

Signal Event List Generation Using Neural Networks

Martin Kovanda

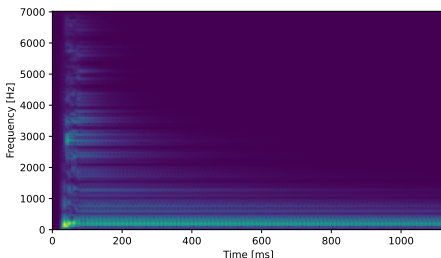
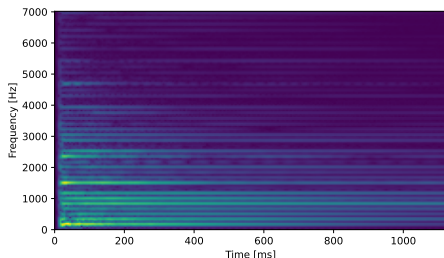
SPMS 2022

June 24, 2022

- 1 Goals of this thesis
 - summarize deep learning methods used for signal processing
 - make a signal source decomposition
 - identify events present in signal
- 2 Signal Decomposition
 - definition
 - dataset
 - experiments
- 3 Event List Generation
 - definition
 - experiments
- 4 Conclusion

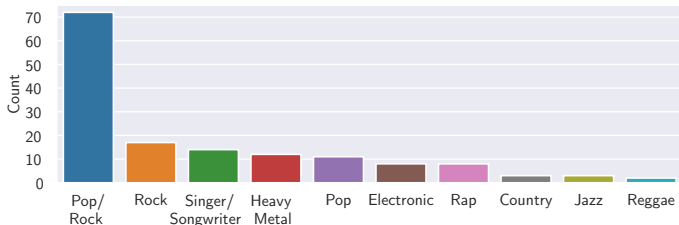
Signal Decomposition Task

- STFT or spectrogram
- L1, L2, AE, MEL-L1, MEL-L2, subjective evaluation

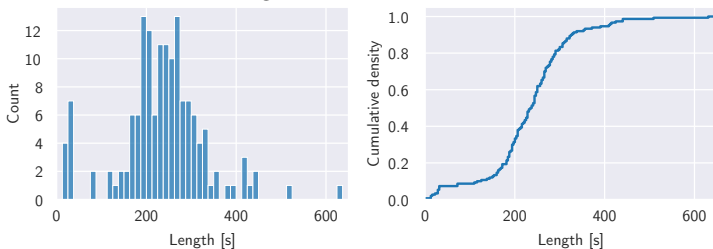


Crop of a spectrogram of a piano E2 tone (left) compared to the bass E2 tone (right).

MUSDB18 Dataset

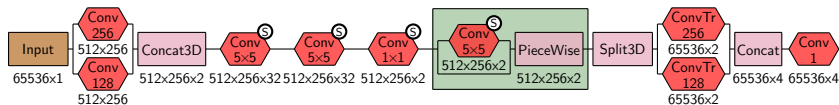


Distribution of genres in the MUSDB18 dataset.



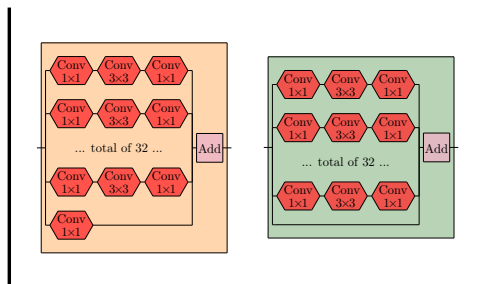
Distribution of song length in the MUSDB18 dataset.

Sepnet + SpecResNeXt



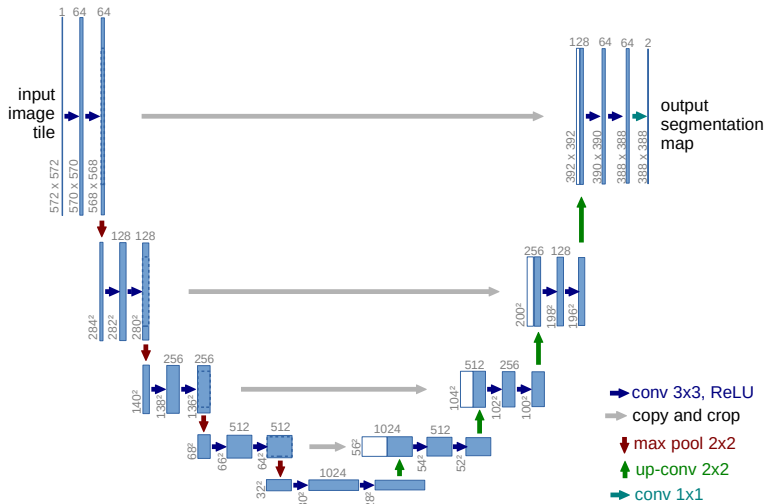
Structure of SepNet. Output shape of each layer is depicted below the corresponding cell. S stands for the SELU activation.

