Master Thesis Proposal

Multi-task Learning of Rhythmic and Tonal Properties in Music

Peter Rjabcsenko, 01228563

Advisor: Assistant Prof. Dipl.-Ing. Dr.techn. Peter Knees

Assistance: Projektass. Dipl.-Ing. Dr.techn. Richard Vogl Bakk.techn.

February 06, 2021

1 Motivation & Problem Statement

Artificial Intelligence is rapidly becoming the dominant approach for solving a variety of multimedia related tasks, be it image classification, text generation, speech recognition or synthesis. The domain of music is no exception with traditional handcrafted signal processing methods slowly being phased out in favour of those heavily based on machine learning, in particular, methods involving deep learning have garnered considerable attention in recent years [3].

Deep learning solutions for chord recognition, utilizing convolutional neural networks [5], and for audio beat tracking, using recurrent neural networks [2] and dilated convolutional neural networks [3] have achieved state of the art results, surpassing their predecessors by a significant margin.

Additionally many music analysis problems overlap in one way or another and it was shown by Sebastian Böck, Matthew Davies and Peter Knees that for example creating a joint model for the task of beat tracking and tempo estimation can have benefits for each individual task. By training the model on both beat and tempo information, evaluating it and then providing additional tempo only information for further training, the beat tracking capabilities of the model were shown to have also further improved [1]. This has two important implications: first of all jointly learning musical

properties can have a positive effect on the outcome of each individual task but also a purely practical benefit is that while tempo information is relatively cheap to produce and is more widely available, high quality beat information is much harder to obtain and by being able to train the model on large amounts of the former it is possible to somewhat mitigate the scarcity of the latter.

In a similar fashion learning a joint model for beat tracking and chord recognition should yield improved results for both. Moreover since high quality annotations for both these tasks are expensive to produce, a joint model could greatly benefit from having the ability to work with either kind of annotations.

2 Expected Results

The aim of this work is to investigate the potential of a deep learning architecture for the joint task of beat tracking and chord recognition.

The proposed model will be first trained and evaluated on beat and chord information separately to demonstrate its feasibility, then a similar experiment will be conducted with both beat and chord information present in order to investigate whether there is any benefit from multi-task learning in this case and finally the model will be further trained on beat information only, to evaluate if its chord recognition capability has improved, and the other way around, it will be further trained on chord information only, to evaluate if the models beat tracking capability has improved.

3 Methodological Approach

- Literature review:

Will be roughly divided into three parts.

Information must be gathered and reviewed about current deep learning architectures for audio beat tracking, the same will be done for chord recognition and finally information on existing approaches for multi-task learning in the music domain have to be gathered and analyzed.

- Model:

Three deep learning models will be proposed.

Two models approximating the state of the art solutions for beat tracking and chord recognition and a third model for solving the joint problem.

The models will follow a common architectural pattern used for modern convolutional neural networks: a sequence of several alternating convolutional and max-pooling layers followed by one or more fully connected layers and a non-linear activation function [5]. In case of beat tracking additional dilated convolutional layers will be introduced between the convolutional and fully connected layers to better capture temporal relations in the data [3]. The input audio signal will be pre-processed using Short-time Fourier transform (STFT) resulting in time-frequency domain features that are well suited as input features for the described models.

- Implementation:

The models described above will be implemented as standalone solutions for their respective tasks with the help of PyTorch, a machine learning library for the Python language. Data preprocessing and the evaluation of results will also be performed in Python.

The input data will be split into training, validation and test sets, where the validation set will be used for intermediate testing during the iterative training process of the model and the test set will only be used after the training is finished to ensure that the final testing of a model only happens on previously unseen data. The discrete Fourier transform needed for the STFT will be computed with the Fast Fourier transform algorithm.

To somewhat reduce the complexity of the chord recognition problem and avoid ambiguities in the input data, as detailed chord classification is not the aim of this work, complex chord annotations will be mapped to a reduced set of thirteen chord classes, one class for each tone of the chromatic scale representing the root note of a chord and a thirteenth class whenever a chord's root is unrecognised.

- Evaluation:

The models will be evaluated using k-fold cross validation and analysis of the results will be performed to determine whether learning a joint model presents an advantage over learning separate models for each task and

whether providing one type of information additionally as training data yields improvement for the task unrelated to this information.

4 State of the Art

With deep learning enjoying a surge in popularity, a variety of problems in the audio and music domain have seen attempts at being tackled with machine learning in mind, many of which have been very successful and produced state of the art solutions. Of particular interest to this work are temporal convolutional networks for beat tracking [3], multi-task approaches for learning rhythmic properties in music [1], joint beat and drum modelling with convolutional recurrent networks [6] and convolutional networks for key classification [4] and chord recognition [5]. These papers present convincing evidence that multi-task learning of rhythmic and tonal properties warrants investigation while also serving as a solid theoretical basis for the task at hand.

5 Relevance to the Curriculum of Logic and Computation (066 931)

This work is closely tied to the module Knowledge Representation and Artificial Intelligence, most notably “184.702 Machine Learning”, “188.501 Similarity Modeling 1” and “188.498 Similarity Modeling 2” as they deal with supervised machine learning, but also “188.502 Media and the Brain 1”, “188.499 Media and the Brain 2” and “188.413 Self-Organizing Systems”.

“194.039 Intelligent Audio and Music Analysis” is also worth mentioning as an additional specialised course dealing with, among other things, supervised learning in the audio domain.

References

[1] S. Böck, M. Davies, P. Knees. (2019). *Multi-Task Learning of Tempo and Beat: Learning One to Improve the Other.* In Proceedings of the 20th International Society for Music Information Retrieval Conference (pp. 486–493). Delft, The Netherlands: ISMIR.

[2] S. Böck, F. Krebs, and G. Widmer. (2016). *Joint beat and downbeat tracking with recurrent neural networks,* in Proc. of the 17th Intl. Society for Music Information Retrieval Conf., pp. 255–261.

[3] M. Davies and S. Böck. (2019). *Temporal convolutional networks for musical audio beat tracking*, 27th European Signal Processing Conference (EUSIPCO), A Coruna, Spain, pp. 1-5, doi: 10.23919/EUSIPCO.2019.8902578.

[4] F. Korzeniowski and G. Widmer. (2018). *Genre-agnostic key classification with convolutional neural networks*, Proc. Int. Soc. for Music Inf. Retr. Conf. (ISMIR), pp. 264-270.

[5] F. Korzeniowski and G. Widmer. (2016). *A fully convolutional deep auditory model for musical chord recognition,* IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), Vietri sul Mare, 2016, pp. 1-6, doi: 10.1109/MLSP.2016.7738895.

[6] R. Vogl, M. Dorfer, G. Widmer, P. Knees. (2017). *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 (pp. 150–157). Suzhou, China: ISMIR.