Machine listening

Peter Karsmakers

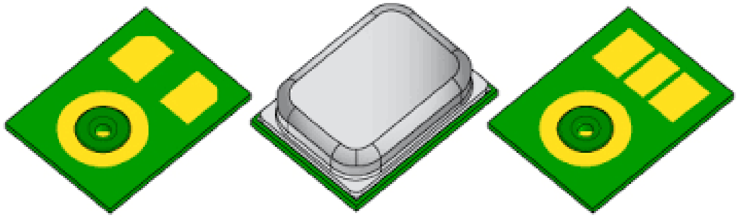
peter.karsmakers@kuleuven.be

Outline

- Information enclosed in acoustic signals
- Why microphones
- Applications
- Challenges
- Machine learning approach
- Case study: smart homes
- Case study: condition based monitoring
- Future research directions
- Conclusion

Information enclosed in acoustics

Why use microphones to sense the environment



analog.com

- Non-destructive & contactless measurement,
- Monitor multiple acoustic sources using a single omni-directional sensor,
- Compared to video, acoustic signals travel through obstacles, and are less affected by environmental conditions such as fog, pollution, rain, and daily changes in light conditions and consume less power,
- Having more than 1 microphone allows for localisation of sources.

Acoustic events and scenes

event

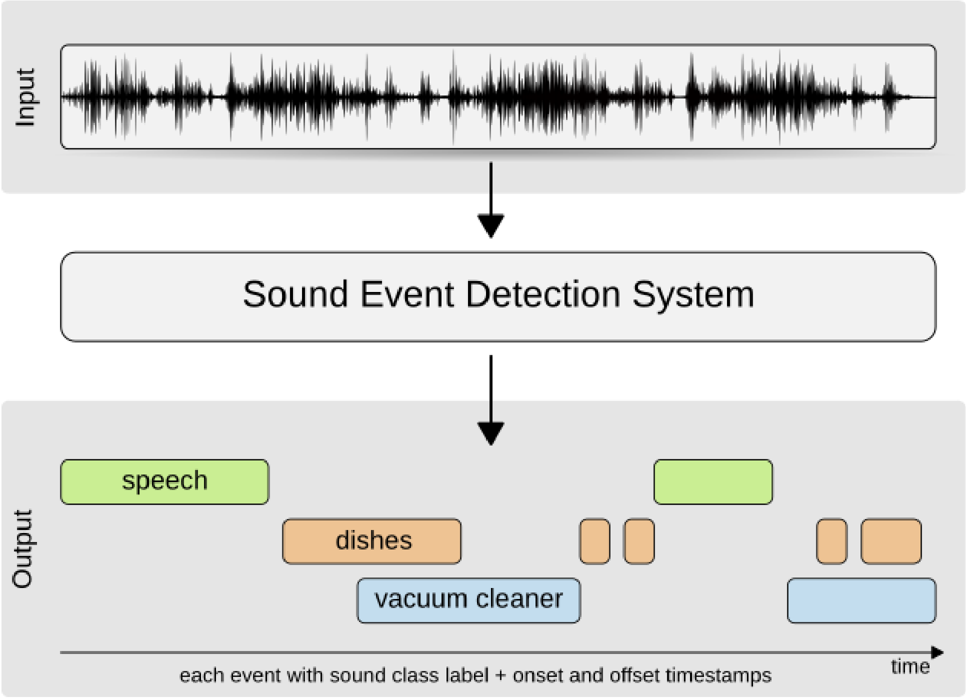
- Single source
- Well-defined brief duration in time

scene

- Mixture of acoustics coming from different sources

Tasks in machine listening

- Classification: describe each event or scene using a textual class label
- Event detection: estimate start- and end-time of each event



http://dcase.community/

Applications

Context awareness in devices

- Computational acoustic analysis endow devices with context awareness
    → <u>Improved QoL</u>

https://www.popsci.com

Smart cities

- Acoustic surveillance

  → Improve city safety

- City noise monitoring

  → Improve urban planning

Precision livestock farming

- Monitor coughs counts in pigs to deliver early warning indicating disease outspread

→ Improve on production yields

Condition monitoring

- Identify modified sounds in assets compared to the normal situation indicating mechanical aging and future failures

  → reduced cost due to unnecessary preventive maintenance and early detection of unexpected failure of elements

