

Automatic Music Transcription



Figure 1. The proposed Jointist framework. Our actual framework can transcribe/separate up to 39 different instruments as defined in Table 7 of Appendix. B : batch size, L : audio length, C : instrument classes, T : number of time steps, K : number of predicted instruments. Dotted lines represent iterative operations for K times. Best viewed in color.

Goal:

Audio \rightarrow MIDI (events note, on/off, time)

Method:

- Focus on vocal part
- Use *Demucs* for extracting vocal and instrumental parts.
- Apply *Basic Pitch*, *Key Detection*, and *BPM Estimation* to enhance vocal transcription.