Mobile and Sensor Systems

Lecture 5: Sensor Data Inference

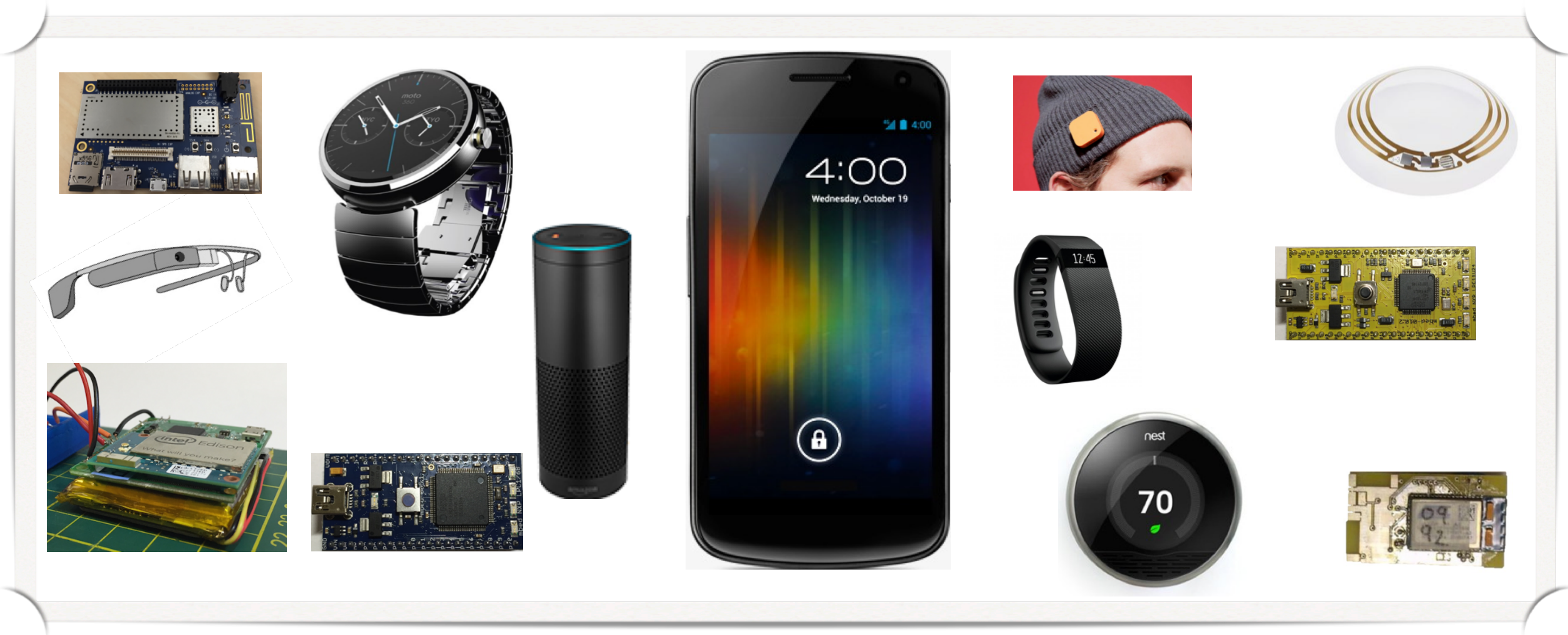
Prof C Mascolo

UNIVERSITY OF CAMBRIDGE

In this lecture

- We will talk about mobile sensing
- We will describe the challenges in sensor inference
- We will talk about the steps involved in the sensor to inference process
- We will introduce traditional feature based and neural network based modelling

Mobile and Wearable Sensing



Sensors

- Microphone
- Camera
- GPS
- Accelerometer
- Compass
- Gyroscope
- WiFi
- Bluetooth
- Proximity
- Light
- NFC (near field communication)