http://www.vibrationconsultants.co.nz

Challenges

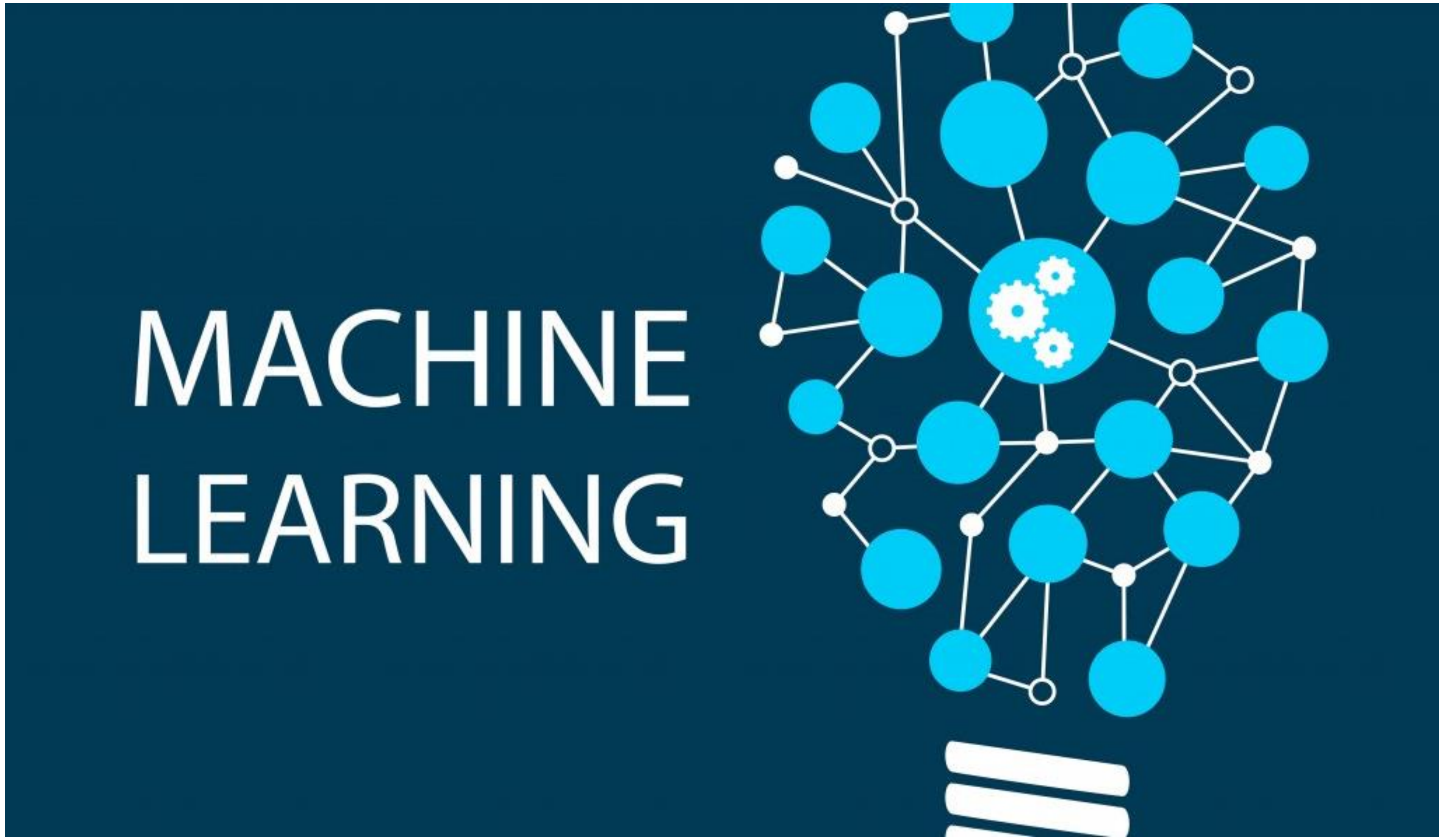
- Large variety of different sounds

- Large acoustic diversity within each class

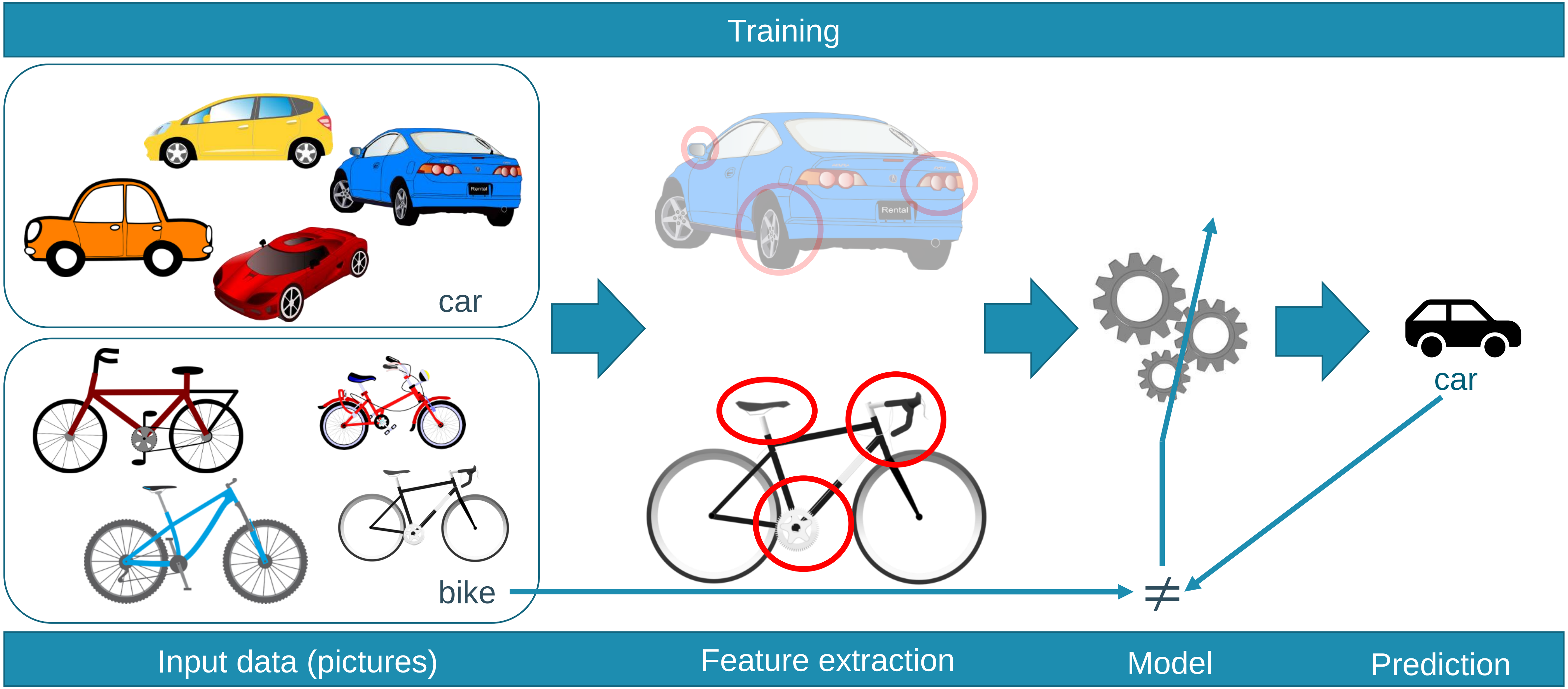
- Overlapping sounds, reverberation



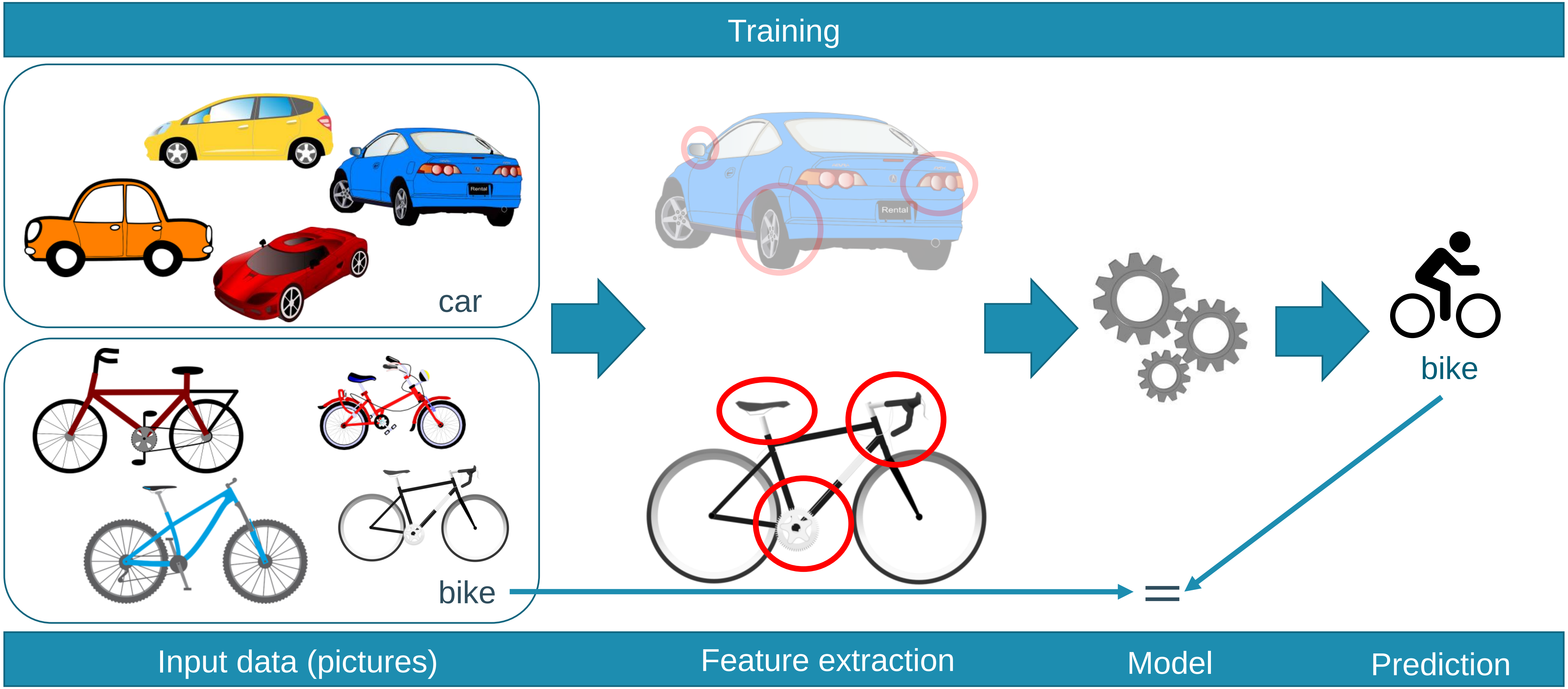
How to let a machine interpret sounds?

