

# Building Symbolic Musical Archives From Audio: Investigating State-of-the-Art Automatic Music Transcription Systems

Lele Liu

University of Würzburg  
lele.liu@uni-wuerzburg.de

Christof Weiß

University of Würzburg  
christof.weiss@uni-wuerzburg.de

## Overview

Recent years have witnessed substantial advancements in automatic music transcription (AMT) technologies. Among its various applications, one prominent use case is music analysis and archiving. Here, AMT tools facilitate the transformation of musical recordings into human- or machine-readable formats like MIDI and MusicXML. These explicit formats, in turn, enable better accessible analysis, preservation, and exchange of musical compositions. In this poster, we investigate several state-of-the-art automatic music transcription algorithms. We explore the models' performance on individual examples and across different musical instruments and styles. We further discuss the challenges and future directions in this field.

## Case Studies

Our exploration centers on several transcription systems. The selected scenarios for investigation encompass three core aspects:

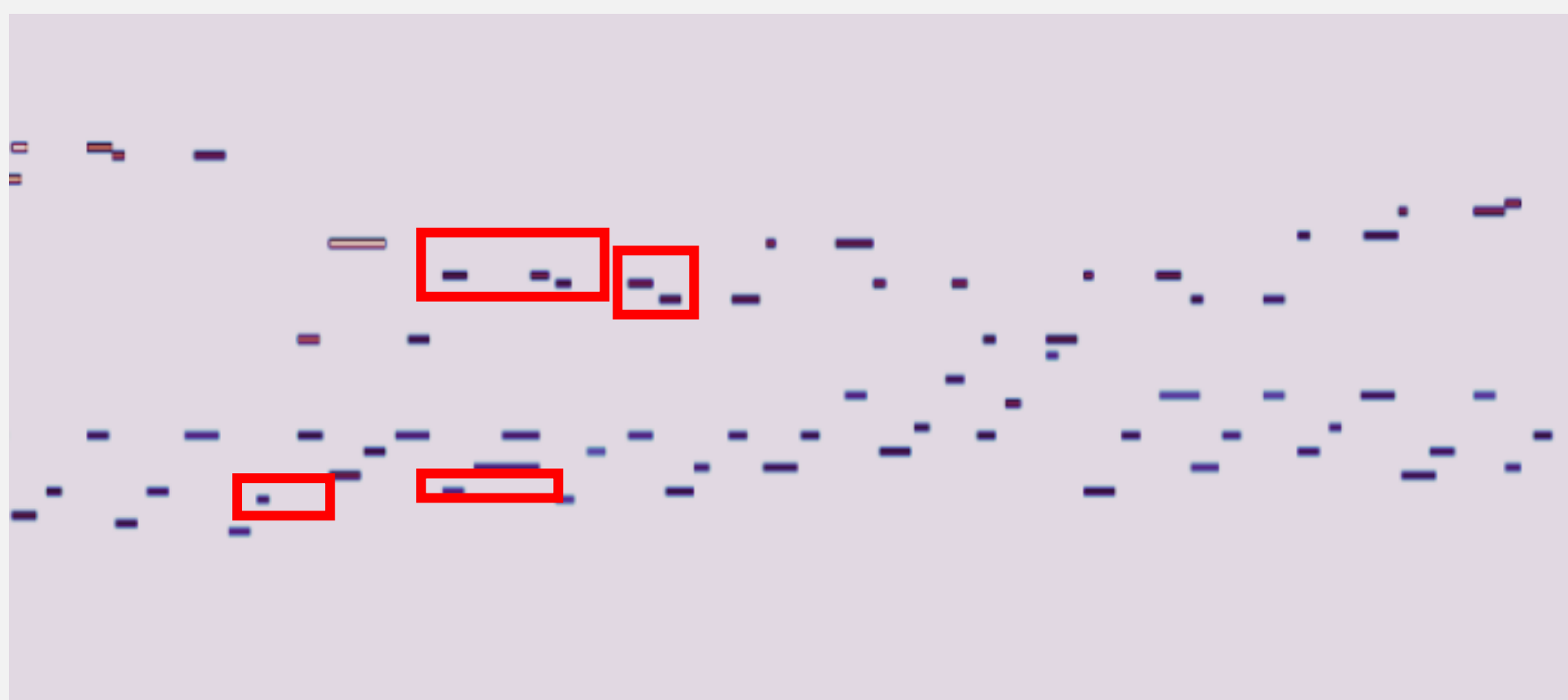
1. *Piano Transcription*: We delve into the high-resolution piano transcription model [1], which converts audio recordings of piano performances into MIDI representations.
2. *Multi-Instrument Transcription*: We investigate the MT3 framework [2] for multi-instrument transcription.
3. *Score-Level Transcription*: We explore the performance MIDI-to-score conversion system in [3].

