Computational Analysis of Sound and Music

Research Project – Tables & Figures

Dr.-Ing. Jakob Abeßer

Fraunhofer IDMT

jakob.abesser@idmt.fraunhofer.de

© Jakob Abeßer, 2024

Tables & Figures

Purpose

- Data visualization
  - present data in a visual format, making complex information easier to understand
- Supporting results
  - provide evidence and support for the results and findings presented in the text
- Enhancing clarity
  - clarify and enhance the interpretation of results by presenting them in a structured and organized manner.
- Comparison and analysis:
  - allow for comparisons between different datasets or experimental conditions

© Jakob Abeßer, 2024

Tables & Figures

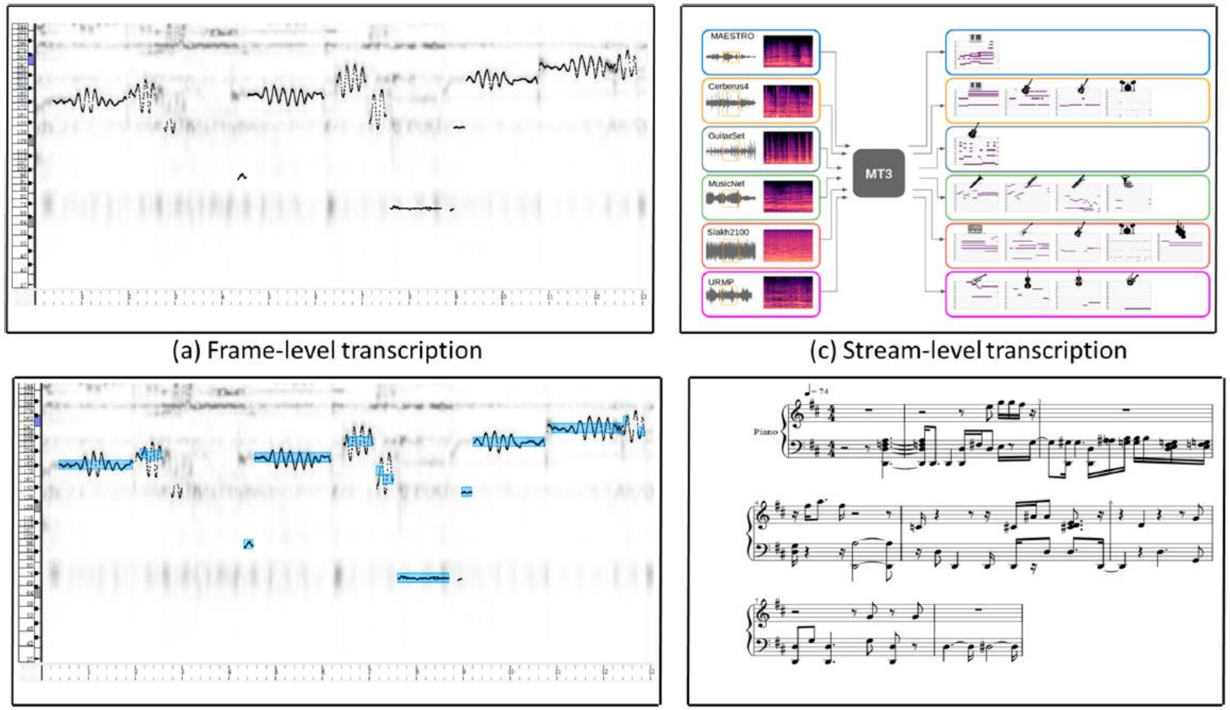
Purpose

- Summarize information:
    - tables summarize large amounts of data concisely
    - figures can illustrate trends, relationships, or distributions
- Reference and replication
    - Figures and tables serve as references for other researchers, enabling them to replicate experiments
- Complementing text
    - Complement the text by providing detailed information that may be cumbersome to explain fully in narrative form
- Highlighting key findings
    - Figures and tables highlight key findings and conclusions of the study, emphasizing important aspects of the research.