- **Define the task** by defining the event/scene labels (classes) in advance
- Practical challenges
  - **large amounts** of **annotated data** are required
  - **computing power** (complex models)

Supervised machine learning

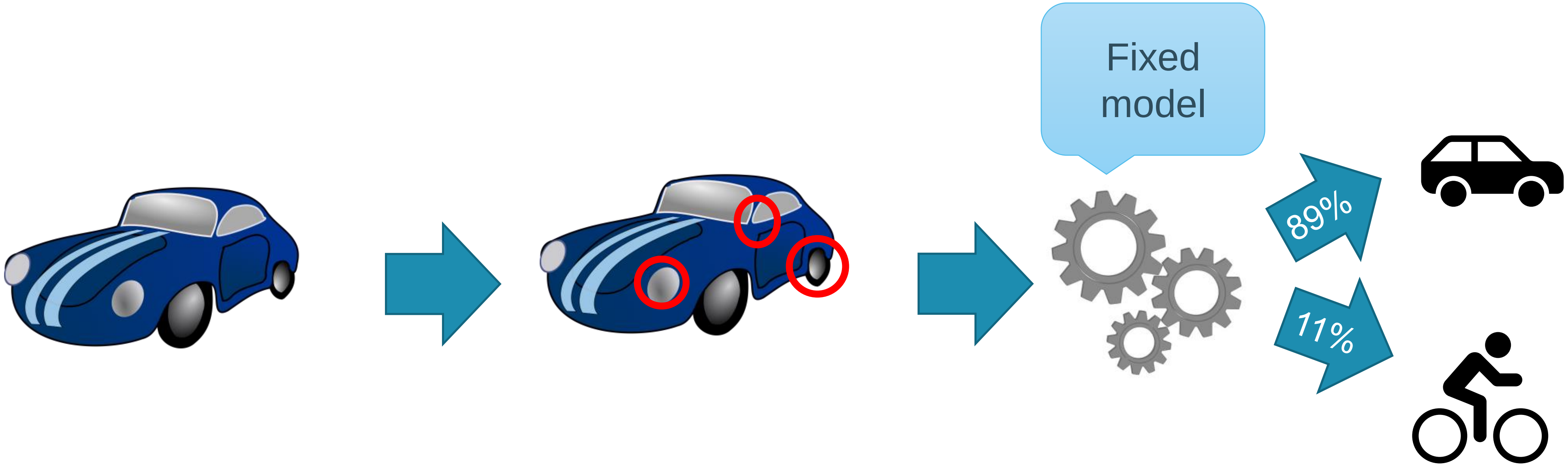


Supervised machine learning



Supervised machine learning

Test

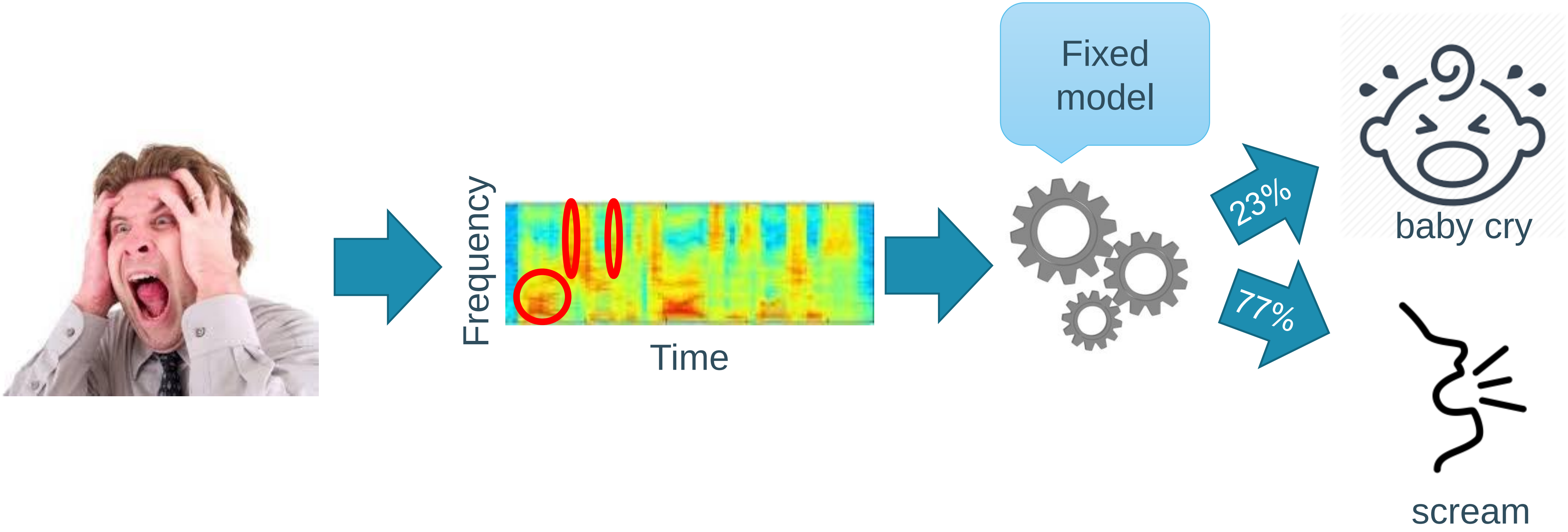


Input data (pictures) | Feature extraction | Model | Prediction

Geel Campus
Faculty of Engineering Technology
KU LEUVEN

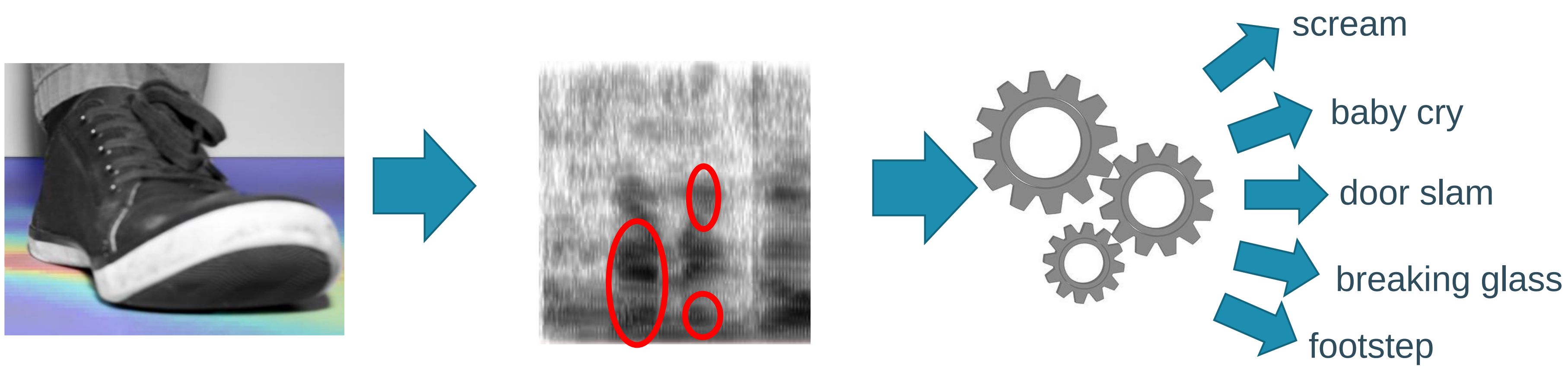
Supervised machine learning

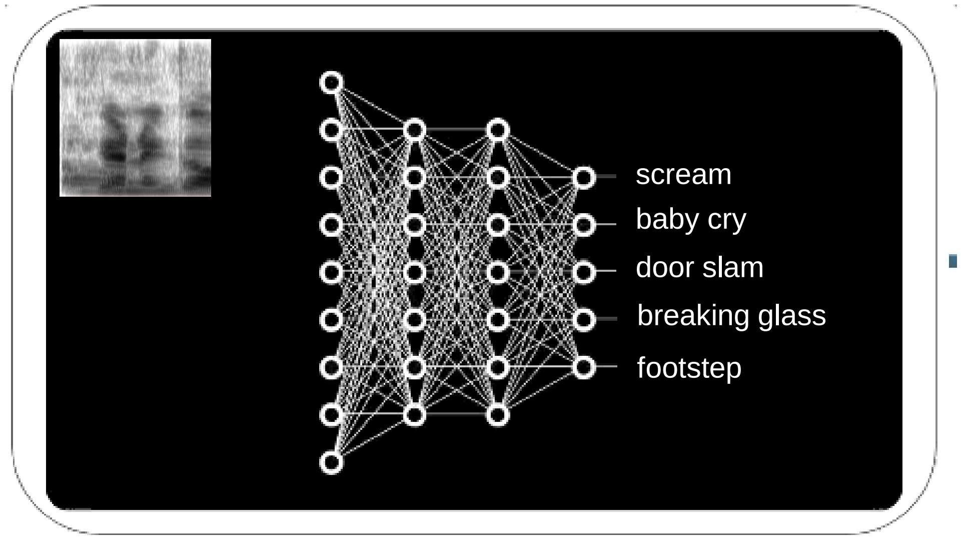
Test



Input data (pictures) | Feature extraction | Model | Prediction

Machine learning subfield: deep learning



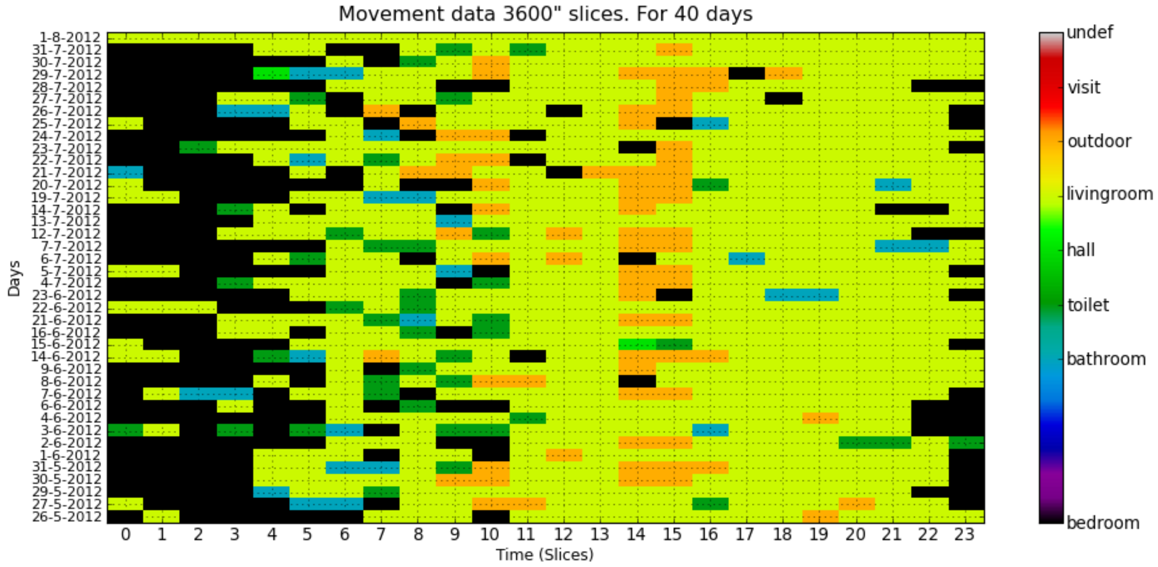


