MIREX SUBMISSIONS FOR CHORD RECOGNITION AND KEY ESTIMATION 2017

Filip Korzeniowski, Sebastian Böck, Florian Krebs and Gerhard Widmer
Department of Computational Perception
Johannes Kepler University, Linz, Austria
filip.korzeniowski@jku.at

ABSTRACT

This abstract describes our chord recognition and key estimation submissions for MIREX 2017. Implementations for the algorithms are part of the *madmom* audio processing framework [1] available at https://github.com/CPJKU/madmom.

1. CHORD RECOGNITION

Our submissions for the chord recognition task follow closely the descriptions in [3, 4], with some minor modifications and additions, which we outline in the following. All submissions predict major and minor chords (and the "no chord" class) only. Additionally, we align the chord segments to beats detected by the `DBNBeatTracker` algorithm provided in *madmom*, which is in turn based on [2] and [6]. The beat alignment is the only difference from our 2016 submissions.

1.1 Deep Chroma Chord Recognition (KBK1)

This program uses a deep neural network to extract chroma vectors as shown in [3]. We simplified the chroma extraction network: it only uses 3 layers of 256 hidden units, and operates on a frequency range of 60 - 2100 Hz.

The chord sequence is then decoded using a linear-chain Conditional Random Field (CRF) as described in [4].

1.2 CNN-CRF Chord Recognition (KBK2)

This program uses features learned automatically by a fully convolutional neural network as described in [4]. The architecture differs slightly from the one presented in the paper: instead of padding the feature maps in the first four convolutional layers, we increased the input size accordingly, such that the feature maps after the fourth layer have the same size as in [4]. The input spectrogram thus comprises a frequency range of 60 - 2600 Hz.

The chord sequence is then decoded using a linear-chain Conditional Random Field (CRF) as described in [4].

1.3 Training

The neural networks and conditional random fields were trained on the following data sets: Beatles, Queen and Zweieck [1], Robbie Williams [2], RWC Popular [3], and the public part of McGill Billboard [4].

2. KEY ESTIMATION (FK1)

The key estimation model is based on [5]. It is a convolutional neural network with 5 convolutional layers, followed by a frame-wise dense projection which is averaged over time. Finally, a softmax classification layer determines the global key of a piece.

2.1 Training

The key estimation model was trained using 1077 pieces from the GiantSteps MTG Key dataset [5] and an in-house dataset of 1751 (mostly piano) classical music pieces. Note that the GiantSteps MTG Key dataset is *distinct* from the GiantSteps Key dataset that is used for evaluation at MIREX.

3. ACKNOWLEDGEMENTS

This work is supported by the European Research Council (ERC) under the EU's Horizon 2020 Framework Programme (ERC Grant Agreement number 670035, project "Con Espressione"). The Tesla K40 used for this research was donated by the NVIDIA Corporation.

4. REFERENCES

[1] Sebastian Böck, Filip Korzeniowski, Jan Schlüter, Florian Krebs, and Gerhard Widmer. madmom: a new Python Audio and Music Signal Processing Library. `arXiv:1605.07008`, 2016.

[2] Sebastian Böck, Florian Krebs, and Gerhard Widmer. A multi-model approach to beat tracking considering heterogeneous music styles. In *Proceedings of the 15th International Society for Music Information Retrieval Conference (ISMIR)*, Taipei, Taiwan, 2014.

[1] http://isophonics.net/datasets

[2] https://www.researchgate.net/publication/260399240_Chord_and_Harmony_annotations_of_the_first_five_albums_by_Robbie_Williams

[3] https://staff.aist.go.jp/m.goto/RWC-MDB/ and https://github.com/tmc323/Chord-Annotations

[4] http://ddmal.music.mcgill.ca/billboard

[5] https://github.com/GiantSteps/giantsteps-mtg-key-dataset

[3] Filip Korzeniowski and Gerhard Widmer. Feature Learning for Chord Recognition: The Deep Chroma Extractor. In *Proceedings of the 17th International Society for Music Information Retrieval Conference (ISMIR)*, New York, New York, USA, August 2016.

[4] Filip Korzeniowski and Gerhard Widmer. A Fully Convolutional Deep Auditory Model for Musical Chord Recognition. In *Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing*, Salerno, Italy, September 2016.

[5] Filip Korzeniowski and Gerhard Widmer. End-to-End Musical Key Estimation Using a Convolutional Neural Network. In *Proceedings of the 25th European Signal Processing Conference (EUSIPCO)*, Kos, Greece, August 2017.

[6] Florian Krebs and Sebastian Böck and Gerhard Widmer. An efficient state-space model for joint tempo and meter tracking. In *Proceedings of the 16th International Society for Music Information Retrieval Conference (ISMIR)*, Mlaga, Spain, 2015.