Acoustic monitoring based on learnable wavelet transform

Frusque Gaëtan, Baorui Dai, Olga Fink

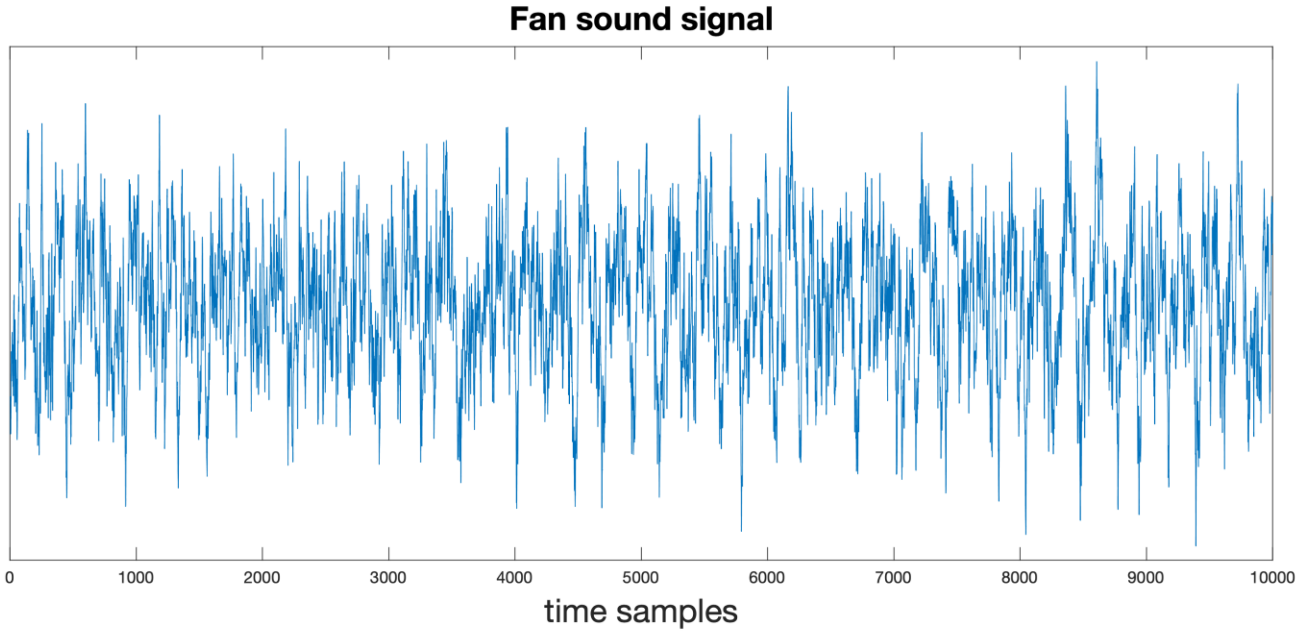
Acoustic monitoring

Audio
Sensor

<u>Advantages :</u>

- Cheap
- Easy to install
- Non-invasive and non-disruptive

Acoustic monitoring



Problems :

- High sampling rate
- High noise level and artefacts
- Sensitive to operating conditions

Plan

I - The Wavelet Packet Transform (WPT)

II - Learnable wavelet transform

III - Application 1: anomaly detection

IV - Application 2: Slab track monitoring

V - Conclusion

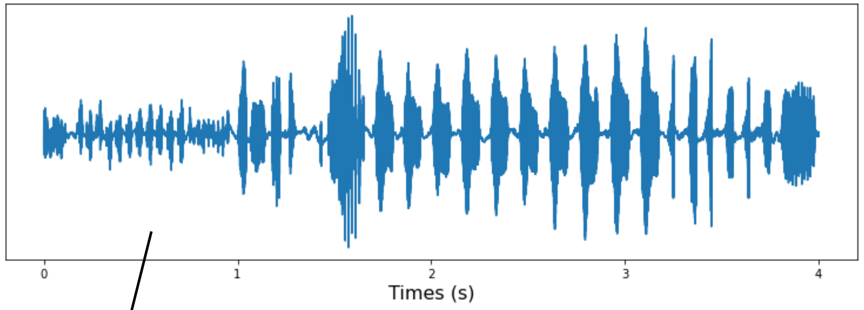
I - The Wavelet Packet Transform (WPT)

II - Learnable wavelet transform

III - Application 1: anomaly detection

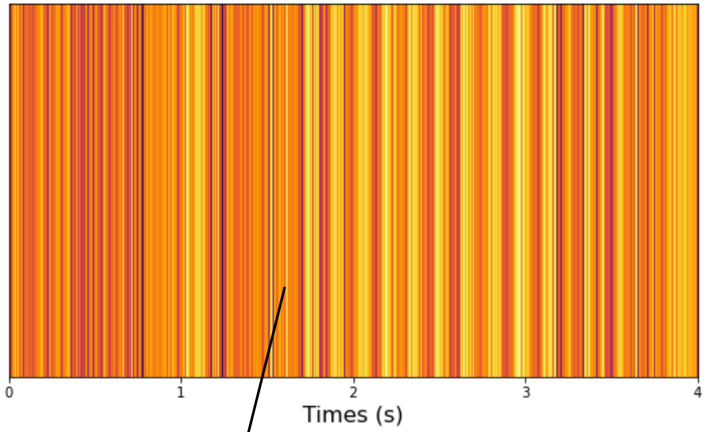
IV - Application 2: Slab track monitoring

V - Conclusion



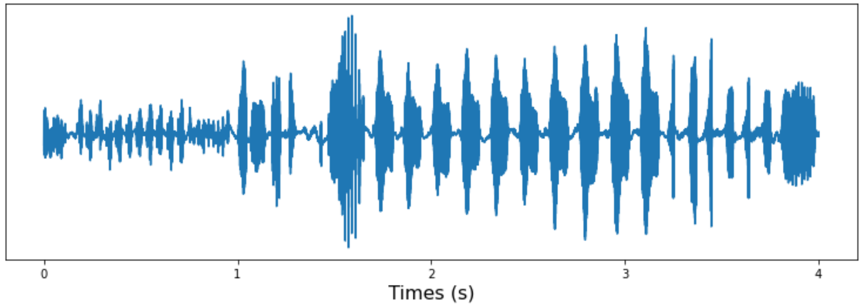
Audio signal

log10(.^2)

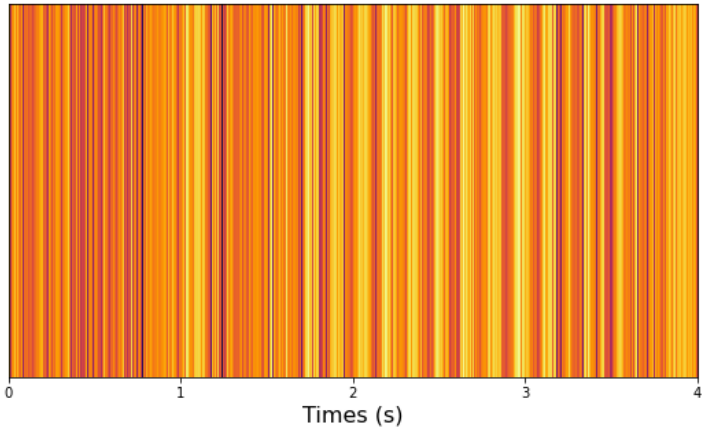


Energy distribution over time

WPT



log10(.^2)