simplilearn.com

Case study: smart homes

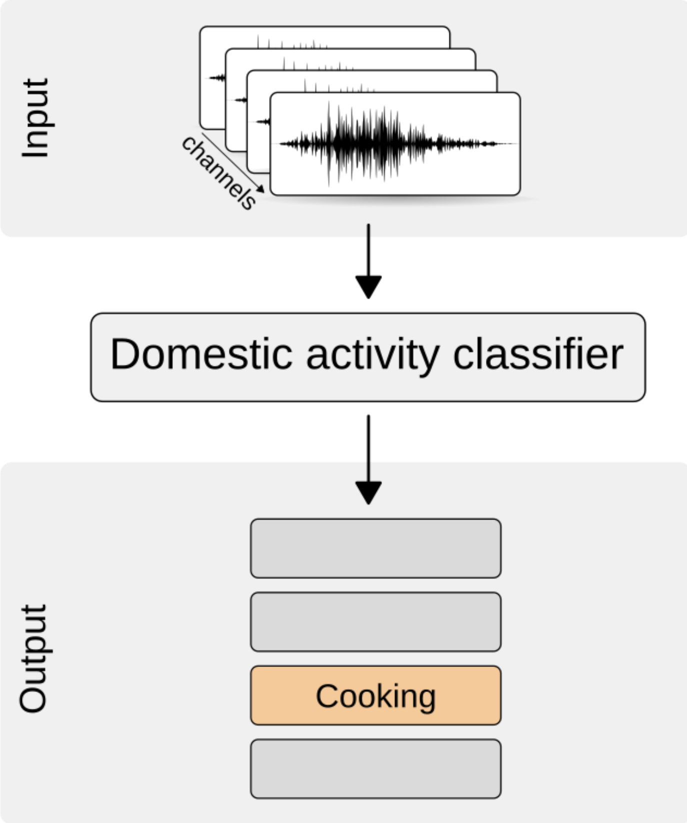
Case study: smart homes

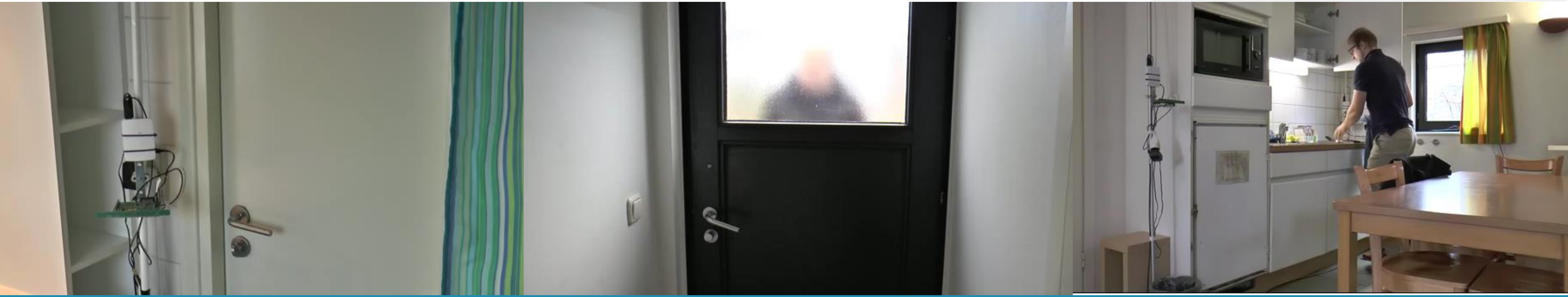
- **Aim:** monitor activities of daily living
  - *cognitive house:*
    - *lower heating when all persons are active e.g. are cooking*
    - *change lighting conditions when all person’s are watching tv*
  - *health care:*
    - *assess self-reliance of elders*



Case study: smart homes

- Task: detect activities of daily living based on acoustics
- First step: collect data

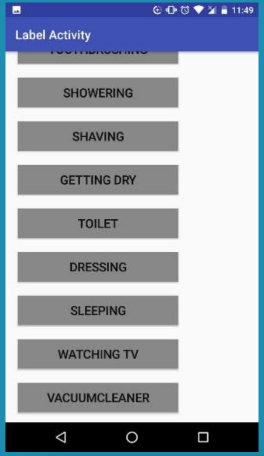


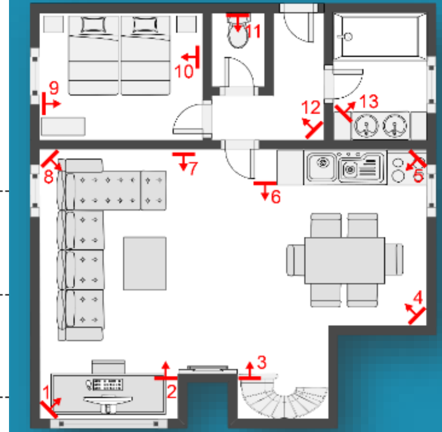


Classification of ADLs

Data set

Room	Activity	Nr. ex.	duration (min.)
Living room	Phone call	22	8.17
	Cooking	19	16.62
	Dishwashing	15	6.37
	Eating	19	7.78
	Visit	9	13.3
