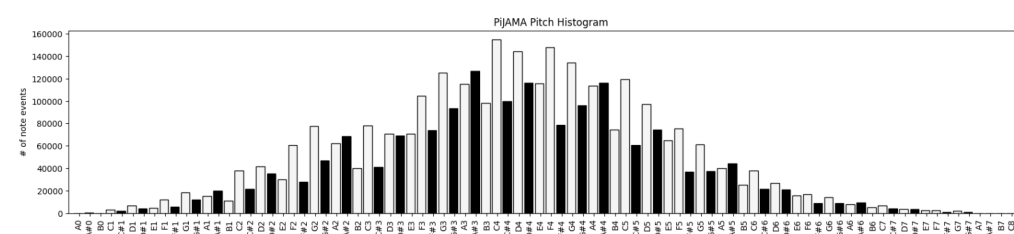


Figure 2: Pitch histogram of all note events in the PiJAMA dataset.

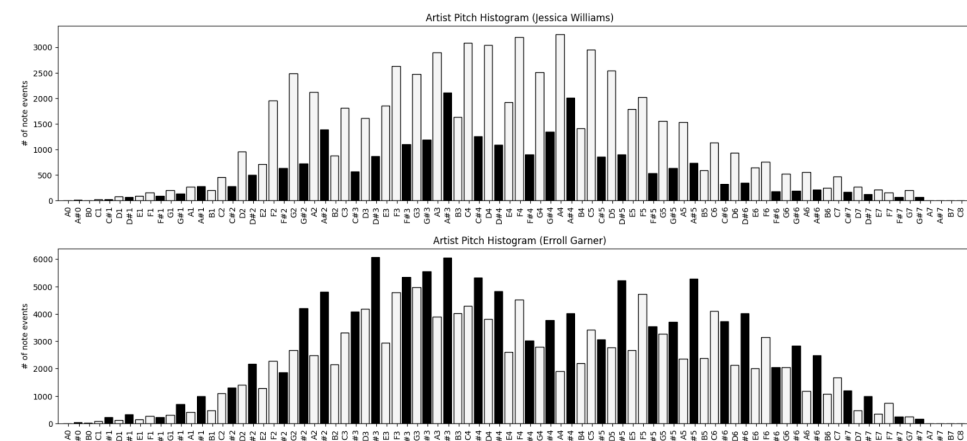


Figure 3: Pitch histograms from pianists Jessica Williams and Erroll Garner.

## Finding

In current S.O.T.A. piano transcription work, data augmentation is reported to hurt test set performance. We have preliminary results of matching S.O.T.A. MAESTRO test set results while also setting new S.O.T.A. results on out-of-distribution evaluation using a combination of data augmentation and automatic data collection.

## Question

Can we use automatic chord and melody detection on a large dataset of jazz piano MIDI to build a system to perform an automatic arrangement from a lead sheet?