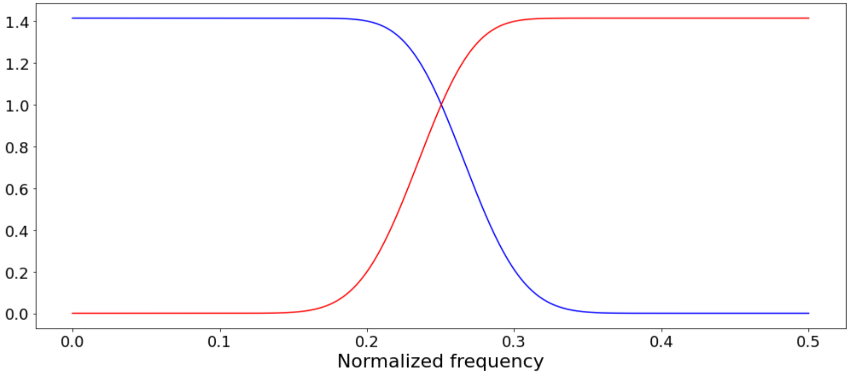
LP

HP

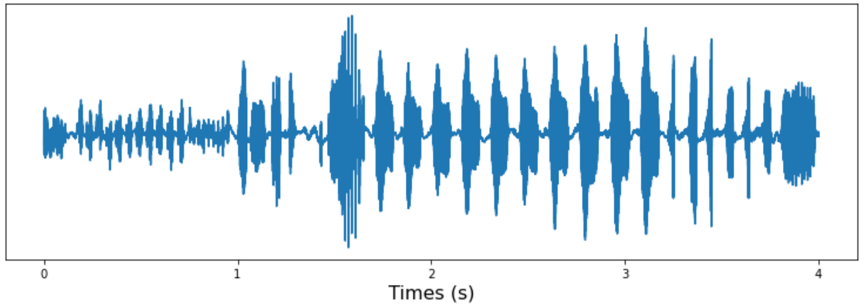
HP = High pass filter

LP = Low pass filter

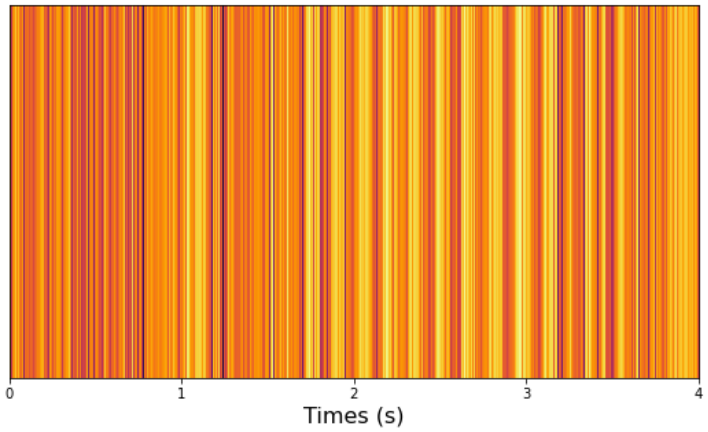
= Sub-sampling by 2

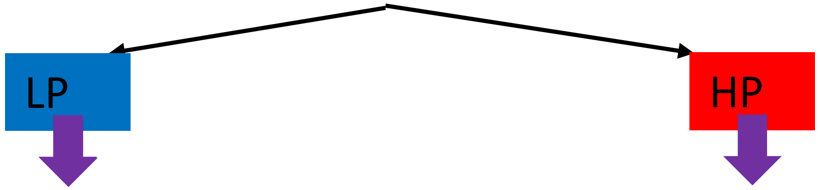


WPT

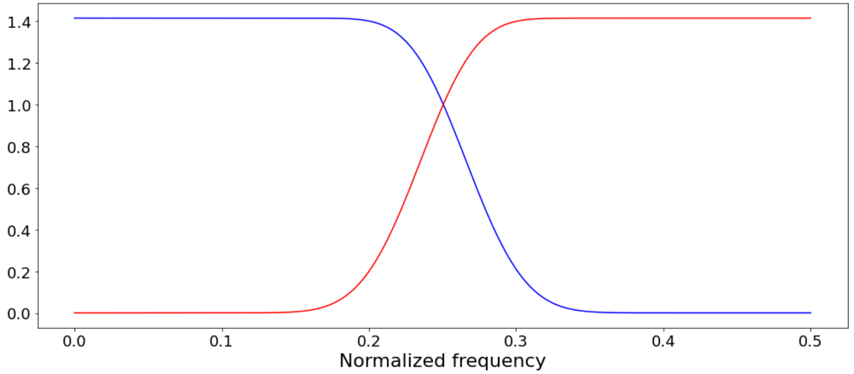


log10(.^2)

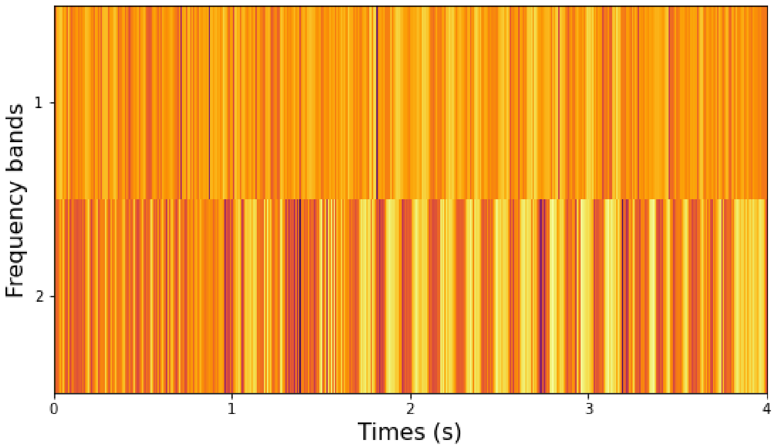




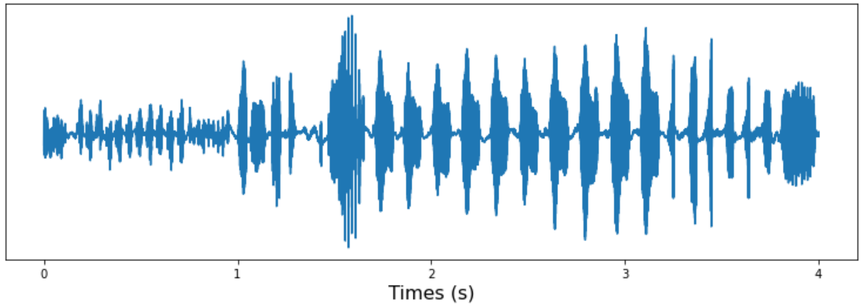
- Increase frequency resolution by 2
- Decrease time resolution by 2



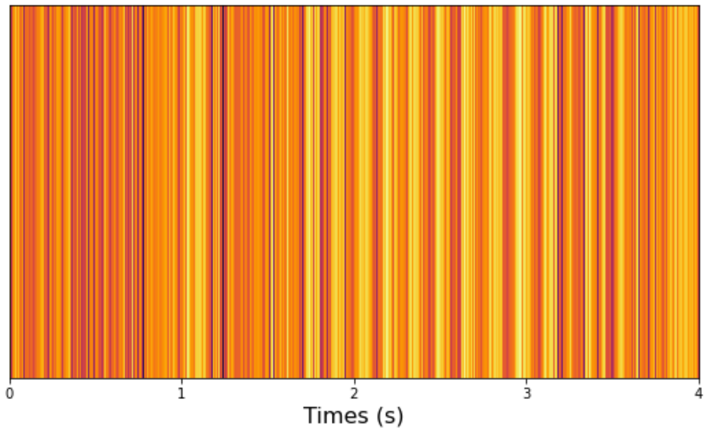
log10(.^2)

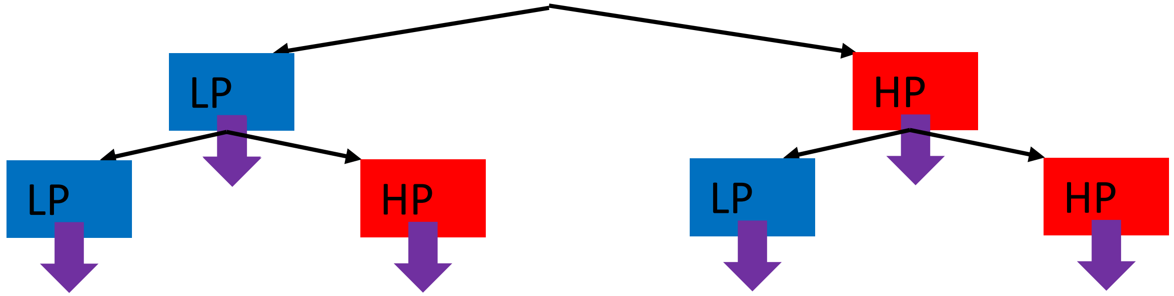


WPT

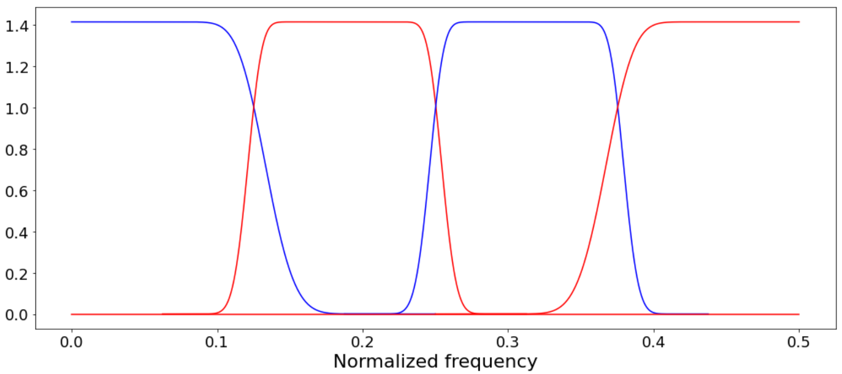


log10(.^2)

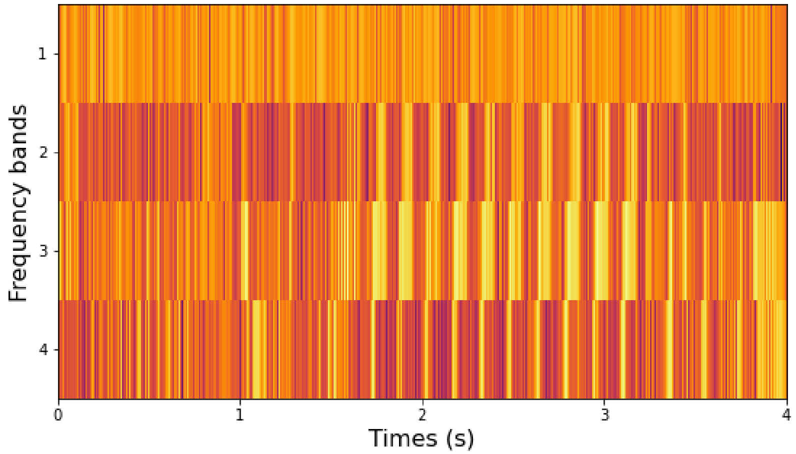




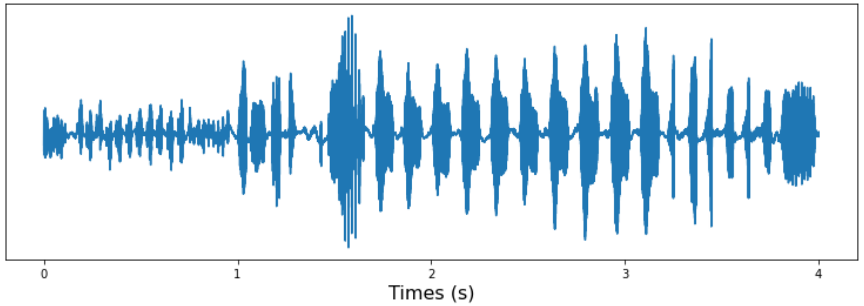
- Increase frequency resolution by 4
- Decrease time resolution by 4



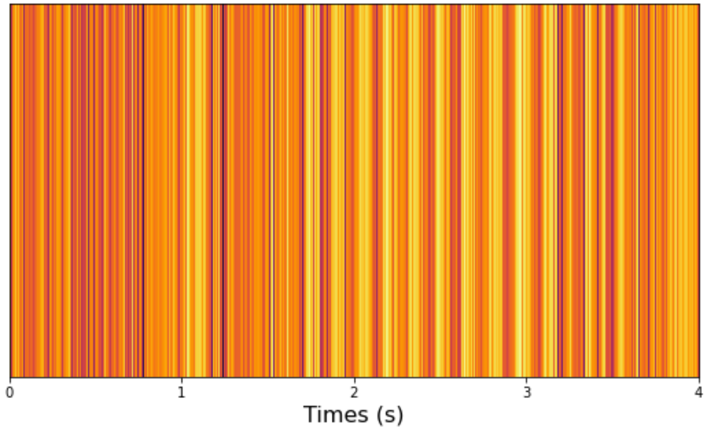
log10(.^2)