	Watching TV	13	155.38
	Working	49	31.24
	Vacuum cleaning	13	4.79
	Other	200	0.75
	Absence	72	66.37
Bathroom	Drying with towel	10	1.67
	Shaving	13	1.91
	Showering	10	6.11
	Toothbrushing	19	1.41
	Vacuum cleaning	9	0.87
	Other	75	0.42
	Absence	35	248.56
Hall	Vacuum cleaning	9	3.31
	Other	164	0.36
	Absence	175	50.17
Toilet	Toilet visit	21	4.74
	Vacuum cleaning	7	0.53
	Absence	31	282.75
Bedroom	Dressing	28	1.53
	Sleeping	7	348.43
	Vacuum cleaning	7	1.04
	Other	22	0.27
	Absence	22	122.28

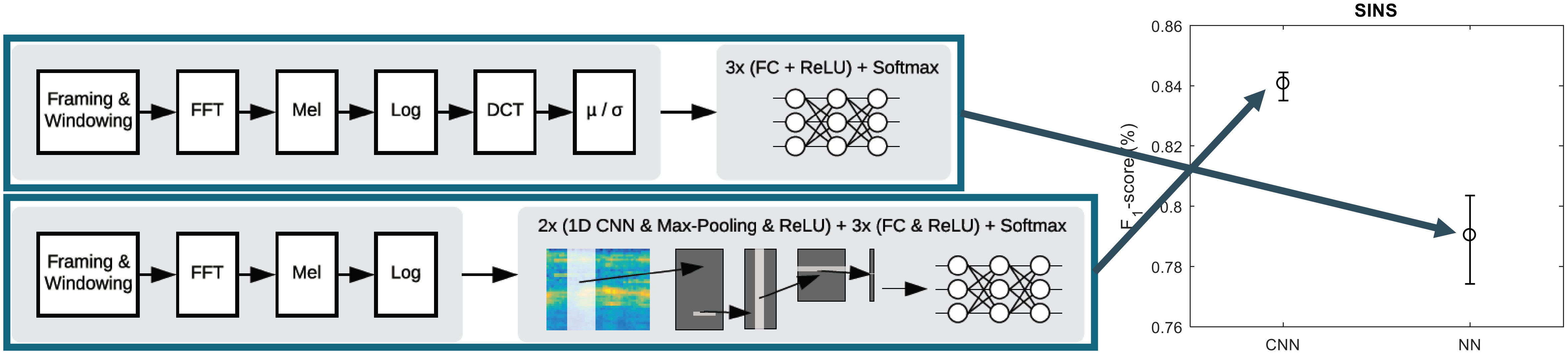




Dekkers et al. (2017). "The SINS database for detection of daily activities in a home environment using an Acoustic Sensor Network" Detection and Classification of Acoustic Scenes and Events 2017, München, Germany, 16-17 November 2017.

