SMOOTHED FRAMEWISE PIANO TRANSCRIPTION

Rainer Kelz
Department of Computational Perception,
Johannes Kepler University Linz, Austria
rainer.kelz@jku.at

Sebastian Böck
Department of Computational Perception,
Johannes Kepler University Linz, Austria
sebastian.boeck@jku.at

ABSTRACT

In [5], framewise piano transcription systems based on convolutional neural nets were trained and evaluated on the MAPS dataset, described in [4]. This submission is intended as a field test of the best performing model reported in the aforementioned paper. Due to the per-note nature of the evaluation of the piano transcription sub-task, we added a simple smoothing and note extraction component.

1. TRANSCRIPTION SYSTEM

The framewise piano transcription system is described in [5], table 4 in the column "ConvNet".

2. SMOOTHING COMPONENT

We are postprocessing the framewise output of the transcription system with a maximum filter of length 2 in time.

3. NOTE EXTRACTION

After smoothing, we extract note onsets by checking if the output of the last network-layer is above a certain threshold θ_{ON}. After a small hysteresis period we detect note offsets by checking if the output is falling below a certain θ_{OFF}.

4. SOFTWARE STACK

To make our lives easier we use Theano [1] and Lasagne [3] for the convolutional neural networks, and Madmom [2] for audio processing and peak picking.

5. REFERENCES

[1] James Bergstra, Olivier Breuleux, Frédéric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio. Theano: a CPU and GPU Math Expression Compiler. In *Proceedings of the Python for Scientific Computing Conference (SciPy)*, 2010.

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

[3] Sander Dieleman, Jan Schlüter, Colin Raffel, Eben Olson, Søren Kaae Sønderby, Daniel Nouri, Daniel Maturana, Martin Thoma, Eric Battenberg, Jack Kelly, Jeffrey De Fauw, Michael Heilman, diogo149, Brian McFee, Hendrik Weideman, takacsg84, peterderivaz, Jon, instagibbs, Dr. Kashif Rasul, CongLiu, Britefury, and Jonas Degrave. Lasagne: First release., August 2015.

[4] Valentin Emiya, Roland Badeau, and Bertrand David. Multipitch estimation of piano sounds using a new probabilistic spectral smoothness principle. *Audio, Speech, and Language Processing, IEEE Transactions on*, 18(6):1643–1654, 2010.

[5] Rainer Kelz, Matthias Dorfer, Filip Korzeniowski, Arzt Andreas Böck, Sebastian and, and Gerhard Widmer. On the potential of simple framewise approaches to piano transcription. In *ISMIR-11th International Society for Music Information Retrieval Conference*, pages ???–???, 2016.