MIREX SUBMISSION FOR DRUM TRANSCRIPTION 2017

Richard Vogl[1,2] **Matthias Dorfer**[2] **Gerhard Widmer**[2] **Peter Knees**[1]
[1] Institute of Software Technology & Interactive Systems, Vienna University of Technology, Austria
[2] Dept. of Computational Perception, Johannes Kepler University Linz, Austria
{richard.vogl, peter.knees}@tuwien.ac.at

ABSTRACT

This extended abstract describes the algorithms submitted for the 2017 MIREX drum transcription task. The implemented algorithms are included in the current version of the madmom [1] library [1].

1. SUBMISSION

A detailed description of the submitted systems can be found in [2]. For the implementation within the madmom framework some minor modifications were implemented which will be described briefly in the following paragraphs.

The *RV1* submission consists of a convolutional recurrent neural network (CRNN). It is built using convolutional input layers followed by recurrent layers with GRU nodes. Compared to [2] we do not use a first order difference spectrogram as additional input features, since the CNN layers can cover this step if necessary. Additionally, only a frequency range from 30 to 15,000 Hz is used resulting in an input vector of length 79. This reduces the model size and enables the use of valid convolutions which simplifies the implementation in madmom.

RV2 is a convolutional neural network (CNN) consisting of the same convolutional building blocks as the CRNN but followed by dense layers instead of the recurrent layers used in the CRNN. The same input features as for the CRNN are used.

For the *RV3* submission, the large BDRNN network from [2] was used without modifications.

RV4 is an ensemble model using all three network models' outputs and averaging them before peak picking.

2. TRAINING

All models are trained the same way as described in [2]. A four-split cross-validation training on the four subsets of the public training data (2005, MEDLEY, RBMA, and GEN) was performed, resulting in four sub-models for each network. These models are then combined using an averaging ensemble model for the final submission.

[1] https://github.com/CPJKU/madmom

	MX'05	RV1	RV2	RV3	RV4
overall	*0.67*	**0.71**	0.67	0.68	0.70
IDMT	*0.75*	**0.66**	0.66	0.62	0.66
KT	*0.61*	**0.65**	0.63	0.62	0.65
MEDLEY	-	0.73	0.66	0.69	0.70
RBMA	-	**0.72**	0.70	0.69	0.72
GEN	-	**0.78**	0.70	0.76	0.76

Table 1. F-measure results for the evaluated models on the individual data sets of the MIREX'17 drum transcription task as well as mean overall performance (first row).

3. RESULTS

Table 1 shows the f-measure results for the overall evaluation as well as the individual data sets. Task winning results are printed in bold letters. The *RV1* system showed the overall best mean f-measure performance. *RV4*, *RV3*, and *RV2* made second, third, and fourth place respectively. The full result table can be found on the results page of the task on the MIREX'17 website [2] .

4. ACKNOWLEDGEMENTS

This work has been partly funded by the Austrian FFG under the BRIDGE 1 project *SmarterJam* (858514).

5. REFERENCES

[1] Sebastian Böck, Filip Korzeniowski, Jan Schlüter, Florian Krebs, and Gerhard Widmer. madmom: a new python audio and music signal processing library. https://arxiv.org/abs/1605.07008, 2016.

[2] Richard Vogl, Matthias Dorfer, Gerhard Widmer, and Peter Knees. Drum transcription via joint beat and drum modeling using convolutional recurrent neural networks. In *Proceedings of the 18th International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, CN, 2017.

[2] http://www.music-ir.org/mirex/wiki/2017:Drum_Transcription_Results