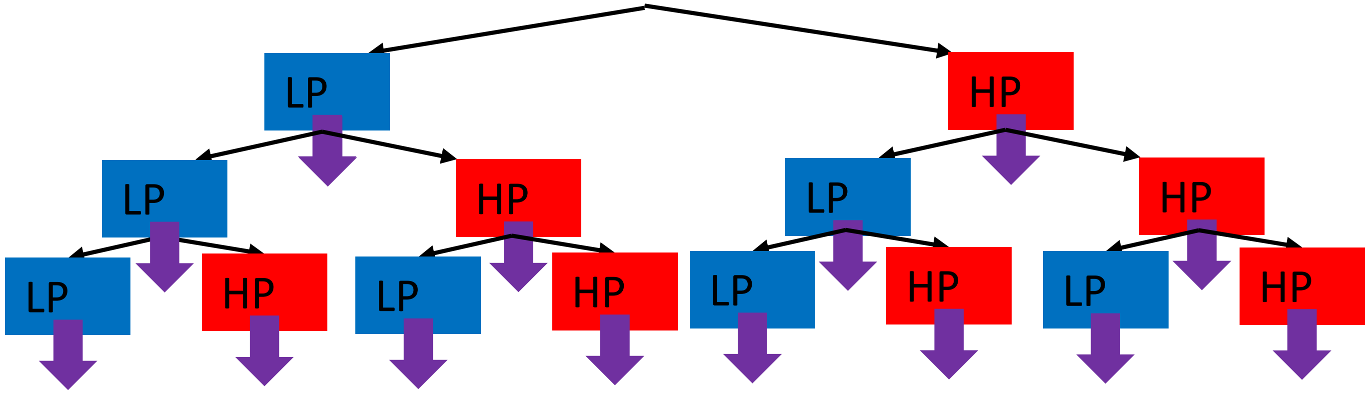


WPT

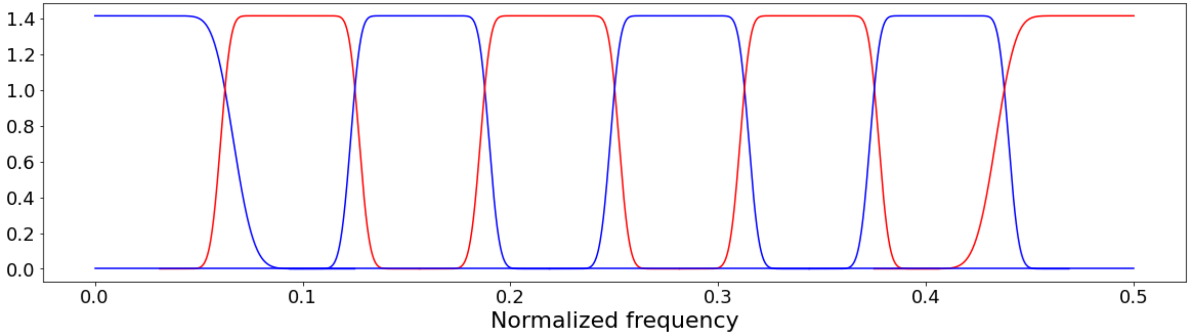


log10(.^2)

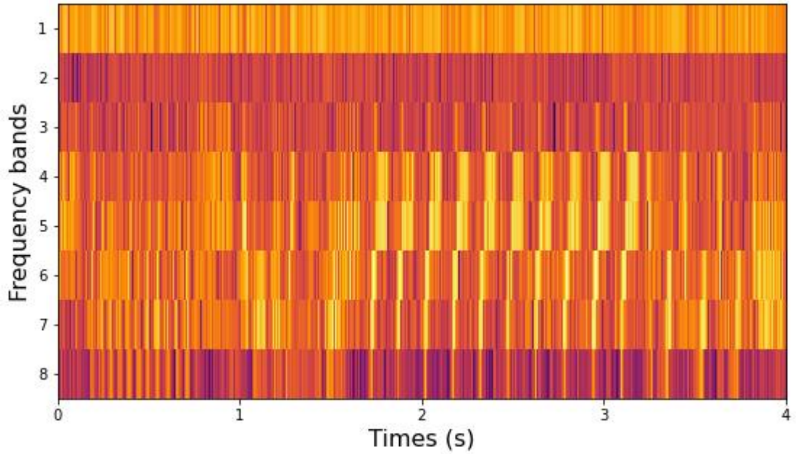




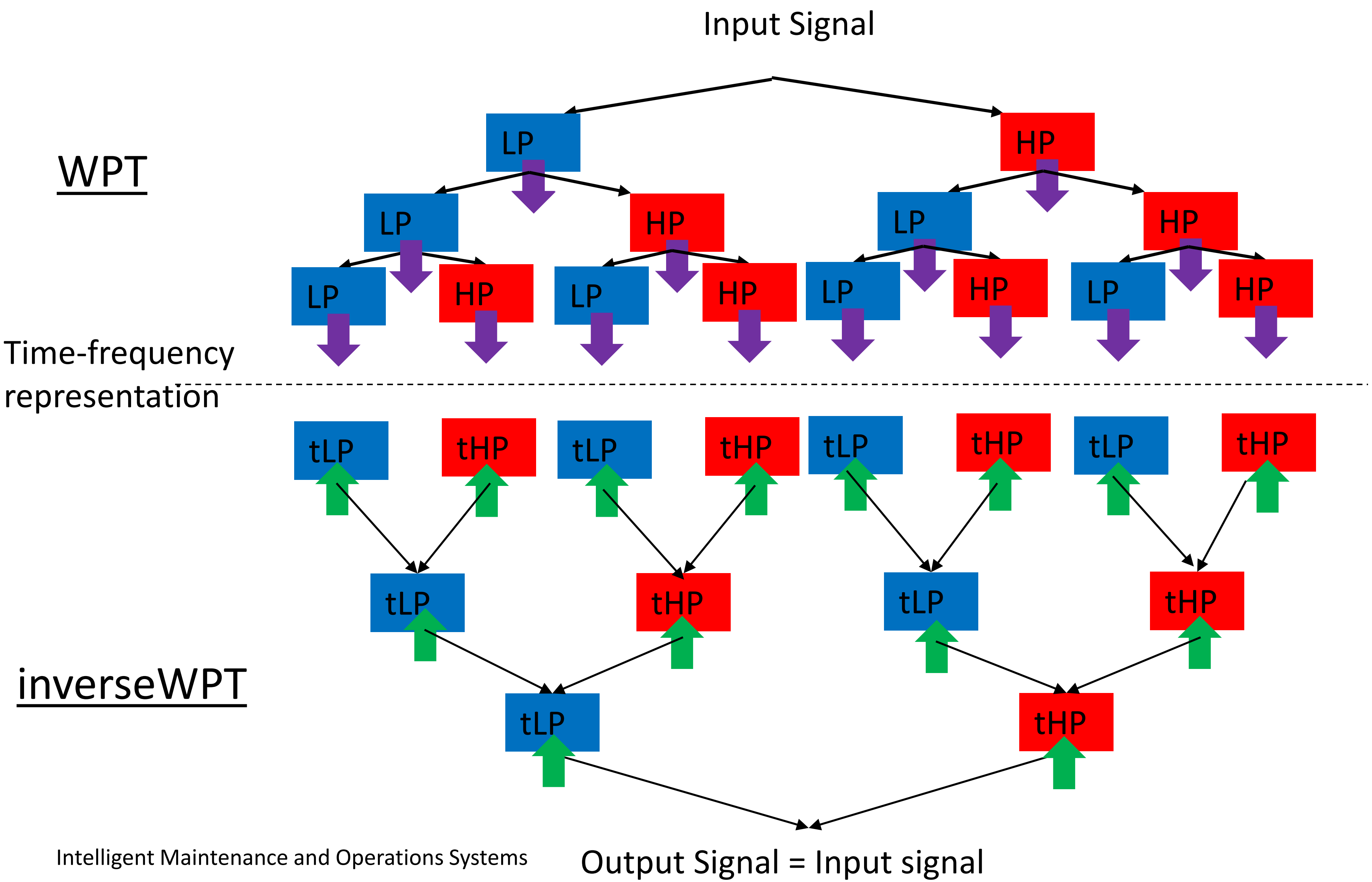
- Increase frequency resolution by 8
- Decrease time resolution by 8



log10(.^2)