STFT



Structure of SpecResNeXt.

U-Net



The U-Net architecture. Source: Ronneberger et al. [2].

Signal Decomposition Task - Results

architecture	id	source	L1	L2	AE	S-L1	S-L2	MEL-L1	MEL-L2	rating
InceptionTime	1	drums	0.0364	0.0052	0.0136	0.2531	2.7035	0.7997	259.04	5.5/10
InceptionTime	1	bass	0.0282	0.0020	0.0094	0.0997	1.1442	0.5578	60.06	6.0/10
InceptionTime	1	others	0.0353	0.0027	0.0135	0.2687	1.6172	0.6376	27.53	4.0/10
InceptionTime	1	vocals	0.0265	0.0019	0.0106	0.2208	1.3599	0.4705	27.73	4.5/10
SepNet	2	drums	0.0261	0.0024	0.0105	0.2614	1.6462	0.6226	124.88	4.5/10
SepNet	2	bass	0.0253	0.0017	0.0085	0.0868	0.8692	0.4628	43.08	5.0/10
SepNet	2	others	0.0320	0.0023	0.0125	0.2424	1.3373	0.5683	28.06	2.0/10
SepNet	2	vocals	0.0241	0.0017	0.0097	0.1998	1.2421	0.4254	20.03	4.5/10
SpecResNeXt	1	drums	0.0356	0.0061	0.0138	0.2930	4.4668	1.4176	4601.90	4.0/10
SpecResNeXt	1	bass	0.0361	0.0043	0.0122	0.1453	2.9934	1.1031	1515.35	4.5/10
SpecResNeXt	1	others	0.0416	0.0053	0.0160	0.3387	3.6196	1.2203	729.89	4.0/10
SpecResNeXt	1	vocals	0.0364	0.0048	0.0141	0.2851	3.5164	1.1121	2252.26	4.0/10
U-Net	1	drums	0.0236	0.0023	0.0093	0.1761	1.3410	0.5776	426.69	8.5/10
U-Net	1	bass	0.0254	0.0024	0.0084	0.0814	1.3522	0.6004	555.03	9.5/10
U-Net	1	others	0.0312	0.0031	0.0118	0.2132	1.8942	0.7212	242.31	7.5/10
U-Net	1	vocals	0.0199	0.0017	0.0078	0.1510	1.1777	0.4041	114.21	6.5/10

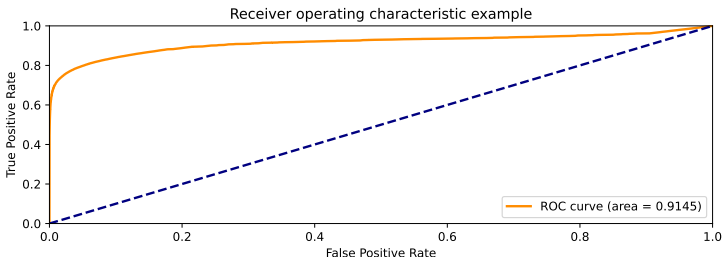
Results of the best architectures on augmented test data for each component.

Event List Generation - Metrics

$$F1 := 2 \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}, \quad (1)$$

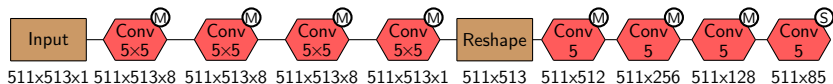
where precision and recall are defined as

$$\text{precision} := \frac{\text{true positives}}{\text{predicted positives}}, \quad \text{recall} := \frac{\text{true positives}}{\text{real positives}}. \quad (2)$$



Example of ROC curve.

Event List Generation - Results



Structure of Spec2Map-big. Output shape of each layer is depicted below the corresponding cell. M stands for Mish activation function.

big version	augmentation	high sr	accuracy	precision	recall	F1 (t=0.5)	roc auc
-1	-1	-1	0.99855	0.76937	0.94451	0.84799	0.99932
1	-1	-1	0.99828	0.72988	0.95656	0.82798	0.99940
-1	1	-1	0.99946	0.94984	0.92290	0.93618	0.99952
1	1	-1	0.99952	0.95359	0.93245	0.94290	0.99974
-1	-1	1	0.99859	0.77263	0.93335	0.84541	0.99835
1	-1	1	0.99841	0.74698	0.94465	0.83426	0.99886
-1	1	1	0.99942	0.95550	0.90165	0.92779	0.99864
1	1	1	0.99949	0.96322	0.91196	0.93689	0.99911

Results of the Spec2Map architecture on augmented test data.