Tables & Figures

Examples

- Comparison of related methods



(a) Frame-level transcription (c) Stream-level transcription

(b) Note-level transcription (d) Notation-level transcription

[1]

Figure 1. An example for the illustration of four different levels of music transcription.

© Jakob Abeßer, 2024

Tables & Figures

Examples

- Dataset(s) metadata

TABLE II
OVERVIEW OF THE DATASETS USED FOR TRAINING AND EVALUATION.

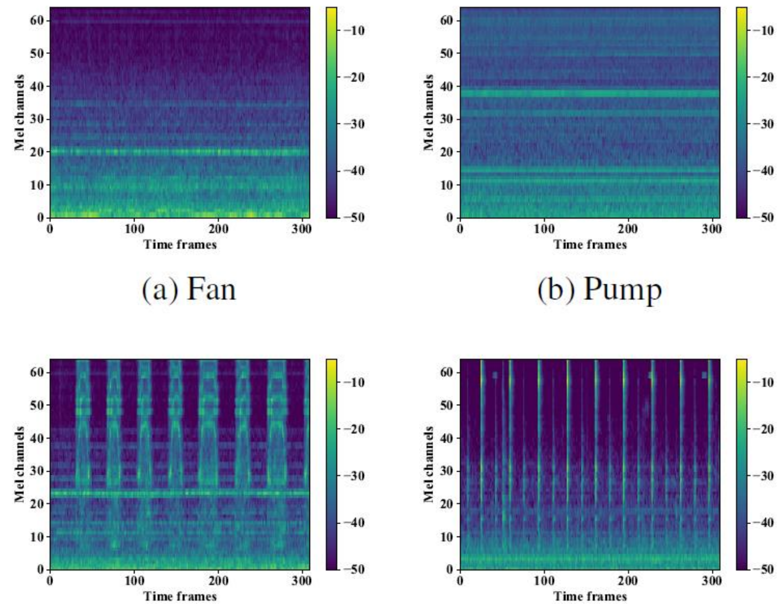
Dataset	# files	length
Ballroom [22], [23] [1]	685	5 h 57 m
Beatles [19]	180	8 h 09 m
Hainsworth [24]	222	3 h 19 m
Simac [25]	595	3 h 18 m
SMC [26]	217	2 h 25 m
GTZAN [20], [21]	999	8 h 20 m

[7]

Tables & Figures

Examples

- Dataset examples (spectrograms)



(a) Fan (b) Pump

(c) Slider (d) Valve

[3]

Fig. 4: Examples of log-Mel spectrograms of the original sound

© Jakob Abeßer, 2024

Tables & Figures

Examples

- Overall system flowchart

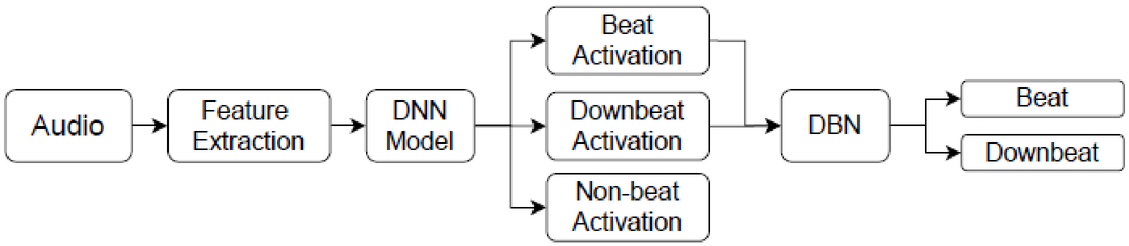


Fig. 1. Pipeline for the beat and downbeat tracking system.

