MIREX 2017
CNN-BASED TEMPO ESTIMATION

Hendrik Schreiber
tagtraum industries incorporated
hs@tagtraum.com

ABSTRACT

This extended abstract describes CNN-based tempo estimation submissions for MIREX 2018.

1. DESCRIPTION

We have submitted three CNN-based systems. Submission 1 is described in detail in [8]. Submission 2 is identical to 1, but has been trained on a larger, more diverse training set consisting of the datasets *LMD-Tempo* [8], *MTG-Tempo* [8], *Extended Ballroom* [4], *ACM MIRUM* [5], *ISMIR 2014 Songs* [1], *GTzan* [9], a subset of tracks from the *Weimar Jazz Database* [6], a subset of *SMC* [3], and finally a subset of *Hainsworth* [2]. Submission 3 has been trained on the same dataset as 2, but uses a different network architecture: shorter filters, dilated convolutions, residual connections, and global average pooling in the final classification layer (see Figure 1). It is thus a *fully convolutional network* (FCN). The same network architecture has also been used as a basis for transfer learning folk music tempo and meter in [7].

2. CODE

Code for running our submissions is available at GitHub: `https://github.com/hendriks73/tempo-cnn`.

3. REFERENCES

[1] Fabien Gouyon, Anssi P. Klapuri, Simon Dixon, Miguel Alonso, George Tzanetakis, Christian Uhle, and Pedro Cano. An experimental comparison of audio tempo induction algorithms. *IEEE Transactions on Audio, Speech, and Language Processing*, 14(5):1832–1844, 2006.

[2] Stephen Webley Hainsworth. *Techniques for the Automated Analysis of Musical Audio*. PhD thesis, University of Cambridge, UK, September 2004.

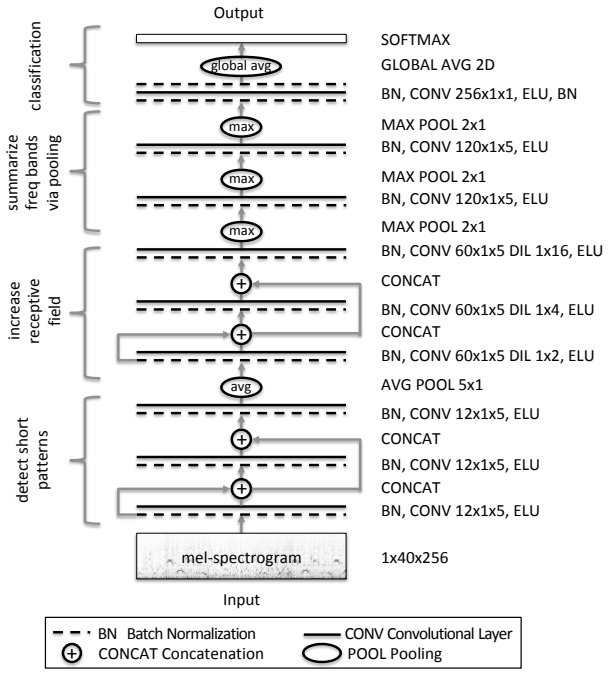

Figure 1. Schematic overview of the network architecture.

[3] Andre Holzapfel, Matthew EP Davies, José R Zapata, João Lobato Oliveira, and Fabien Gouyon. Selective sampling for beat tracking evaluation. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(9):2539–2548, 2012.

[4] Ugo Marchand and Geoffroy Peeters. The extended ballroom dataset. In *Late Breaking Demo of the International Conference on Music Information Retrieval (ISMIR)*, New York, NY, USA, 2016.

[5] Geoffroy Peeters and Joachim Flocon-Cholet. Perceptual tempo estimation using GMM-regression. In *Proceedings of the second international ACM workshop on Music information retrieval with user-centered and multimodal strategies (MIRUM)*, pages 45–50, New York, NY, USA, 2012. ACM.

[6] Martin Pfleiderer, Klaus Frieler, Jakob Abeßer, Wolf-Georg Zaddach, and Benjamin Burkhart, editors. *Inside the Jazzomat - New Perspectives for Jazz Research*. Schott Campus, 2017.

[7] Hendrik Schreiber. Technical report: Tempo and me-

ter estimation for greek folk music using convolutional neural networks and transfer learning. In *International Workshop on Folk Music Analysis (FMA), Tempo/Meter Challenge*, Thessaloniki, Greece, June 2018.

[8] Hendrik Schreiber and Meinard Müller. A single-step approach to musical tempo estimation using a convolutional neural network. In *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR)*, Paris, France, September 2018.

[9] George Tzanetakis and Perry Cook. Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, 10(5):293–302, 2002.