Build model: deep learning

Shift from regular machine learning in 2013 into deep learning in 2018



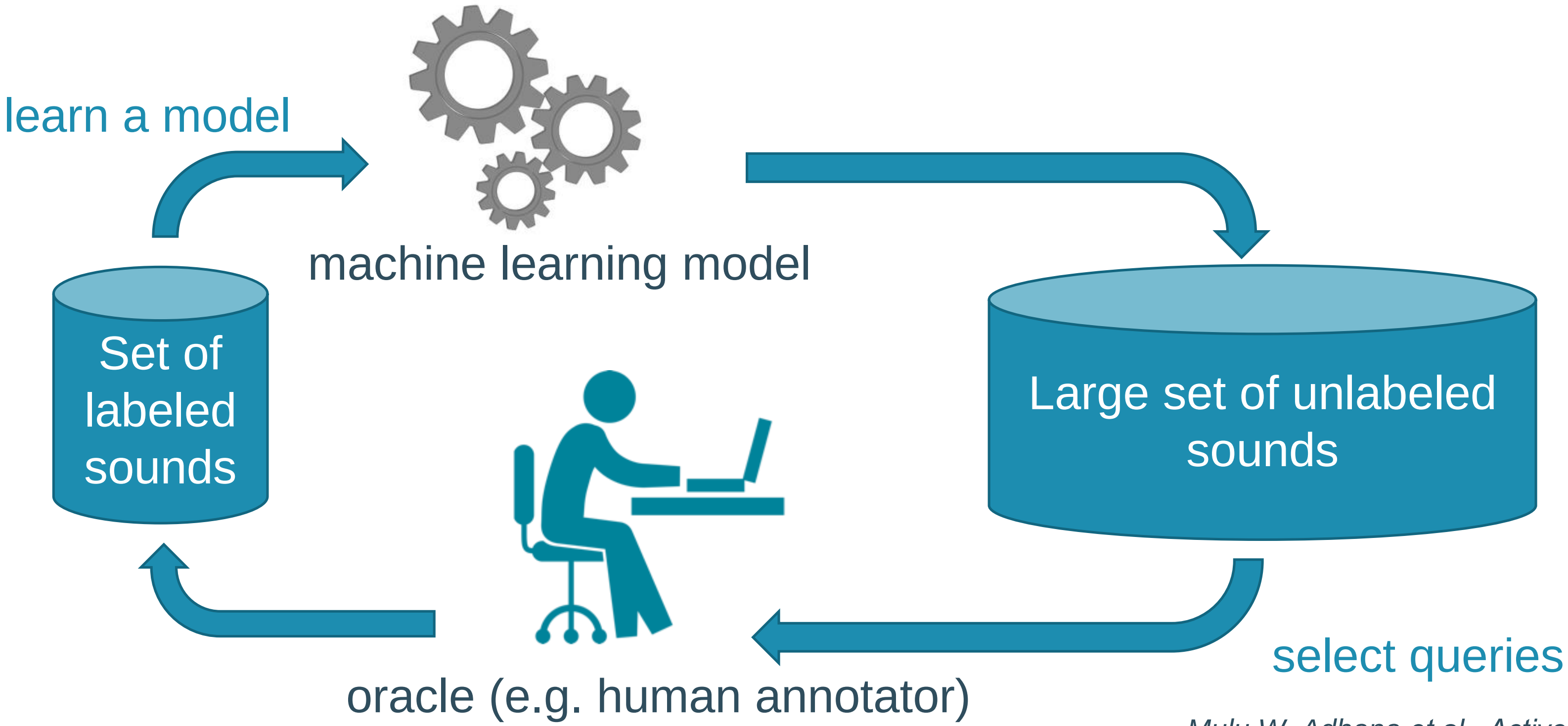
- *Gert Dekkers, Steven Lauwereins, Bart Thoen, Mulu Weldegebreal Adhana, Henk Brouckxon, Toon van Waterschoot, Bart Vanrumste, Marian Verhelst, and Peter Karsmakers. The SINS database for detection of daily activities in a home environment using an acoustic sensor network. In Proceedings of the Detection and Classification of Acoustic Scenes and Events 2017 Workshop (DCASE2017), 32–36. November 2017.*
- *Gert Dekkers, Lode Vuegen, Toon van Waterschoot, Bart Vanrumste, and Peter Karsmakers.DCASE 2018 Challenge - Task 5: Monitoring of domestic activities based on multi-channel acoustics. Technical Report, KU Leuven, 2018. URL: https://arxiv.org/abs/1807.11246, arXiv:1807.11246.*

The pursuit for data

- Supervised deep learning gives very powerful models
- But large amounts of annotated data are required
    - Relax the need on annotated data using
        - semi-supervised learning
        - unsupervised learning

Semi-supervised learning
Active learning

Maximize the value obtained for the expense of human labeling by ensuring they are shown the most important examples.



Mulu W. Adhana et al., Active "Learning for Audio-based Home Monitoring", Benelearn, 2016.

Semi-supervised learning
Opportunistic data collection

Use existing data sources such as YouTube

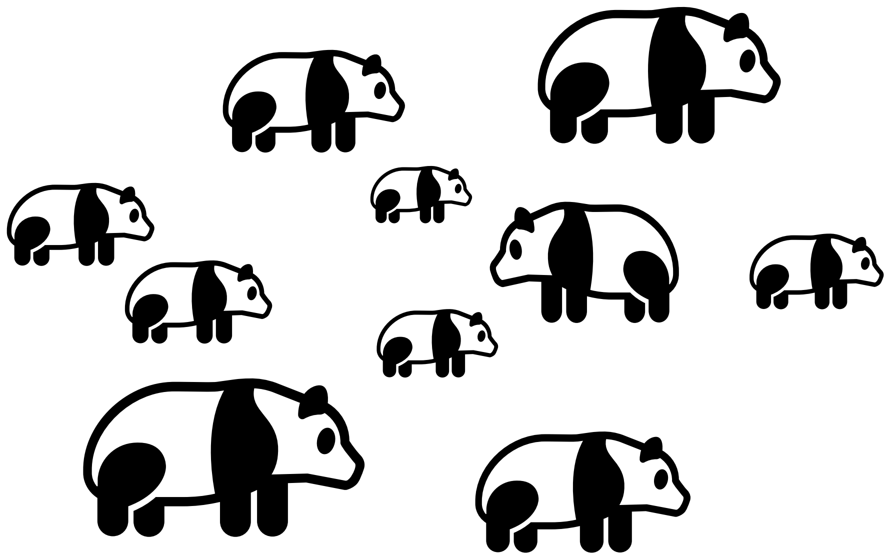
(maybe) somewhere it contains a door slam

weak labels

Vuegen L. et al., "Acoustic Event Classification using Low-resolution Multi-label Non-negative Matrix Deconvolution", Journal of the Audio Engineering Society, 2017