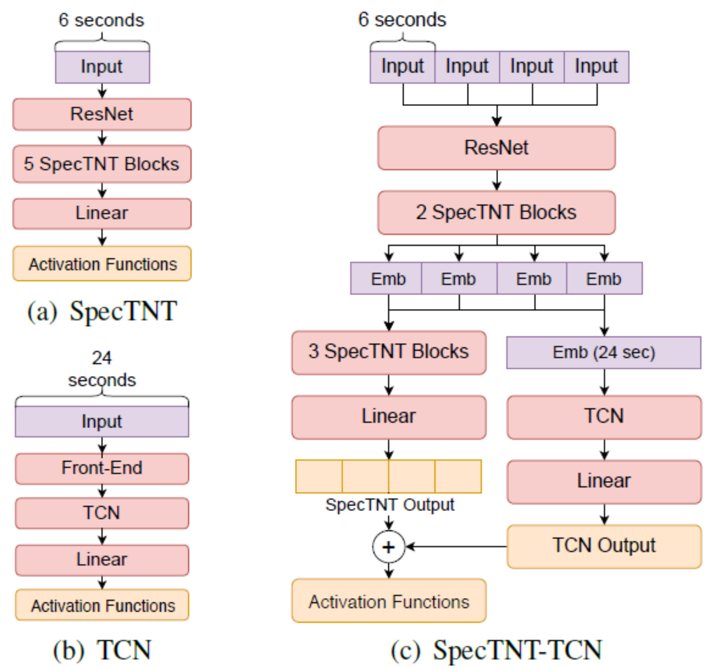
[9]

© Jakob Abeßer, 2024

Tables & Figures

Examples

- DNN architecture comparison (flowcharts)



[10]

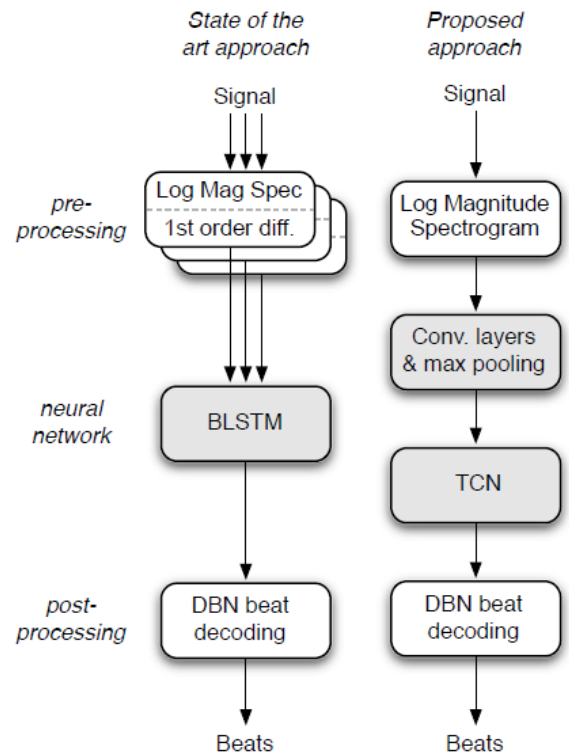
Fig. 2. Model architecture overview.

© Jakob Abeßer, 2024

Tables & Figures

Examples

- DNN architecture (flowchart)



[5]

Fig. 1. Comparison between existing state of the art (left) with our proposed approach (right). The neural network blocks are shaded light grey.

© Jakob Abeßer, 2024

Tables & Figures

Examples

- DNN architecture (table)

Table 1: Modified ResNet architectures

RB Number	RB Config	
	RN1	*RN2*
	Input 5×5 stride=2	
1	3×3, 1×1, P	3×3, 1×1, P
2	3×3, 3×3, P	3×3, 3×3, P
3	3×3, 3×3,	3×3, 3×3
4		3×3, 1×1, P
5	3×3, 1×1, P	1×1, 1×1
6		1×1, 1×1
7		1×1, 1×1
8		1×1, 1×1
9	1×1, 1×1	1×1, 1×1
10		1×1, 1×1
11		1×1, 1×1
12		1×1, 1×1

RB: Residual Block, P: 2×2 max pooling after the block.
RB number 1-4 have 128 channels.
RB number 5-8 have 256 channels.
RB number 9-12 have 512 channels.

