

Automatic Music Transcription

Дмитрий Протасов

Научный руководитель: Иван Матвеев
МФТИ

16 декабря 2023

Постановка задачи

Проблема

Генеративные музыкальные модели довольно удобно строить в пространстве MIDI-файлов. Проблема – нет большого количества таких MIDI-датасетов, для большинства песен в интернете есть только их аудиоформат. Эту проблему предлагается решать алгоритмом преобразования аудио-представления песен в её MIDI-представление.

Цель

Исследование и улучшение существующих алгоритмов извлечения MIDI из песен

Задачи работы

- Собрать базу данных песен и их MIDI-представлений, сгенерировать синтетические датасеты
- Изучить и протестировать существующие модели, понять их главные недостатки
- Реализовать свои методы извлечения MIDI из аудио

Постановка задачи

Сама задача распознавания нот делится на три этапа

- Разделение на отдельные инструментальные дорожки (Music-Source-Separation)
- Распознавание инструмента (Instrument-Recognition)
- Транскрибация в ноты (Note-Transcription)

Рассмотрим основные работы, посвященные одному или нескольким из этих этапов



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.

Обзор литературы: Transcription

MT3 (2022) [link]

SOTA в Multi-instrument, основана на модели T5, учится end-to-end

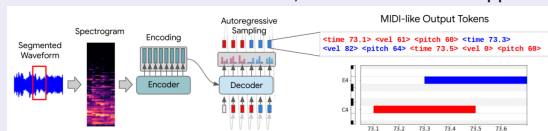


Figure 2: Tokenization/detokenization, as described in Section 3.2. MIDI data (left, represented here as a multitrack "pianoroll") can be tokenized into MIDI-like target tokens for training (right). Output tokens using the same vocabulary can be deterministically decoded back into MIDI data.

Jointist (2022) [link]

Учатся отдельно блоки MSS, Instrument Recognition, Note Transcription

Crepe (2018) [link]

Находит фундаментальную частоту по аудио. Может быть полезно для извлечения нот из вокала

Обзор литературы: Music Source Separation

Benchmarks and leaderboards for sound demixing tasks [\[link\]](#)

Demucs [\[github\]](#)

Based on a U-Net convolutional architecture

MDX-Net [\[github\]](#)

MUSDB18 Dataset [\[link\]](#)

150 music tracks (10h duration) with isolated drums, bass, vocals, others

Проведенные на данный момент эксперименты

Мелспектрограммы инструментальных дорожек, полученные через нейросеть demucs, а также выделение фундаментальной частоты через встроенный метод в librosa, а также просто выделение частоты с максимальной энергией

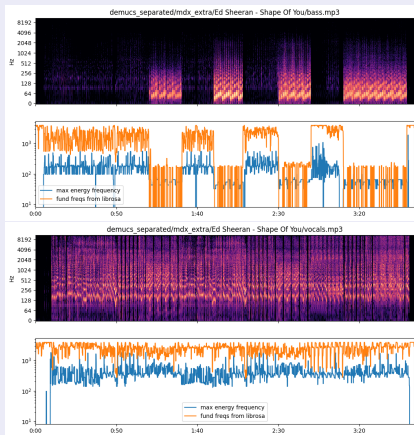


Рис.: Bass (сверху), Vocal (снизу)

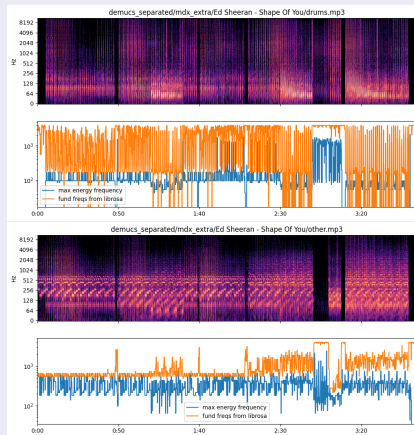


Рис.: Drums (сверху), Other (снизу)

Планы на будущее

- 1 Сделать эксперименты по замеру качества существующих методов
- 2 Собрать базу данных песен и их MIDI-представлений, сгенерировать свои датасеты (возможно, используя языковую MIDI-модель)
- 3 Придумать и реализовать более объективную меру качества, более удобную токенизацию MIDI, лосс при обучении
- 4 Гипотеза: для разных инструментов надо использовать свои различные модели, методы
- 5 Реализовать и протестировать несколько своих методов извлечения MIDI из аудио