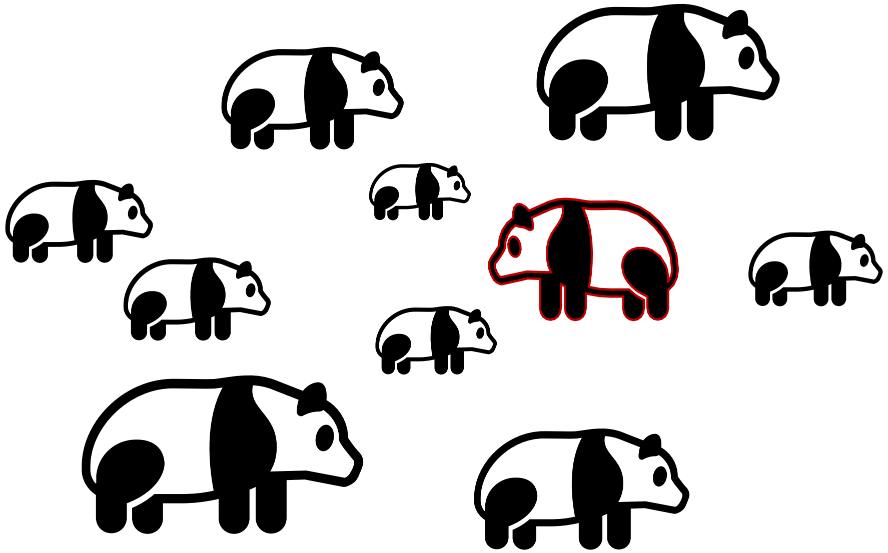
Unsupervised learning
Anomaly detection

Acute anomalous events



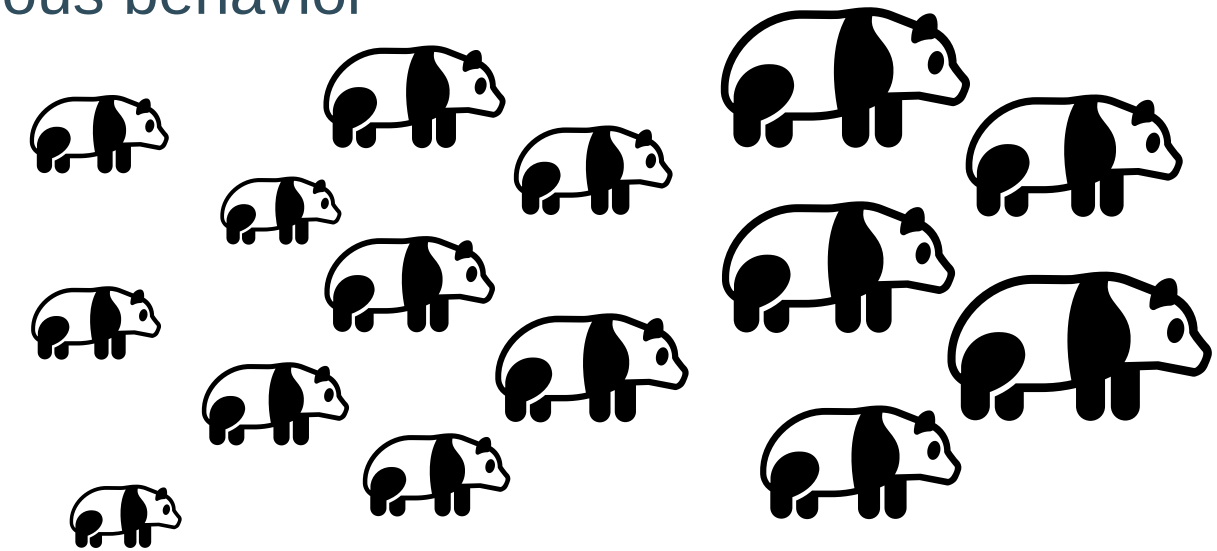
Unsupervised learning
Anomaly detection

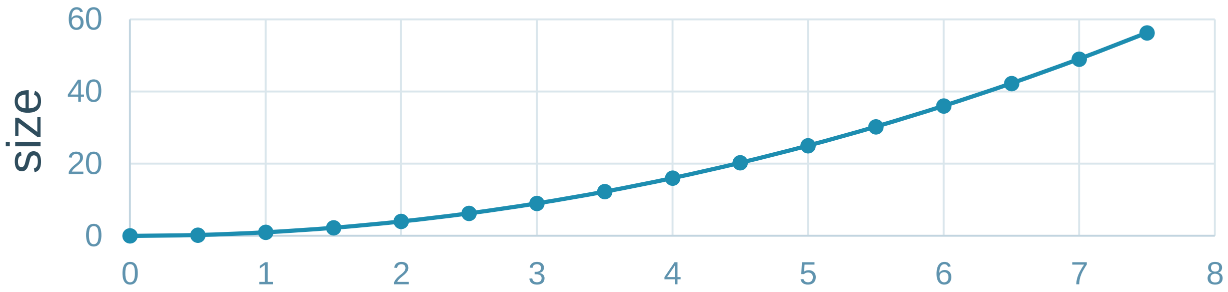
Acute anomalous events

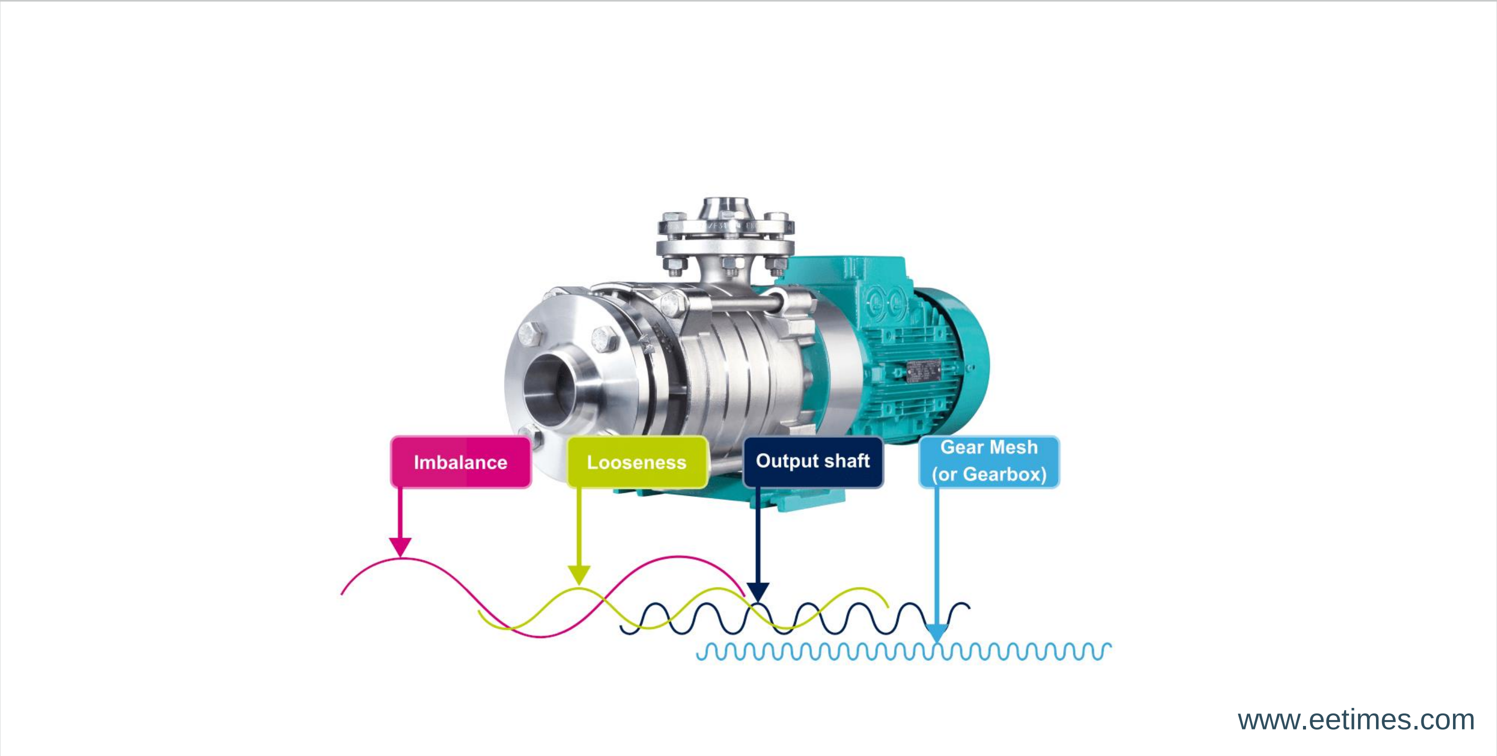


Unsupervised learning

Gradual anomalous behavior





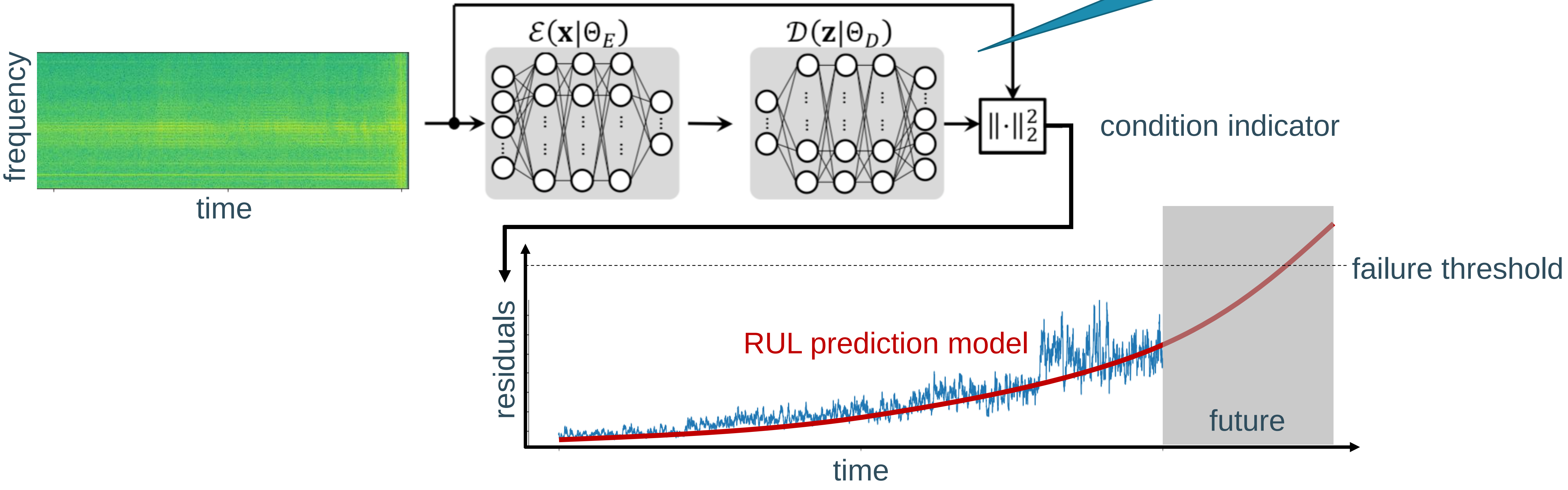


Case study: condition based monitoring

(Gradual) anomaly detection

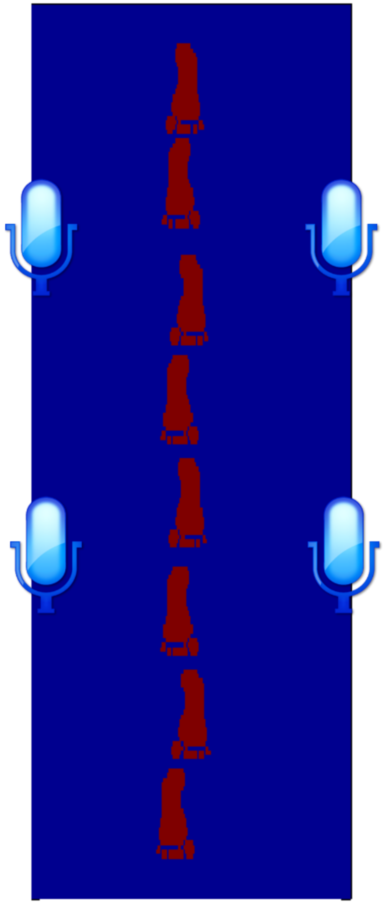
Fix model

No annotation required



Include spatial information

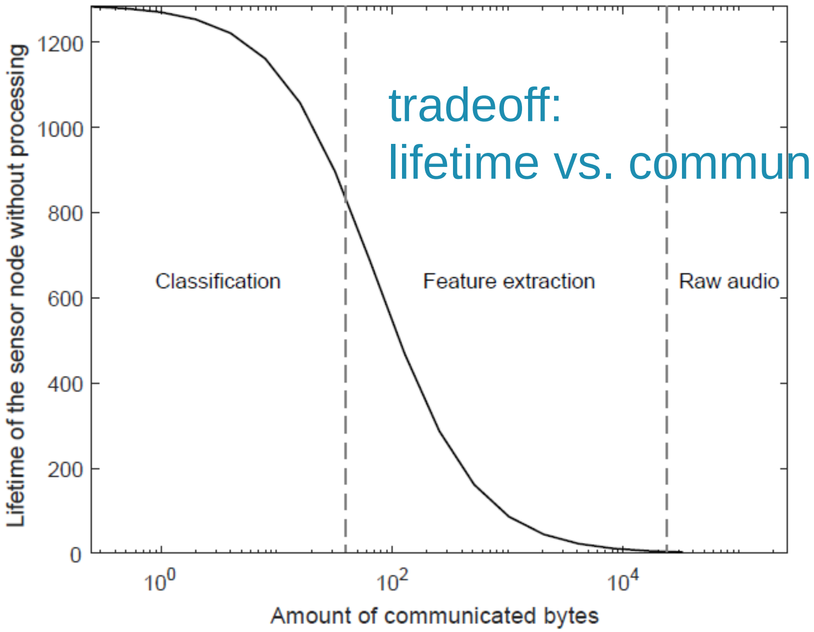
Source
- localization
- tracking
- separation



Bert Van Den Broeck et al. (2016), "Noise robust footstep location estimation using a wireless acoustic sensor network", Journal of Ambient Intelligence and Smart Environments (JAISE)

Other trend:
move intelligence up to the (extreme) edge

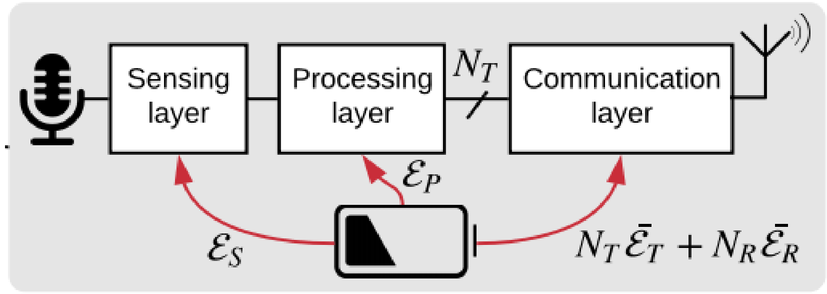
- Why? Assume wireless microphone node



tradeoff:
lifetime vs. communication BW

Energy model

$$\bar{\mathcal{E}}_{\text{node}} = \mathcal{E}_{\text{S}} + \mathcal{E}_{\text{P}} + N_T\bar{\mathcal{E}}_{\text{T}} + N_R\bar{\mathcal{E}}_{\text{R}}$$



- Other reasons: latency, power, privacy, scalability
- Limited resources → need for easy-to-compute models

G. Dekkers, F. Rosas, et al, "A multi-layered energy consumption model for smart wireless acoustic sensor networks," KU Leuven, Tech. Rep., December 2018.

Future research directions

- Transfer learning
- Real-time energy efficient (adaptive) classifier models to be deployed on embedded (IoT) systems
- Data fusion: acoustics combined with other sensing modalities

To conclude

Acoustic event detection: emerging research field

Several potential applications

Scientific challenges: robust classification, dealing with overlapping sounds, reverberation

Practical challenges: acquisition of annotated data, computing power

Deep learning enables to search for suitable representations and give state of the art performance

We can safely assume high-accuracy automatic sound scene and event recognition in the near future

Thanks to

Mulu Weldegebrael Adhana
Active learning of probabilistic classifier models

Gert Dekkers
Robust and energy efficient audio-based classification systems

Yonas Yehualashet Tefera
Anomaly detection in (accelerometer- and audio-based) time-series data

Maarten Meire
Acoustically based rotational machine breakdown prediction using deep learning strategies

Lode Vuegen
Acoustical classification and descriptive models for human monitoring