[2]

Tables & Figures

Examples

- Performance comparison of 3 models and 4 datasets (1 metric: AUC)

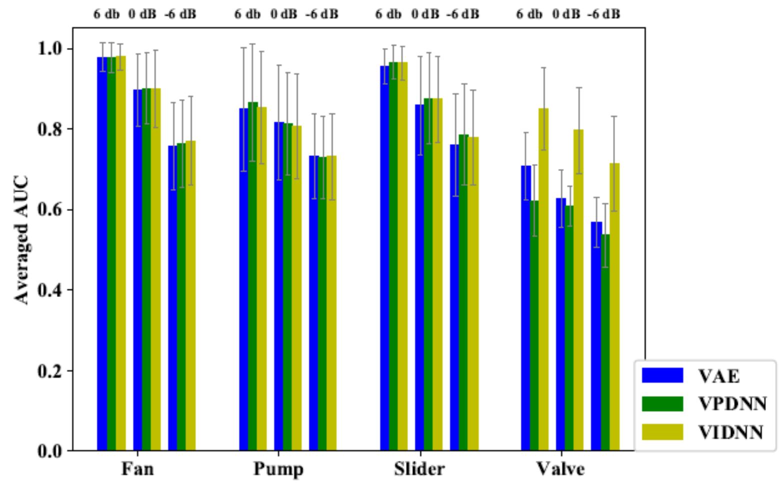


Fig. 6: Averaged AUC of the VAE, VIDNN, and VPDNN

[3]

© Jakob Abeßer, 2024

Tables & Figures

Examples

- List of hyperparameters

TABLE I
OVERVIEW OF SIGNAL PROCESSING AND LEARNING PARAMETERS

Signal Conditioning	
Audio sample rate	44.1 kHz
Window shape	*Hann*
Window & FFT size	2048 *samples*
Hop size	10 ms
Filterbank freq. range	30 . . . 17000 Hz
Sub-bands per octave	12
Total number of bands	81
Conv. Block	
Number of filters	16, 16, 16
Filter size	3×3, 3×3, 1×8
Max. pooling size	1×3, 1×3, —
Dropout rate	0.1
Activation function	*ELU*
TCN	
Number of stacks	1
Dilations	$2^{0,\ldots,10}$
Number of filters	16
Filter size	5
Spatial dropout rate	0.1
Activation function	*ELU*
Training	
Optimizer	*Adam*
Learning rate	0.001
Batch size	1
Output activation function	*sigmoid*
Loss function	*binary cross-entropy*

[6]

© Jakob Abeßer, 2024

Tables & Figures

Examples

- Dataset(s) metadata

TABLE II
OVERVIEW OF THE DATASETS USED FOR TRAINING AND EVALUATION.

Dataset	# files	length
Ballroom [22], [23] [1]	685	5 h 57 m
Beatles [19]	180	8 h 09 m
Hainsworth [24]	222	3 h 19 m
Simac [25]	595	3 h 18 m
SMC [26]	217	2 h 25 m
GTZAN [20], [21]	999	8 h 20 m

[7]

© Jakob Abeßer, 2024

Tables & Figures

Examples

- Performance comparison of 3 models and 4 datasets (multiple metrics)

TABLE III
OVERVIEW OF BEAT TRACKING PERFORMANCE.