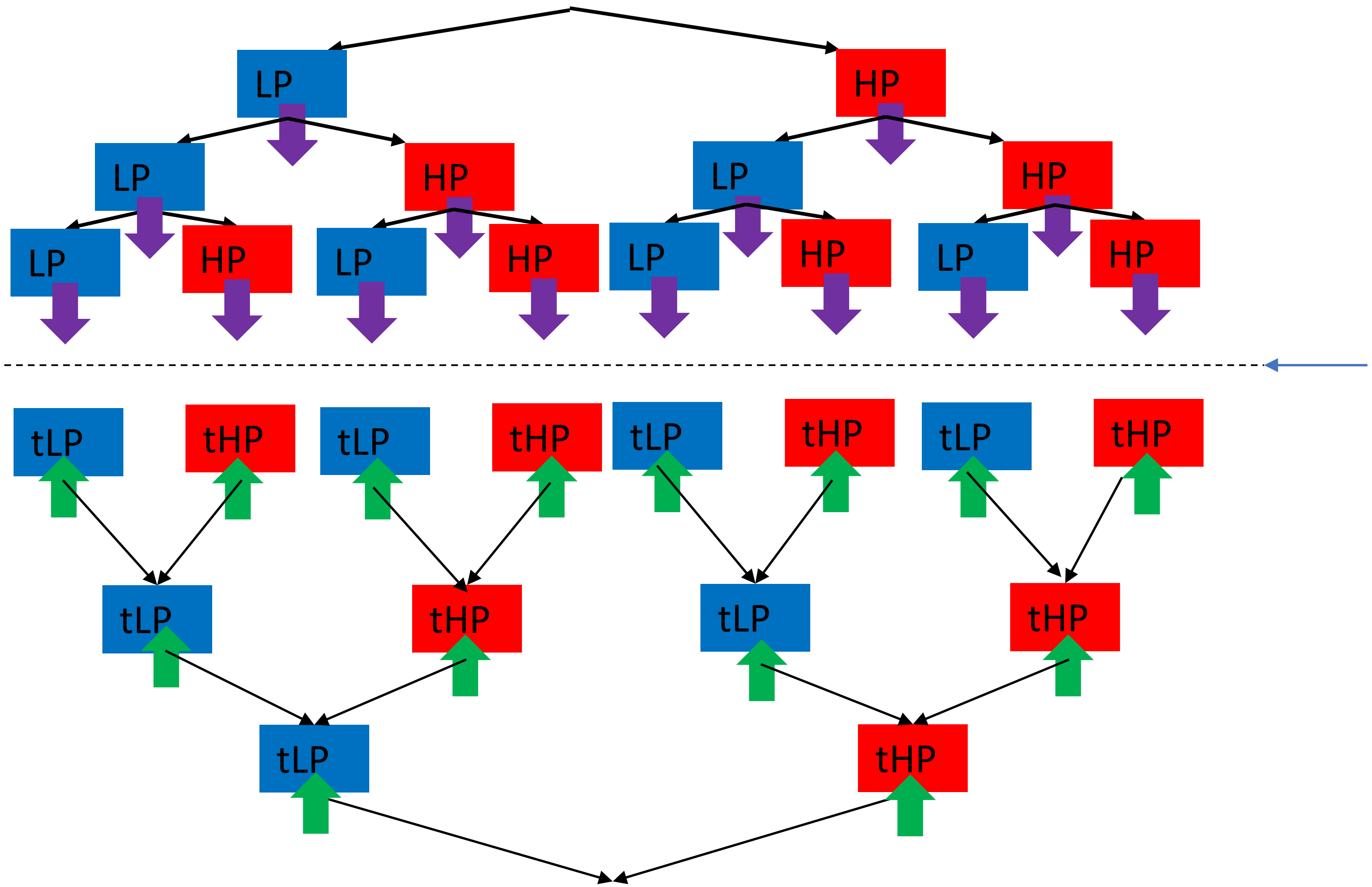
Perfect reconstruction



Denoising with WPT

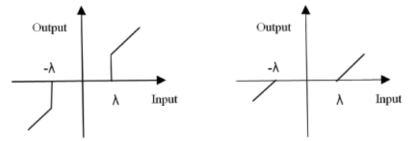
Input Signal

WPT



Time-frequency representation

Hard threshold function



inverseWPT

Output Signal = Input signal

Plan

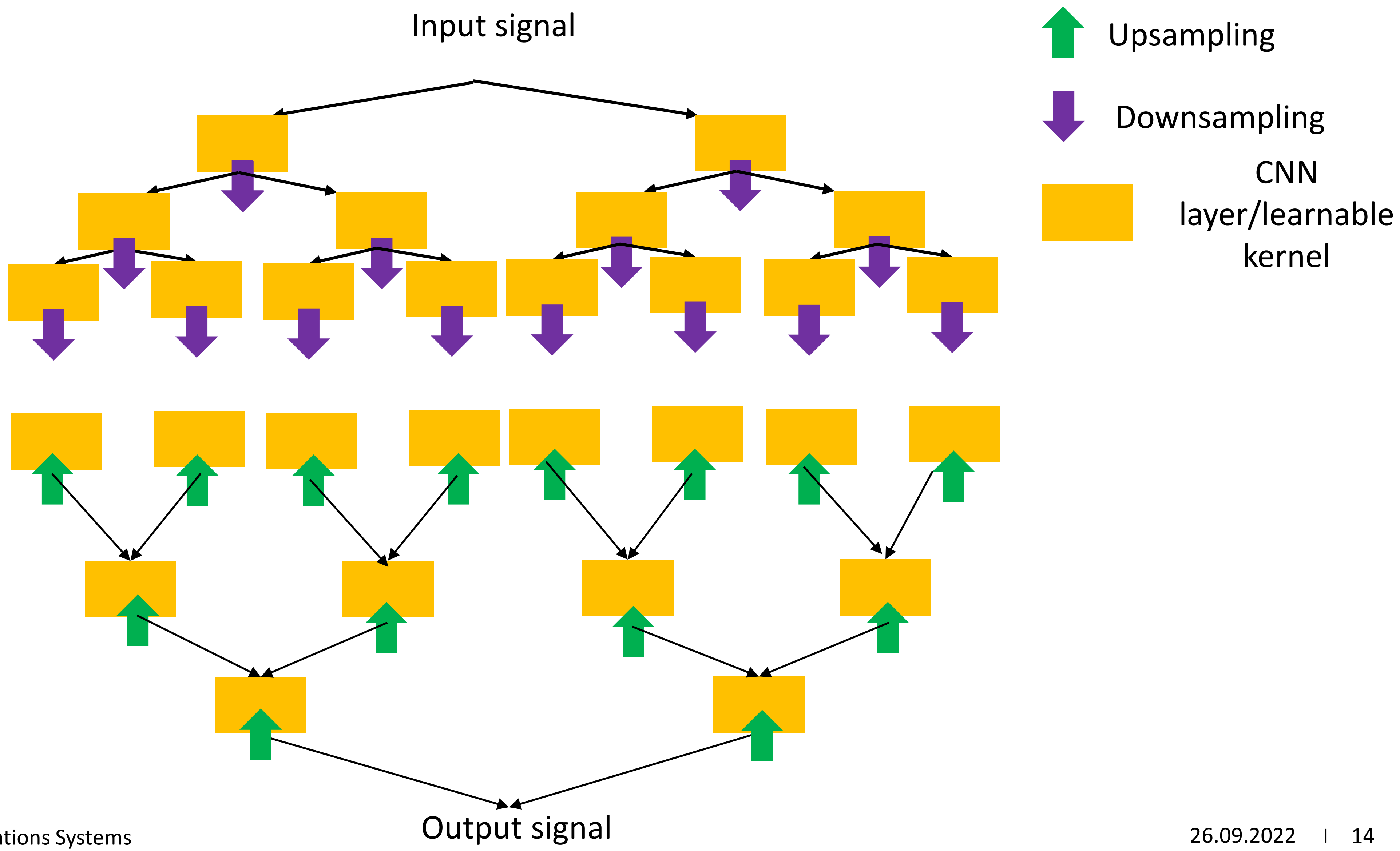
I - The Wavelet Packet Transform (WPT)

II - Learnable wavelet transform

III - Application 1: anomaly detection

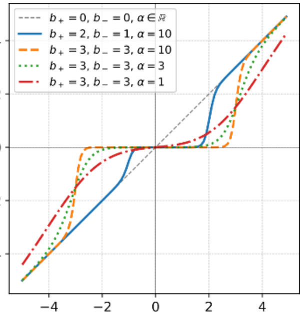
IV - Application 2: Slab track monitoring