## Piano Transcription

Ground truth

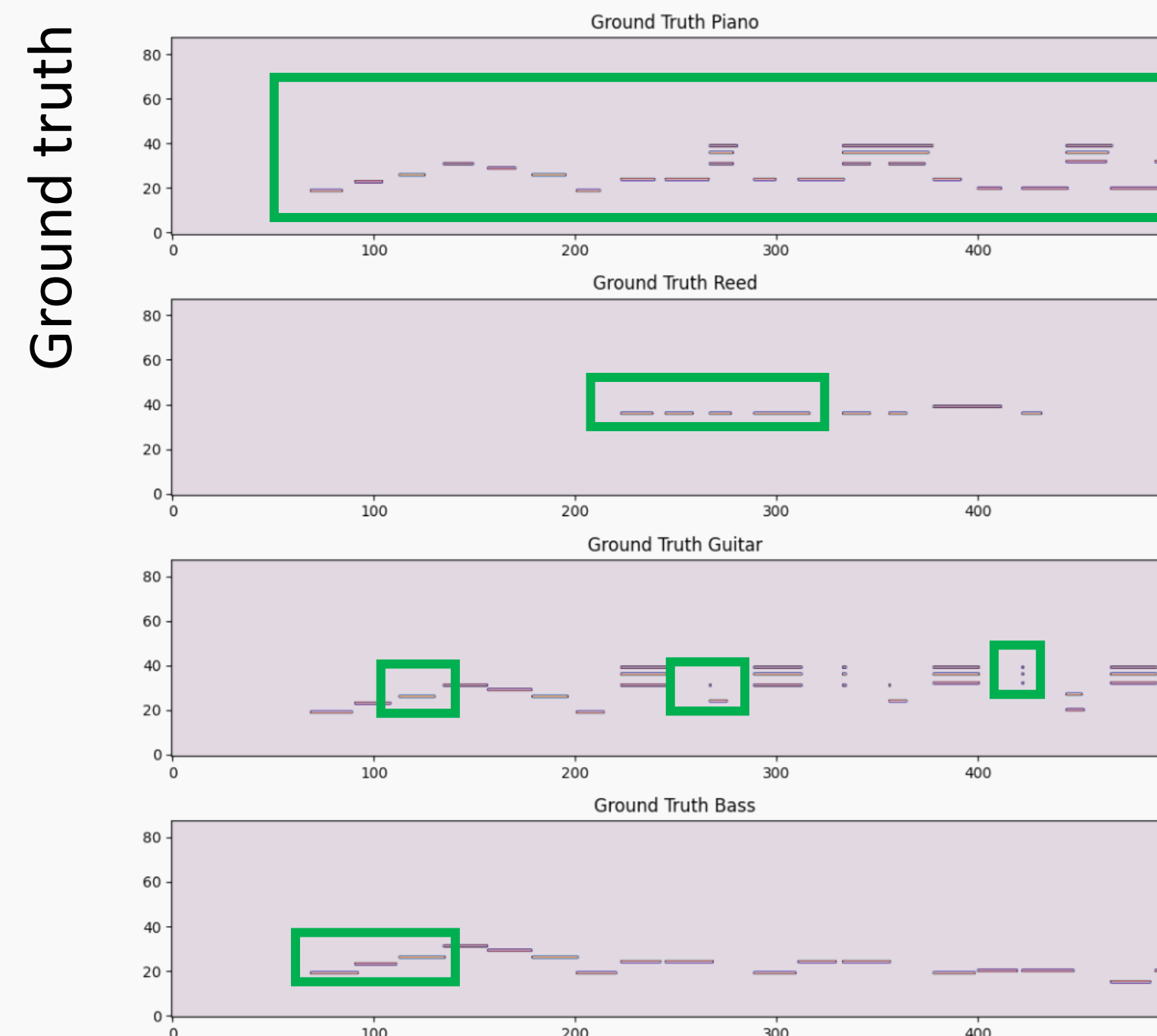


Transcribed

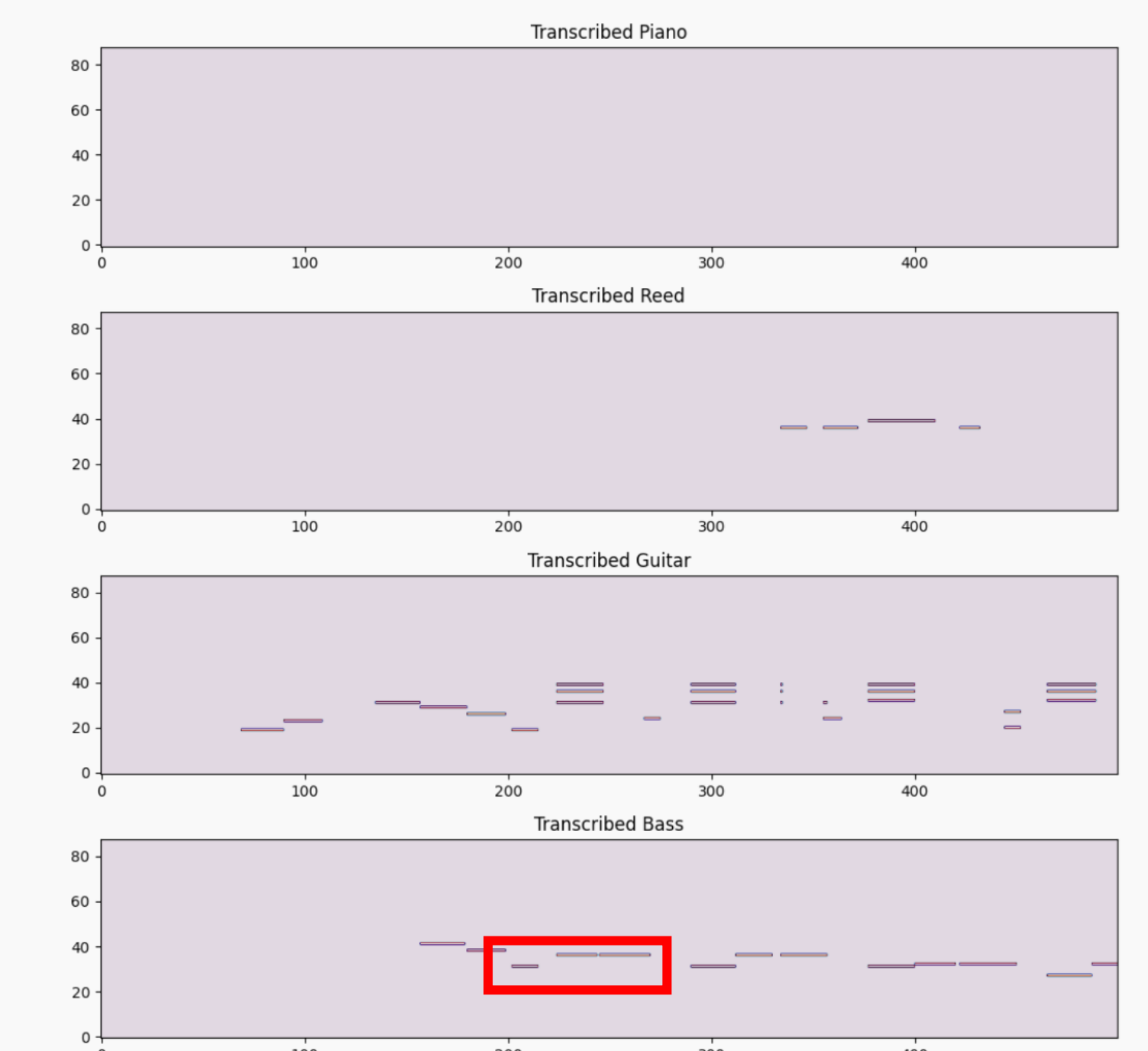


## Multi-Instrument Transcription

Ground truth



Transcribed



## Automatic Music Transcription

## Discussion

Displayed are just a few examples using the three transcription algorithms. More examples, including those using different musical instruments/styles, and listening examples, are available on our demo page.