	F-measure	CMLc	CMLt	AMLc	AMLt	D
Ballroom						
TCN	0.933	0.864	0.881	0.909	0.929	3.456
BLSTM [5]	0.917	0.832	0.849	0.905	0.926	**3.539**
BLSTM [6]	**0.938**	**0.872**	**0.892**	**0.932**	**0.953**	3.397
Hainsworth						
TCN	0.874	0.755	0.795	**0.882**	**0.930**	**3.518**
BLSTM [5]	**0.884**	**0.769**	**0.808**	0.873	0.916	3.507
BLSTM [6]	0.871	0.732	0.784	0.849	0.910	3.395
SMC						
TCN	**0.543**	**0.315**	**0.432**	**0.462**	**0.632**	**1.574**
BLSTM [5]	0.529	0.296	0.428	0.383	0.567	1.460
BLSTM [6]	0.516	0.307	0.406	0.429	0.575	1.514
GTZAN						
TCN	0.843	0.695	0.715	0.889	0.914	**3.096**
BLSTM [5]	**0.864**	**0.750**	**0.768**	**0.901**	**0.927**	3.071
BLSTM [6]	0.856	0.716	0.744	0.876	0.919	3.019

[8]

© Jakob Abeßer, 2024

References

[1] Bhattarai, B., & Lee, J. (2023). A Comprehensive Review on Music Transcription. Applied Sciences, 13(21), 11882. https://doi.org/10.3390/app132111882, Fig. 1, p. 3

[2] Koutini, K., Eghbal-zadeh, H., & Widmer, G. (2019). CP-JKU submissions to DCASE'19: Acoustic scene classification and audio tagging with receptive-field-regularized CNNs (Technical Report), Tab. 1, p. 3

[3] Suefusa, K., Nishida, T., Purohit, H., Tanabe, R., Endo, T., & Kawaguchi, Y. (2020). Anomalous Sound Detection Based on Interpolation Deep Neural Network. In ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 271-275). Barcelona, Spain. https://doi.org/10.1109/ICASSP40776.2020.9054344, Fig. 4, p. 3

[4] Suefusa, K., Nishida, T., Purohit, H., Tanabe, R., Endo, T., & Kawaguchi, Y. (2020). Anomalous Sound Detection Based on Interpolation Deep Neural Network. In ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 271-275). Barcelona, Spain. https://doi.org/10.1109/ICASSP40776.2020.9054344, Fig. 6, p. 3

[5] Davies, E. P. M., & Böck, S. (2019). Temporal convolutional networks for musical audio beat tracking. In 2019 27th European Signal Processing Conference (EUSIPCO) (pp. 1-5). https://doi.org/10.23919/EUSIPCO.2019.8902578, Fig. 1, p. 2

[6] Davies, E. P. M., & Böck, S. (2019). Temporal convolutional networks for musical audio beat tracking. In 2019 27th European Signal Processing Conference (EUSIPCO) (pp. 1-5). https://doi.org/10.23919/EUSIPCO.2019.8902578, Tab. 1, p. 3

[7] Davies, E. P. M., & Böck, S. (2019). Temporal convolutional networks for musical audio beat tracking. In 2019 27th European Signal Processing Conference (EUSIPCO) (pp. 1-5). https://doi.org/10.23919/EUSIPCO.2019.8902578, Tab. 2, p. 4

© Jakob Abeßer, 2024

References

[8] Davies, E. P. M., & Böck, S. (2019). Temporal convolutional networks for musical audio beat tracking. In 2019 27th European Signal Processing Conference (EUSIPCO) (pp. 1-5). https://doi.org/10.23919/EUSIPCO.2019.8902578, Tab. 3, p. 4

[9] Hung, Y.-N., Wang, J.-C., Song, X., Lu, W.-T., & Won, M. (2022). Modeling beats and downbeats with a time-frequency transformer. In ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 401-405). https://doi.org/10.1109/ICASSP43922.2022.9747048, Fig. 1, p. 2

[10] Hung, Y.-N., Wang, J.-C., Song, X., Lu, W.-T., & Won, M. (2022). Modeling beats and downbeats with a time-frequency transformer. In ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 401-405). https://doi.org/10.1109/ICASSP43922.2022.9747048, Fig. 2, p. 2

© Jakob Abeßer, 2024