V - Conclusion

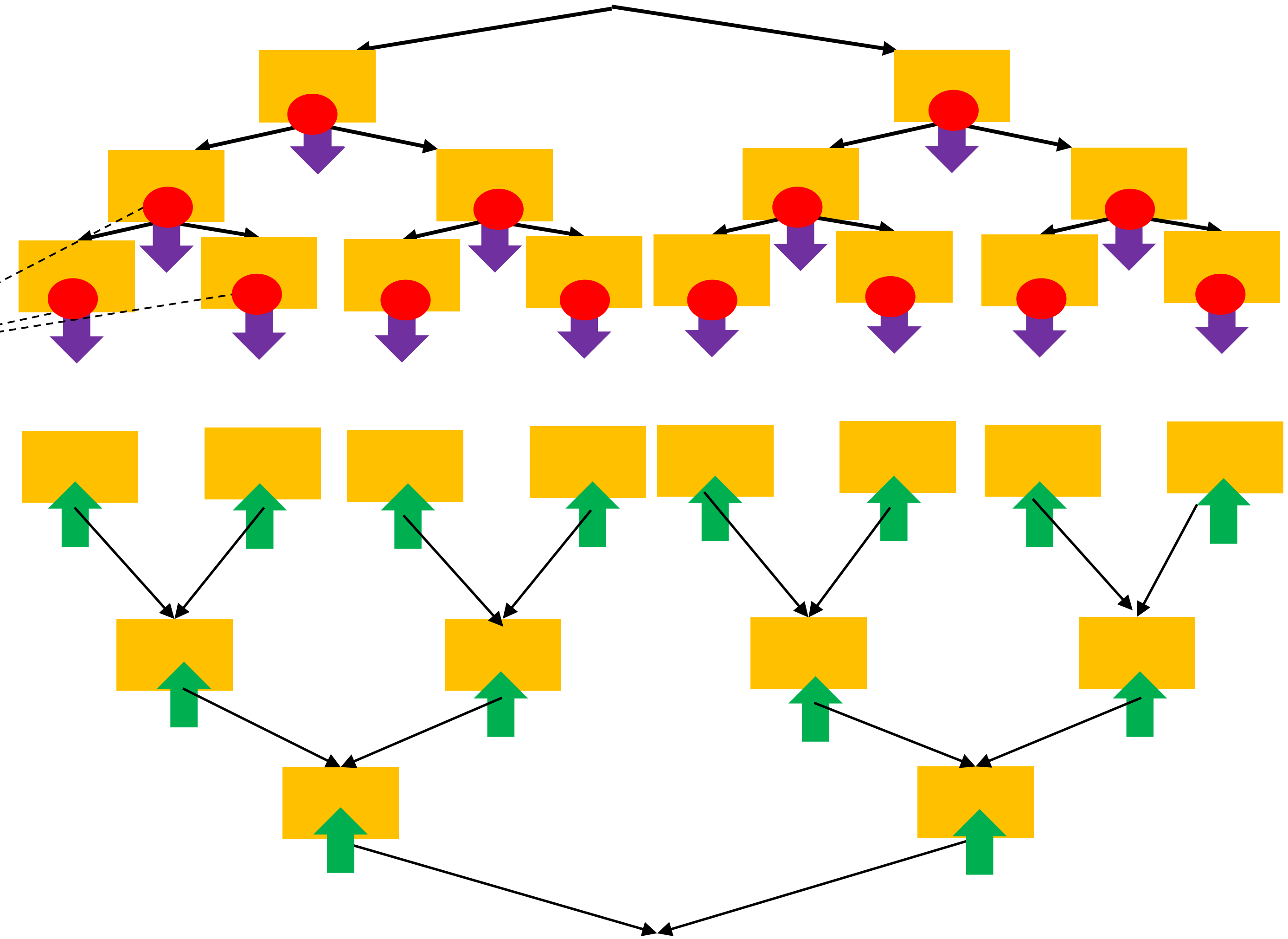

Input signal
Upsampling
Downsampling
CNN layer/learnable kernel
Output signal

L-WPT

HT based on two sigmoid functions



Input signal



Output signal

Upsampling

Downsampling

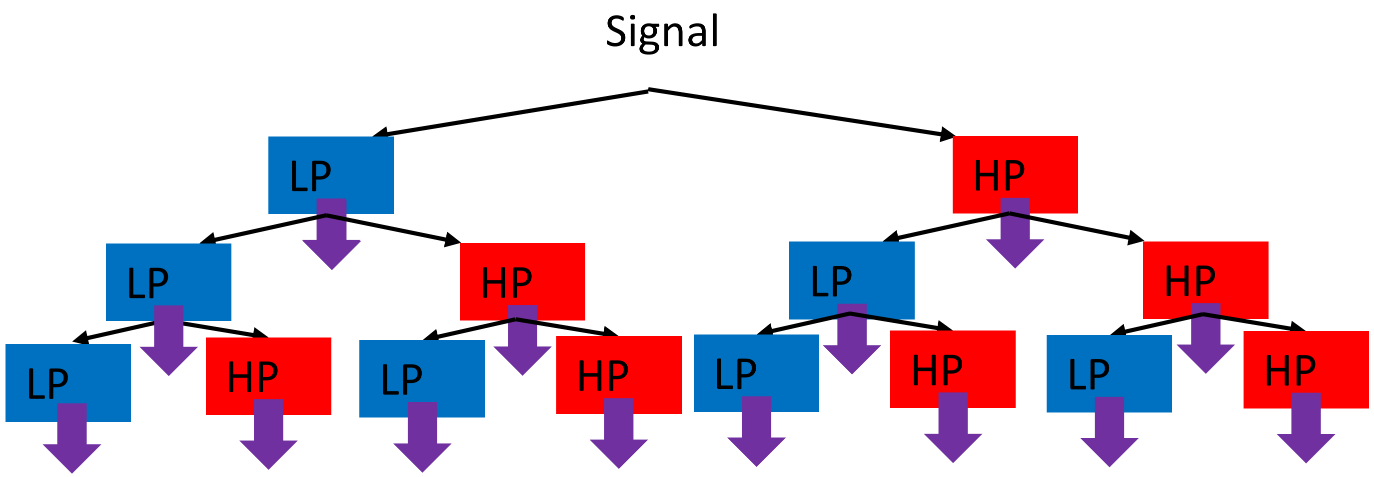
CNN layer/learnable kernel

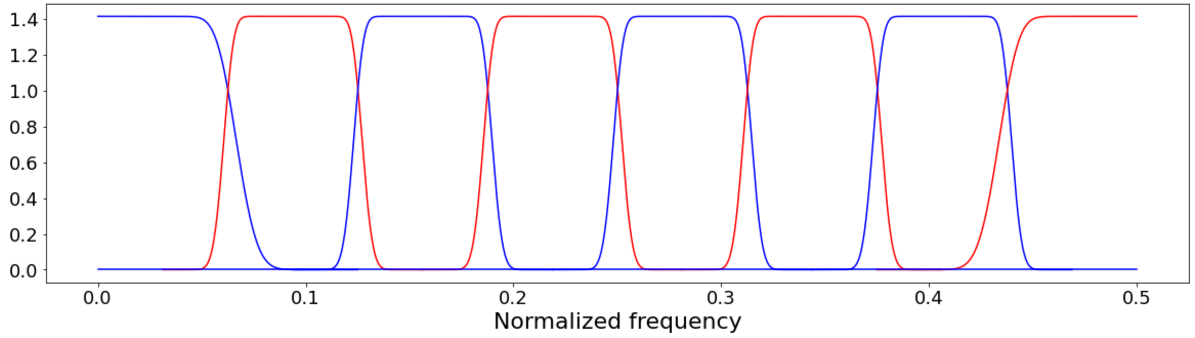
Learnable denoising

Frusque, G., & Fink, O. (2022). Robust Time Series Denoising with Learnable Wavelet Packet Transform.

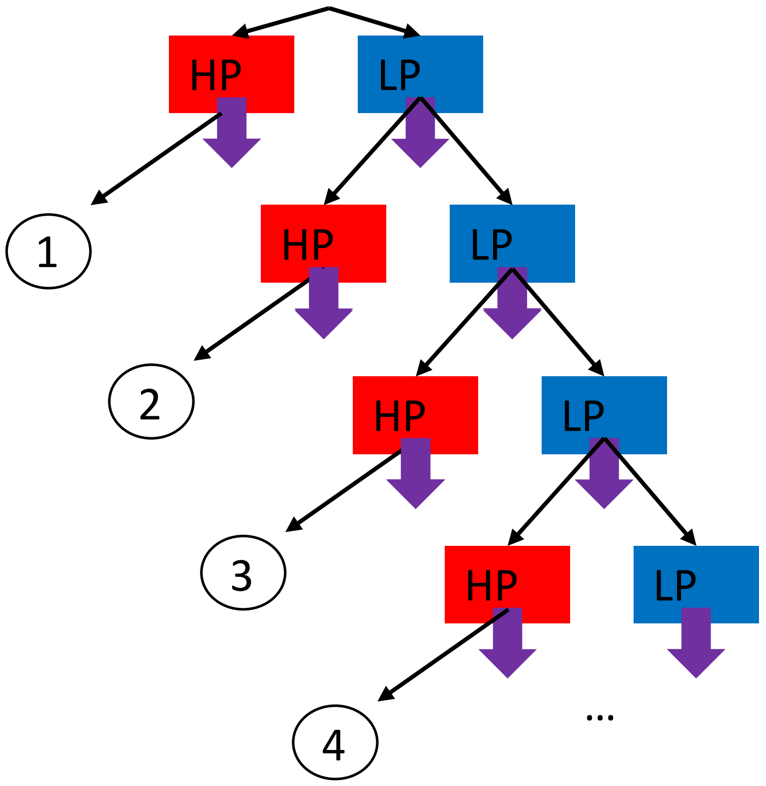
Advantages of the architecture

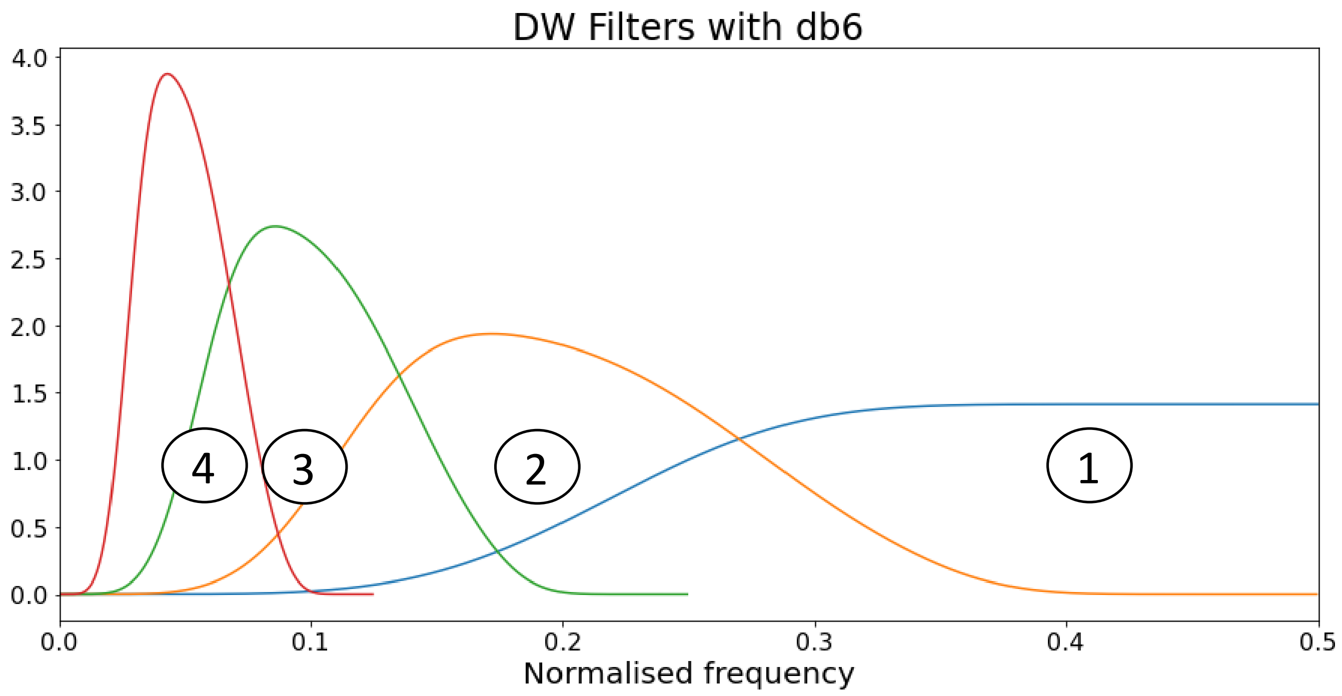
- 1) Huge learning capabilities with few parameters.
- 2) A natural way to initialise parameters so L-WPT starts behaving like a WPT.
- 3) Clear interpretation of each parameter in the structure.