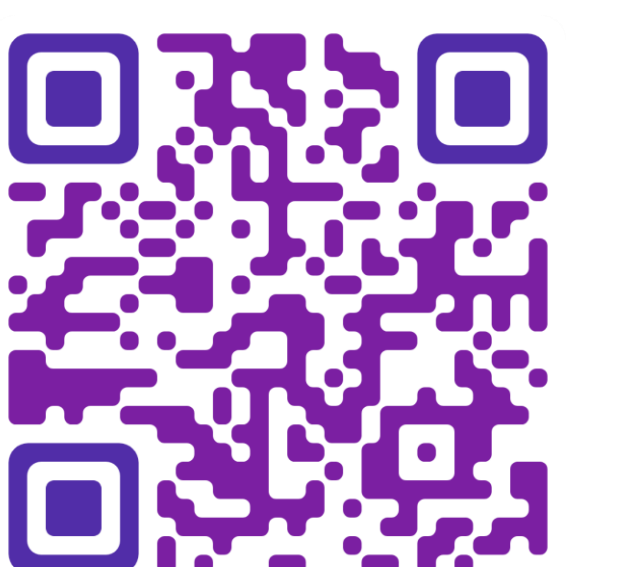
The high-resolution piano transcription algorithm generally preserves the notes and performance details in transcribed files with good quality. Interestingly, although the model is trained with piano recordings, it can also somewhat transcribe guitar performances. One obstacle towards complete transcription is that the note offsets are not transcribed to be the key release, but the time when the piano sustain pedal is released.

The MT3 multi-instrument transcription model tends to have missing notes for specific instruments (especially piano in a mixed recording). When it is exposed to real recordings, the transcription quality drops, which suggests an important future direction for the development of AMT algorithms: System development and evaluation for real music performances.

The score transcription system, on the other hand, tends to make errors when predicting time signatures and metrical structures. It presents better transcription quality when there is no syncopation in the rhythmic pattern. It also presents better results when there is a clear beat indicated by the bass voice.

Overall, there are still challenges in the development of AMT methods, including Dealing with multiple instruments, rhythmic structures, and further expressive performance annotations. Moreover, better evaluation methods are needed in order to avoid bias introduced by centrally distributed datasets.

[LINK TO DEMO](#)

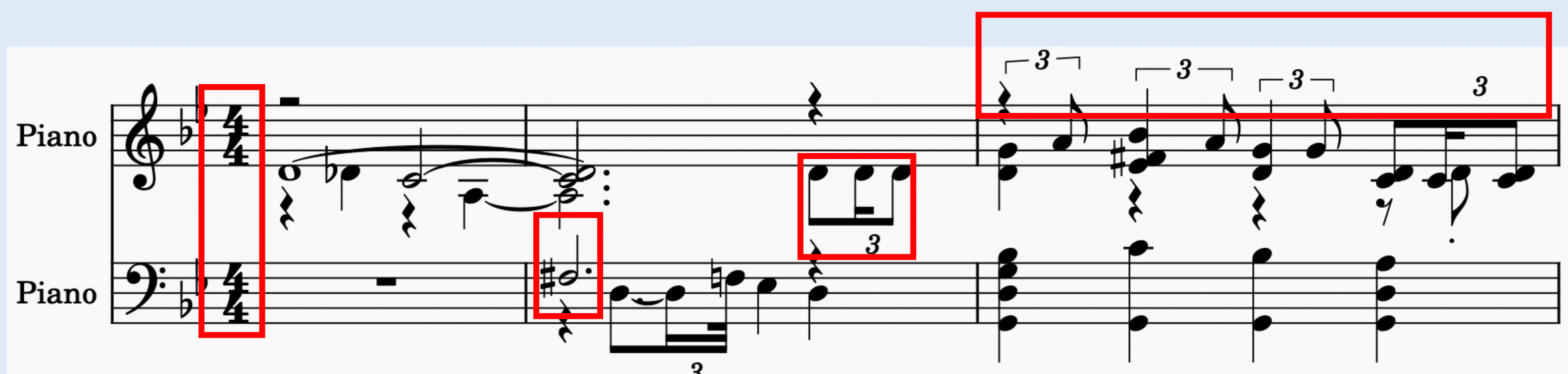


## Score-Level Transcription

Ground truth



Transcribed



## References

- [1] Q. Kong et al., High-resolution piano transcription with pedals by regressing onset and offset times. *IEEE/ACM TASLP*, 29:3707-3717, 2021.
- [2] J. P. Gardner et al., MT3: Multi-task multitrack music transcription. In *ICLR*, 2022.
- [3] L. Liu et al., Performance MIDI-to-score conversion by neural beat tracking. In *ISMIR Conference*, pp. 395-402, 2022.