Event List Generation - Acoustic Emission

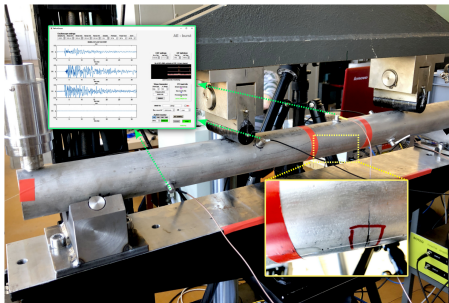
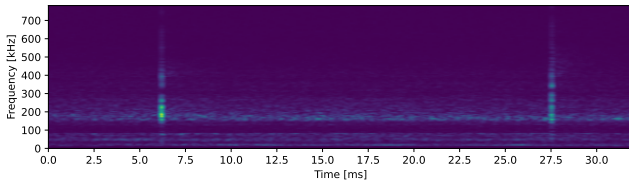
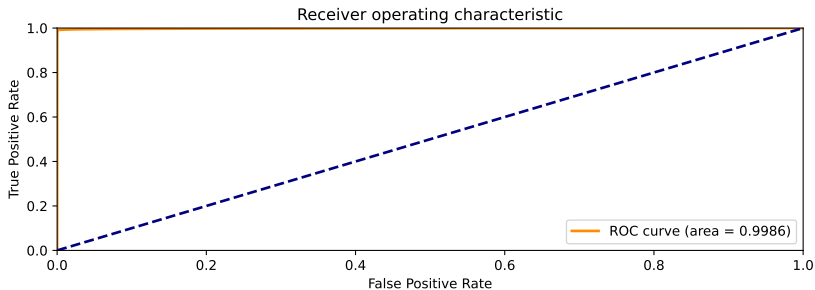


Diagram of tensile testing machine experiment.

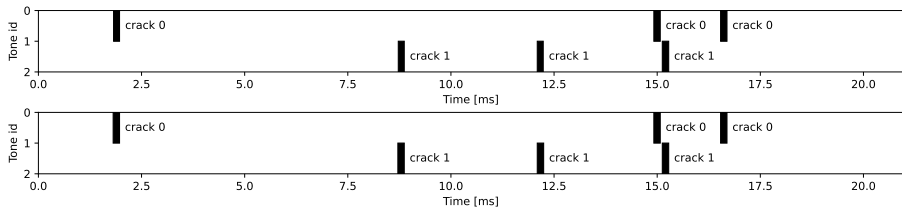


Spectrogram of the measured signal containing 2 bursts.

Event List Generation - Results



ROC AUC of the winning InceptionTime architecture.



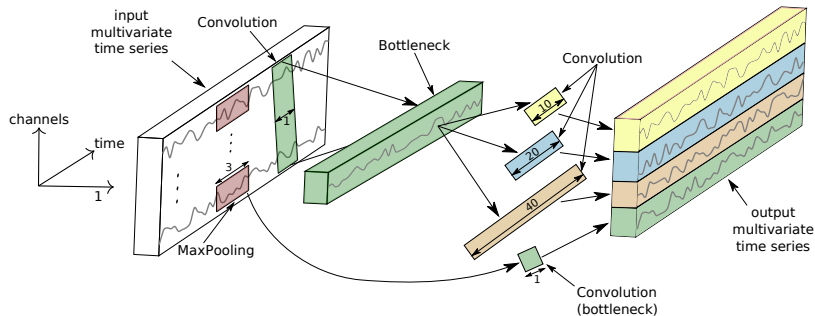
Conclusion

- U-Net appears to be a good choice for signal decomposition
- networks based on raw signal may have high computational demand
- spectrogram networks are simpler and seem to perform better on many signal decomposition and event list generation tasks

Bibliography:

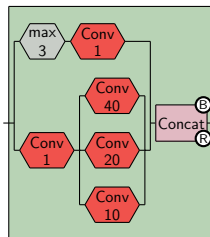
- [1] Hassan Ismail Fawaz et al. "InceptionTime: Finding AlexNet for Time Series Classification". In: *Data Mining and Knowledge Discovery* (2020).
- [2] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: *arXiv:1505.04597 [cs]* (May 2015). arXiv: 1505.04597. URL: <http://arxiv.org/abs/1505.04597> (visited on 04/03/2022).

Question



InceptionTime architecture.

Source: Fawaz et al. [1].



Question

L1	1.000	0.769	0.942	0.641	0.711	0.784	0.381	-0.463
L2	0.769	1.000	0.749	0.555	0.983	0.950	0.749	-0.346
AE	0.942	0.749	1.000	0.853	0.717	0.739	0.340	-0.507
S-L1	0.641	0.555	0.853	1.000	0.571	0.514	0.202	-0.489
S-L2	0.711	0.983	0.717	0.571	1.000	0.927	0.761	-0.413
MEL-L1	0.784	0.950	0.739	0.514	0.927	1.000	0.787	-0.257
MEL-L2	0.381	0.749	0.340	0.202	0.761	0.787	1.000	-0.191
rating	-0.463	-0.346	-0.507	-0.489	-0.413	-0.257	-0.191	1.000
	L1	L2	AE	S-L1	S-L2	MEL-L1	MEL-L2	rating

Correlation matrix of used metrics based on all the results.