Signal
LP
HP
LP
HP
LP
HP
LP
HP
LP
HP
LP
HP
LP
HP

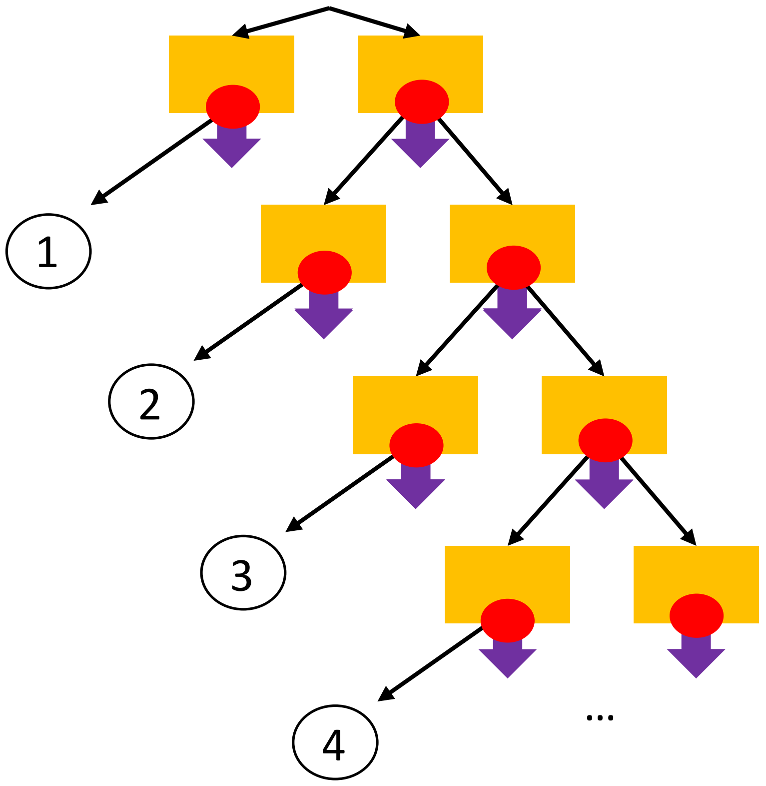

1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.1
0.2
0.3
0.4
0.5
Normalized frequency

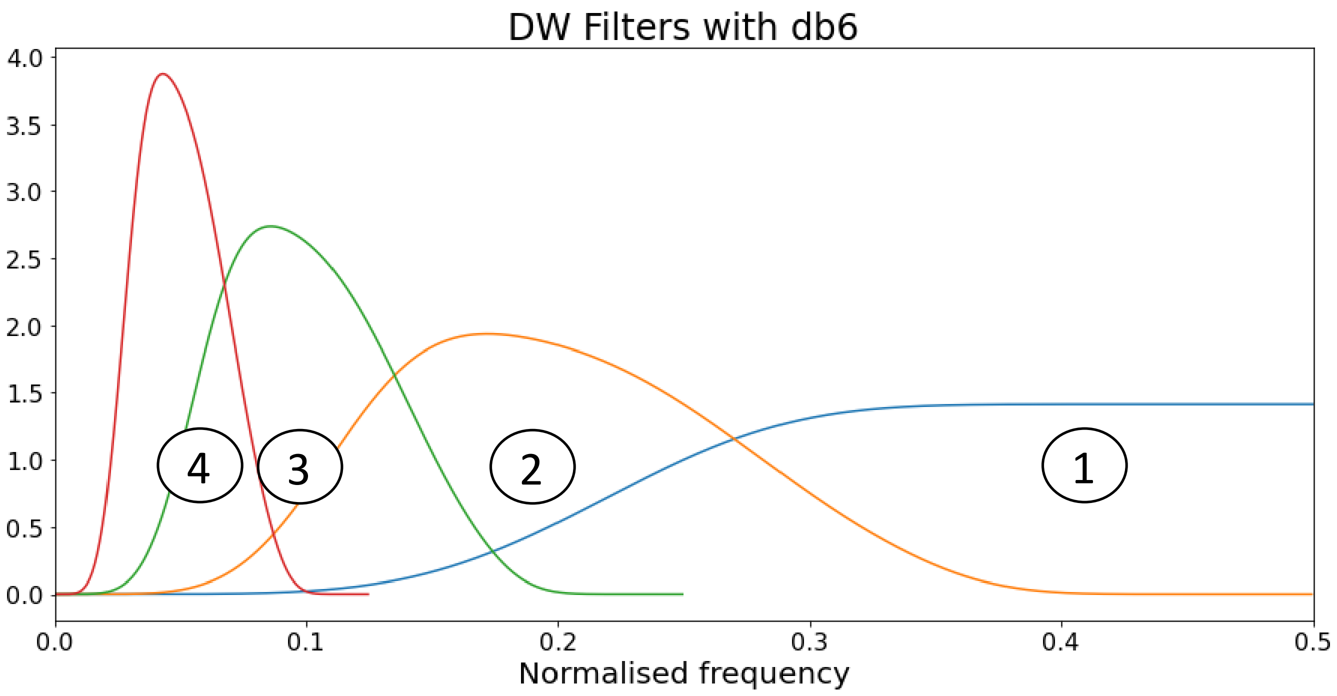
Discrete Wavelet Transform





DeSpaWN architecture





Michau, G., Frusque, G., & Fink, O. (2022). Fully learnable deep wavelet transform for unsupervised monitoring of high-frequency time series. *Proceedings of the National Academy of Sciences*, *119*(8), e2106598119.

I - The Wavelet Packet Transform (WPT)

II - Learnable wavelet transform

III - Application 1: anomaly detection

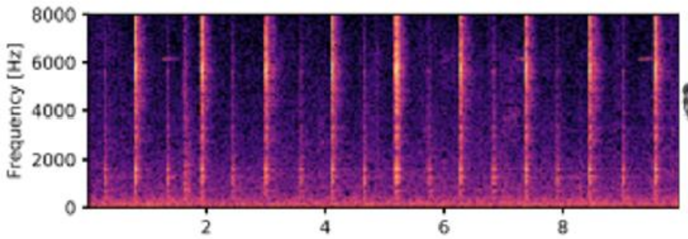
IV - Application 2: Slab track monitoring

V - Conclusion

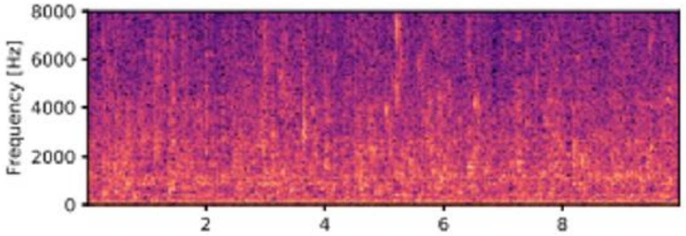
MIMII DATASET: SOUND DATASET FOR MALFUNCTIONING INDUSTRIAL MACHINE INVESTIGATION AND INSPECTION

Harsh Purohit, Ryo Tanabe, Kenji Ichige, Takashi Endo, Yuki Nikaido, Kaori Suefusa, and Yohei Kawaguchi

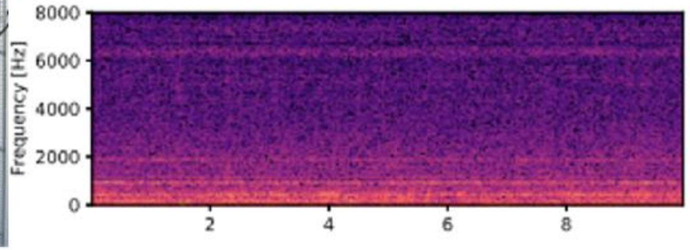
Valve



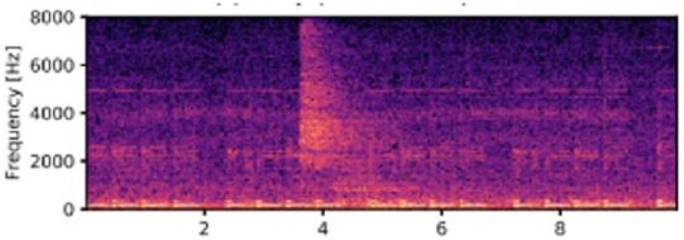
Pump

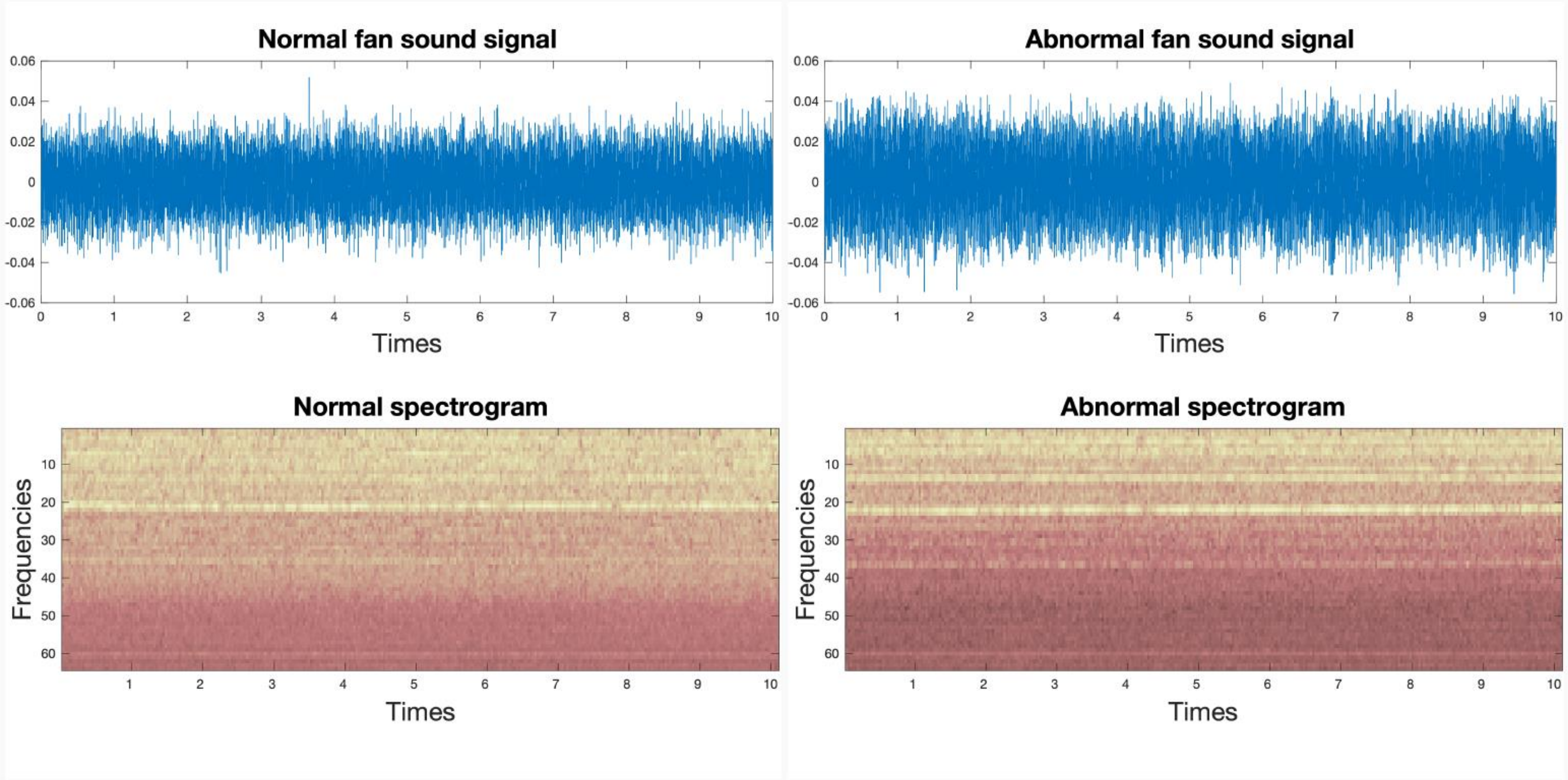


Fan



Slide rail



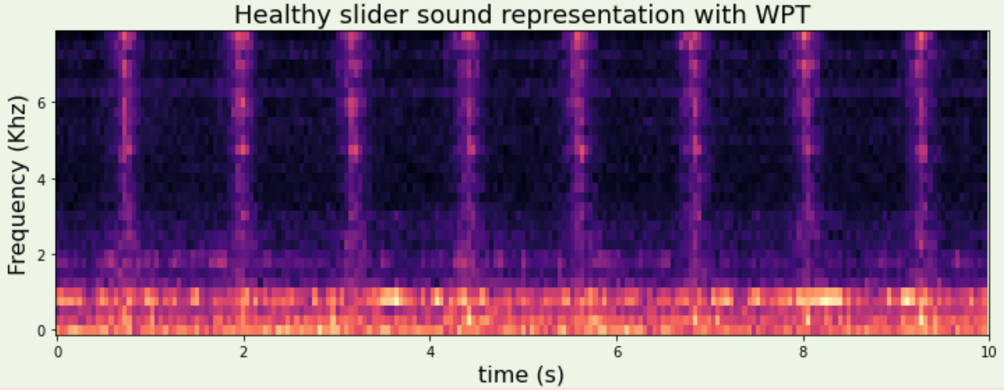


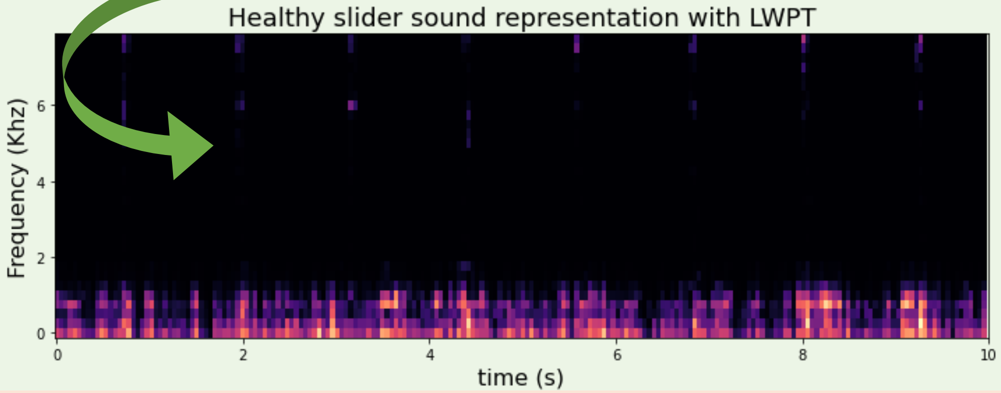
Normal sound

Abnormal sound

Healthy sound

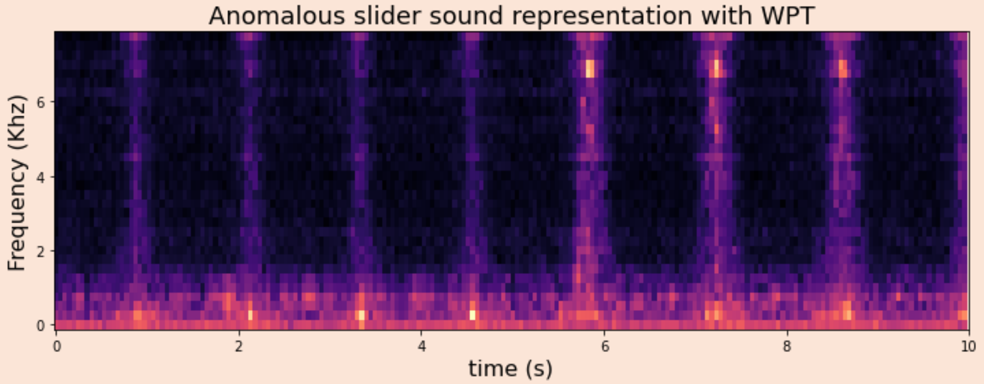
- L-WPT is trained on healthy sound only, we encourage parsimonious healthy sound representations

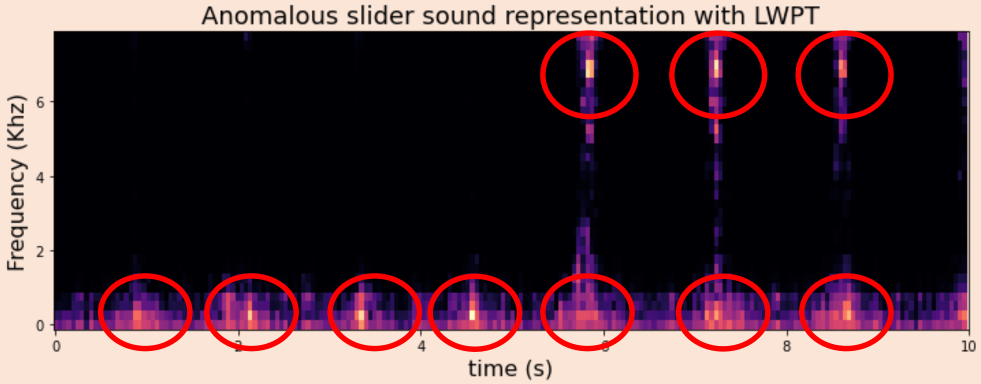




Abnormal sound

- Main anomalous features are then highlithed thanks to the L-WPT representation
- The anomaly detection task is then simpler with L-WPT





Selected feature:

- Mean/Max of each node of the time-frequency representation
- Mean/Max of the residual

Machine	Fan	Pump	Slider	Valve	All
WPT	91.3	85.6	97.8	97.0	92.9
DesPawn	92.4	82.2	92.5	94.0	90.3
WPT-CNN	91.8	86.8	98.5	**97.3**	93.6
L-WPT	**94.9**	**87.7**	**99.5**	97.0	**94.8**

Table 1: mean AUC for the different machine types

Frusque, G, & Fink, O. (2022, May). Learnable Wavelet Packet Transform for Data-Adapted Spectrograms. In *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 3119-3123). IEEE.

Plan

I - The Wavelet Packet Transform (WPT)

II - Learnable wavelet transform

III - Application 1: anomaly detection

IV - <u>Application 2: Slab track monitoring</u>

V - Conclusion

Background

Acoustic signal radiated by the target object is often impacted by:

- Environmental noise
- Other sounds radiated by irrelevant structures or mechanisms

Solution to this problem:

