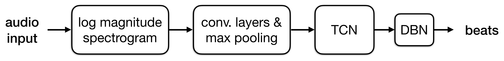
Deconstruct, Analyse, Reconstruct: How to Improve Tempo, Beat, and Downbeat Estimation

Sebastian Böck[1] and Matthew E. P. Davies[2]

[1]enliteAI, Vienna, Austria
[2]University of Coimbra, CISUC, DEI, Portugal

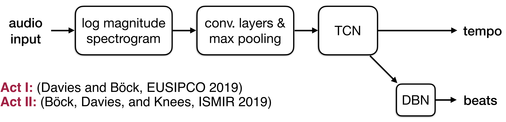
Act I: Beat Tracking with TCNs

Introduce temporal convolutional networks for beat tracking



Act II: Multi-task learning

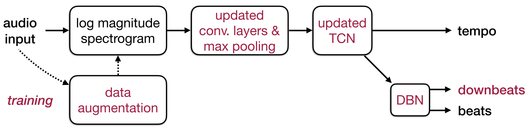
Add a tempo classification layer to the output



Act I: (Davies and Böck, EUSIPCO 2019)
Act II: (Böck, Davies, and Knees, ISMIR 2019)

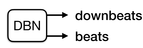
Act III: Deconstruct, Analyse, Reconstruct

Include downbeats as a new multitask learning target
Update the convolution and max pooling layers
Incorporate two dilation rates that are multiples of each other
Incorporate data augmentation when training

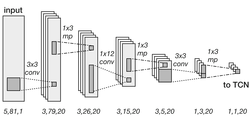


Downbeat multitask learning target

Joint estimation: beats and downbeats
Sequential estimation: beats then downbeats



Updated conv. & max pooling

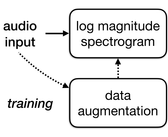


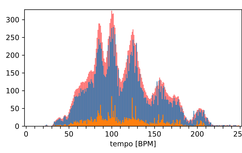
Updated TCN

Introduce a "double dilation" rate to allow longer term temporal dependencies to be captured by the network

Data augmentation

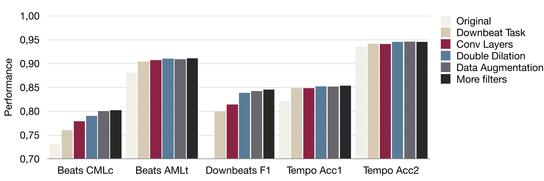
To increase the network's exposure to a wider range of tempi, we generate multiple versions of the log magnitude spectrogram at different hop sizes and adjust the beat, downbeat, and tempo targets accordingly





Ablation Study

To demonstrate the benefit across each task of each introduced modification we present an ablation study



Main Results: Unseen Datasets

Tempo Estimation

	Accuracy 1	Accuracy 2
ACM Mirum		
Gkiokas et al. [50]	0.725	0.979
Percival and Tzanetakis [44]	0.733	0.972
Schreiber and Müller [17]	0.781	0.976
Böck et al. [20]	0.749	0.974
Foroughmand & Peeters [18]	0.733	0.965
Ours	**0.841**	**0.990**
GiantSteps		
Gkiokas et al. [50]	0.721	0.922
Percival and Tzanetakis [44]	0.506	0.956
Schreiber and Müller [17] *	0.821	0.971
Böck et al. [20]	0.764	0.958
Foroughmand & Peeters [18] *	0.836	0.979
Ours	**0.870**	**0.965**
GTZAN		
Gkiokas et al. [50]	0.651	0.931
Percival and Tzanetakis [44]	0.658	0.924
Schreiber and Müller [17]	0.769	0.926
Böck et al. [20]	0.673	0.938
Foroughmand & Peeters [18]	0.697	0.891
Ours	**0.830**	**0.950**

Beat Tracking

	F-measure	CMLt	AMLt
GTZAN			
Böck et al. [5]	0.864	0.768	0.927
Davies and Böck [22]	0.843	0.715	0.914
Ours (beat tracking)	**0.883**	**0.808**	**0.930**
Ours (joint tracking)	**0.885**	**0.813**	**0.931**

Downbeat Tracking

	F-measure	CMLt	AMLt
GTZAN			
Böck et al. [28]	0.640	0.577	0.824
Durand et al. [8]	0.607	0.480	0.774
Ours (sequential tracking)	**0.654**	**0.619**	**0.817**
Ours (joint tracking)	**0.672**	**0.640**	**0.832**

Tempo
Acc 1 > 0.8

Beat
CMLt > 0.8

Downbeat
CMLt > 0.6

Main Findings

We establish a new state of the art in all three tasks, with the most prominent gains coming in totally unseen test datasets.

We observe a promising "closing of the gap" between stricter and weaker evaluation methods indicating our approach is better able to reproduce the metrical level chosen by the annotator.

Funding Acknowledgements

This work is funded by national funds through the FCT - Foundation for Science and Technology, I.P., within the scope of the project CISUC - UID/CEC/00326/2020 and by European Social Fund, through the Regional Operational Program Centro 2020 as well as by Portuguese National Funds through the FCT - Foundation for Science and Technology, I.P. under the project IF/01566/2015.

Cofinanciado por: