BRNO UNIVERSITY OF TECHNOLOGY FACULTY OF INFORMATION TECHNOLOGY

Excel@FIT 2025

http://excel.fit.vutbr.cz

Generation of guitar tabs from signal

Tobiáš Adamčík*

Abstract

The work addresses the problem of Automatic Guitar Transcription (AGT), with a particular focus on non-standard (alternative) tunings. The problem is approached by using virtual instruments to create a large, labeled dataset and employing convolutional neural networks for the transcription task. This work is among the first AGT tools designed to support non-standard tunings. The results (from a guitarist's) perspective range from accurate transcription to almost unusable, depending on the techniques used during the performance.

*xadamc08@vut.cz, *Faculty of Information Technology, Brno University of Technology*

1. Introduction

Accurate guitar notation is invaluable to musicians at all skill levels. While beginners benefit from clear and precise guidance, experienced players also rely on accurate transcription to document and share their work. However, most modern tools for guitar transcription, such as TabCNN [1] or TabInception [2] focus on standard-tuned guitars, leaving players of genres such as hardcore or metal, where alternative tunings are commonly used, behind.

The task itself is as follows: produce accurate guitar notation from Direct Input recordings of an electric guitar with the support for a variety of tunings. A good solution should detect all notes played with correct timing and accurately assign the correct string and fret a note has been played on.

Existing solutions for this specific problem have not been found. The closest similar projects include instrument-agnostic solutions for transcription such as Basic Pitch [3] and tools created specifically for guitars in standard tuning, such as TabCNN [1].

The proposed solution is a multi-task convolutional neural network detecting the note onsets, note presence, and string activity in guitar recordings represented by constant-Q transform spectrograms. The outputs of the network are post-processed to create guitar tablature. The neural network is trained on a dataset created via synthesis of guitar tablature utilizing MIDI virtual guitars.

This work is among the first AGT tools supporting various non-standard tunings. The accuracy of the produced transcription ranges from great with minor mistakes to inaccurate, depending mostly on the techniques used.

2. Data

The main dataset used for training the network has been created via the synthesis of existing guitar tablature accessed from Ultimate Guitar. It consists of songs mainly from heavier genres of music, such as metalcore, hardcore and metal. Each tablature was converted into the `.jams` format. The `.jams` files are then transposed across 16 different tunings and saved. Corresponding `.midi` files are created, from which audio recordings are synthesized using the ODIN III virtual guitar by Solemn Tones, utilizing various presets in order to combat overfitting. The resulting dataset thus contains pairs of `.jams` and `.wav` files for each song in 16 different tunings. In total, the synthesized dataset includes 755 pairs of `.jams` and `.wav` files containing more than 45 hours of audio.

3. Proposed System

The proposed system consists of three main stages:

1. **Data Pre-processing:** Audio files are converted into frame-wise Constant-Q Transform (CQT) spectrograms. Corresponding annotations are extracted from the `.jams` files.
2. **Model:** The core of the system is a Convolutional Neural Network (CNN) with multi-task

learning, performing three interconnected tasks:

- **Tone Identification:** The spectrogram is processed through three convolutional layers, each followed by max-pooling and dropout, and then flattened. A final linear layer maps the output to 65 distinct tones. This branch serves both as an independent output and as input to the next two tasks.
- **Note Onset Detection:** This branch detects whether a note onset occurs in a given frame. It starts with a convolutional layer applied to the original spectrogram. The result is concatenated with the tone identification output and passed through another convolutional layer, followed by a linear layer mapping to 65 possible tone onsets.
- **String Activity Detection:** This branch detects which strings are active in a given frame. It takes as input the original spectrogram and the embedded tuning representation. After an initial convolutional layer, its output is concatenated with the tone identification output. Another convolutional layer and a final linear layer predict the presence of each of the 8 strings.

3. **Post-processing:** The model outputs (note presence, note onset, and string activity) are combined and converted into guitar tablature through several steps:
   First, onsets are added to longer portions of tone activity that do not begin with an onset. Then, the onsets combined with tone activity are converted into notes. Notes shorter than a given threshold are removed.
   Non-overlapping notes are assigned to strings by selecting the string with the highest activity within the time frame of the note. Overlapping notes are grouped into clusters, and these overlaps are resolved based on string activity and playability constraints using a recursive algorithm.

4. Results

The results are gathered from both the evaluation on the training portion of the synthesized dataset and by evaluation on the author's guitar recordings.

The evaluation on the synthesized dataset is performed by measuring the precision, recall, and subsequent F-measure of each task of the network and combining them into a single accuracy score.

The evaluation on real-life guitar recordings is carried out by comparing the post-processed, note-level predictions of the system with ground truth annotations. The predictions are correlated with the annotations and shifted in time to achieve alignment. The presence of notes is then compared with a tolerance in the range of tens of milliseconds.

Recall, precision, and F-measure values are collected. The results of the evaluation on real-life recordings range from an F-measure of 0.45 to an F-measure of 0.93, depending on the specific recording.

5. Conclusions

The results of the project appear promising, but further work is necessary to develop it into a valuable guitar transcription tool: the creation of a more diverse dataset, enabling the network to fine-tune itself based on recordings of a specific guitar, and making adjustments to the network architecture itself.

Acknowledgements

I would like to thank my supervisor prof. Dr. Ing. Jan Černocký for his guidance and help during the creation of this project. I would also like to thank my friend Bc. Jakub Sychra for sharing his experience during the creation of his own thesis on a similar subject.

References

[1] Andrew Wiggins and Youngmoo Kim. Guitar tablature estimation with a convolutional neural network. In *Proceedings of the 20th International Society for Music Information Retrieval Conference*, pages 284–291. ISMIR, November 2019.

[2] Charbel El Achkar, Talar Atechian, Raphaël Couturier, and Abdallah Makhoul. Leveraging computer vision networks for guitar tablature transcription. pages 3–15, 01 2024.

[3] Rachel M. Bittner, Juan José Bosch, David Rubinstein, Gabriel Meseguer-Brocal, and Sebastian Ewert. A lightweight instrument-agnostic model for polyphonic note transcription and multipitch estimation, 2022.