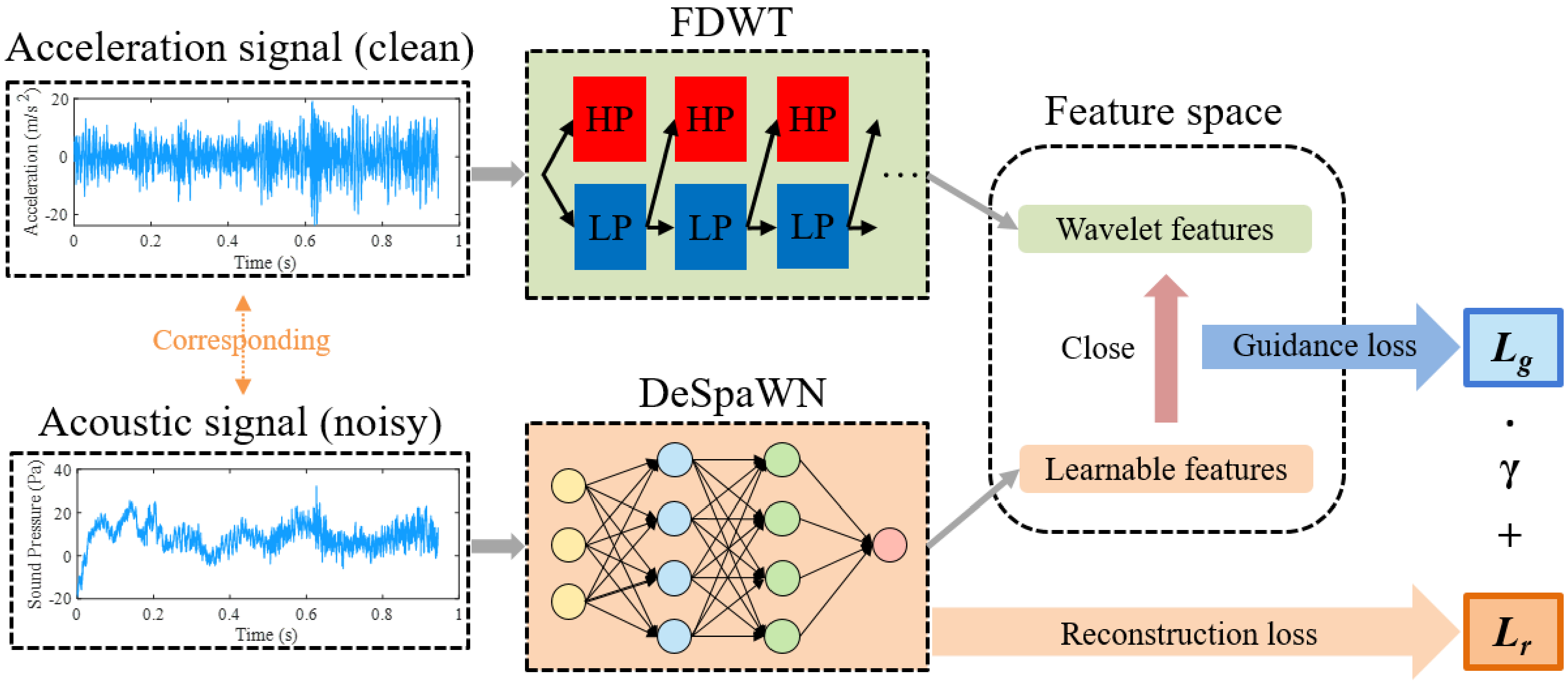
- Unsupervised machine learning method: Performance limited
- Supervised deep denoising model: Often impossible to record the pure acoustic signals

Our method:

- An acoustic denoising framework that can learn a strong denoising transform under the guidance of the corresponding acceleration signals

Architecture of Acceleration-guided acoustic signal denoising framework (AG-ASDF) :

- In training stage: Requiring both acoustic signals and corresponding acceleration signals
- In application stage: Requiring only acoustic signals



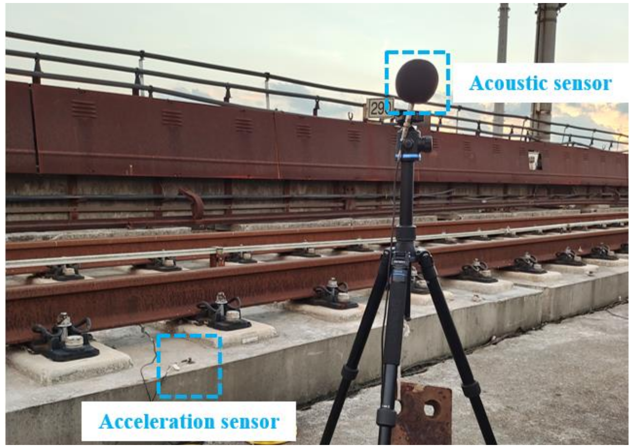
Architecture of AG-ASDF

Experiment

Three types of slab track with different supporting conditions :

- Imitating different healthy and unhealthy states of slab tracks

Metro train



Acceleration and acoustic sensors

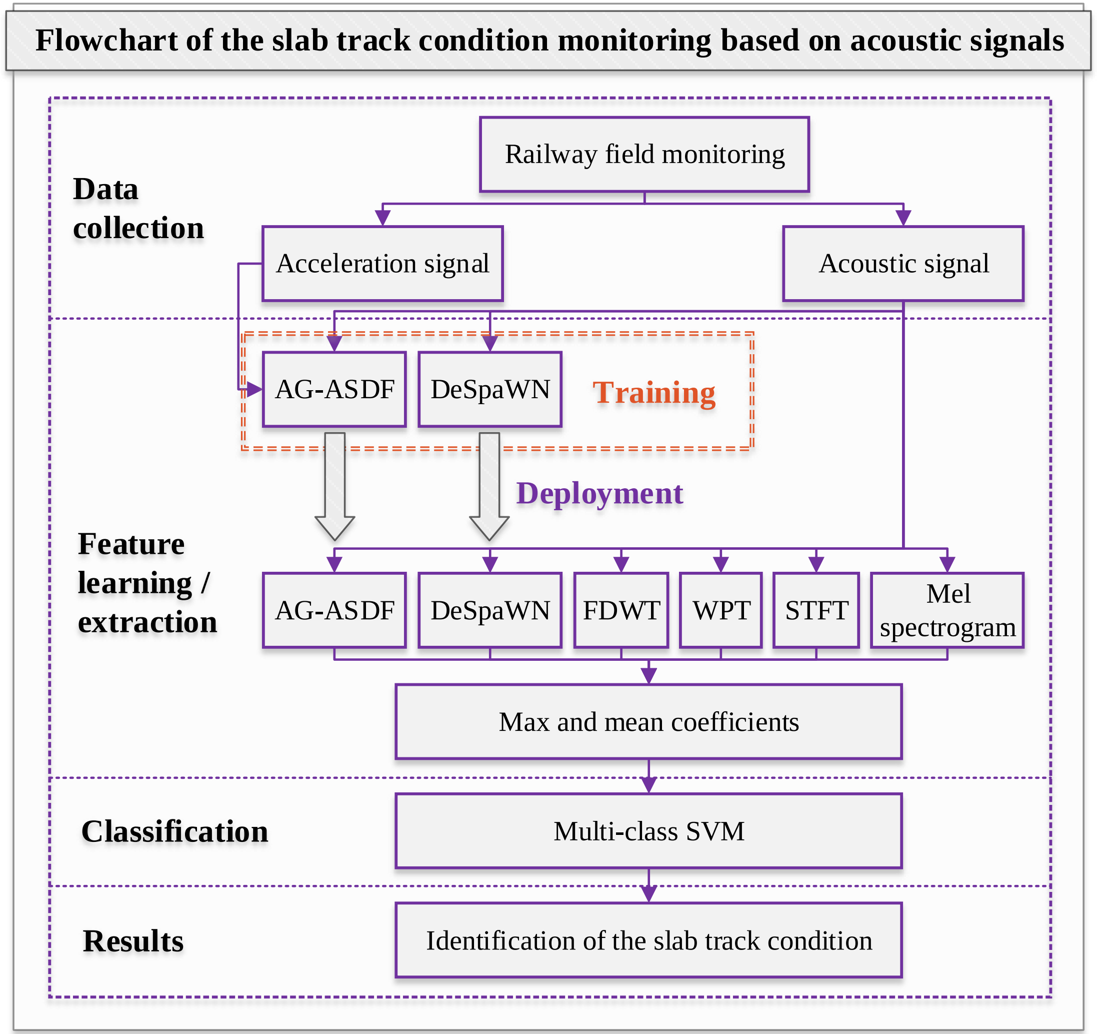
Supporting conditions of three types of slab track: (a) No degradation; (b) Intermediate degradation; (c) Severe degradation.

Test conditions

Speed (km/h)	Number of train passes	Average duration of effective signals (s)		
		Slab Track 1	Slab Track 2	Slab Track 3
20	6	21.84	22.34	22.04
40	6	11.41	11.47	11.43
60	6	7.94	7.85	7.90
80	6	6.11	6.12	6.20

Comparative study

Flowchart of the slab track condition monitoring :



- AG-ASDF: Acceleration-guided acoustic signal denoising framework
- DeSpaWN: Denoising Sparse Wavelet Network
- FDWT: Fast discrete wavelet transform
- WPT: Wavelet Packet Transform
- STFT: Short-time Fourier transform
- Mel spectrogram: Mel spectrogram

Results

Preliminary study:

- Classification accuracy based on the acceleration signals is 100% in all classification tasks

Classification results based on acoustic signals:

Average classification accuracy based on acoustic signals

Training dataset	Test dataset	Classification accuracy (%)					
		AG-ASDF	DeSpaWN	FDWT	WPT	STFT	Mel spectrogram
20, 40, 60, 80 km/h	20, 40, 60, 80 km/h	**95.4**	93.2	87.9	65.8	82.4	88.0
20, 40, 60 km/h	80 km/h	**90.1**	82.1	81.5	48.2	57.4	50.9
40, 60, 80 km/h	20 km/h	**57.4**	49.6	36.1	40.7	46.7	39.8
20, 60, 80 km/h	40 km/h	**94.4**	83.4	79.6	72.2	74.1	76.8
20, 40, 80 km/h	60 km/h	**94.7**	85.2	84.3	67.6	79.6	79.6
Average		**86.4**	78.7	73.9	58.9	68.0	67.0

Results analysis:

- AG-ASDF reaches a superior performance compared to other feature extraction and learning methods
- Speed interpolation regimes perform better than speed extrapolation regimes in classification tasks
- Poor performance in Training 40, 60, 80 km/h and test 20 km/h: Low SNR of acoustic signals with 20 km/h

EPFL

Plan

I - The Wavelet Packet Transform (WPT)

II - Learnable wavelet transform

III - Application 1: anomaly detection

IV - Application 2: Slab track monitoring

V - Conclusion

Conclusion

- Deep model based on Wavelet Packet Transform: Good learning capabilities with few parameters, meaningful weights, natural initialisation.

- Able to learn data adapted spectrograms from training dataset.

- Outperform other similar approaches on anomaly detection and slab track monitoring

Perspectives:

- Propose end-to-end deep model on a supervised task

- Analysing how to better interpret the obtained data-adapted spectrogram

Thanks for your attention !