UNIVERSITY OF CAMBRIDGE

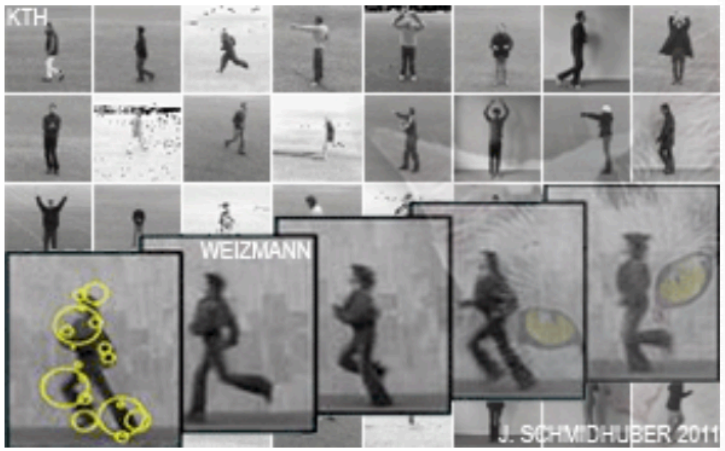
Applications

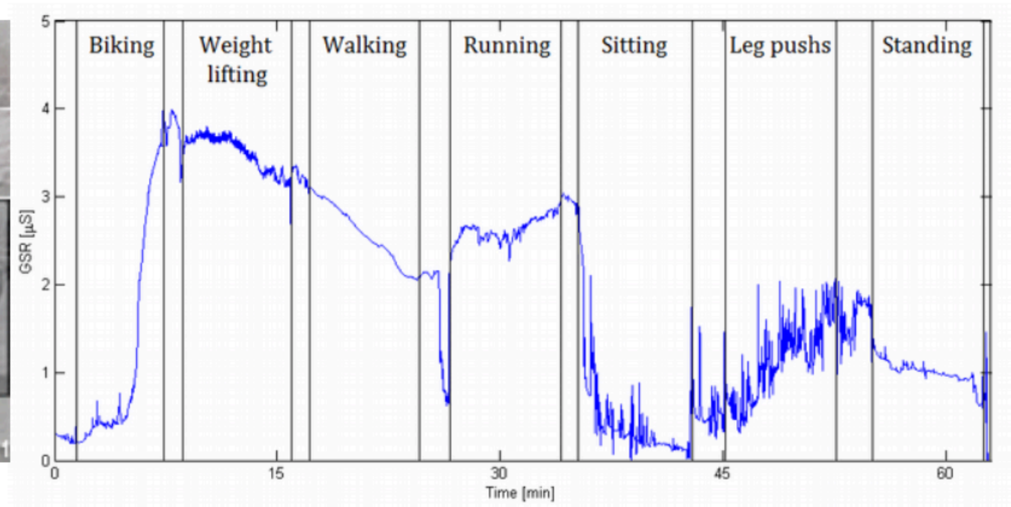
- Individual sensing:
  - fitness and health applications
  - behaviour intervention applications
- Group/community sensing:
  - groups to sense common activities and help achieving group goals
  - examples: assessment of neighbourhood safety, environmental sensing, collective recycling efforts
- Urban-scale sensing:
  - large scale sensing, where large number of people have the same application installed
  - examples: tracking speed of disease across a city, congestion and pollution in a city

Sensor Based Inference Systems Characteristics

- Offline/Online Inference
- Continuous/Periodic/Isolated Inference
- In all cases, collecting ground truth is key:
  – data needs precise labels…

[Bulling et al 2013]



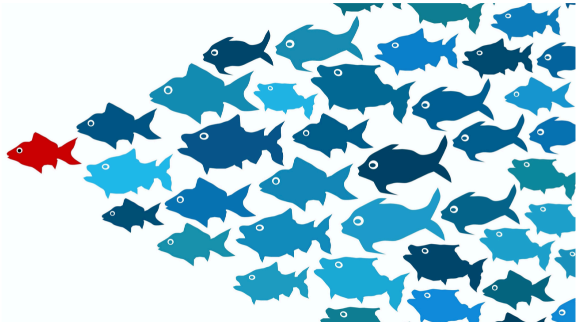


Sensor Data Analysis Challenges: Differences in Users and Device Positions

- Users have differences which influence the readings on the sensors (eg different gait)

Sensor Data Analysis Challenges: Range of Activities

- Activities (classes) can often be many and sometimes very similar (or similar for some users)
- Collection of balanced data among those classes (for ground truth) can be challenging
  - The nature of the data collection might impose imbalances in the data
- Ground Truth Annotation is hard….



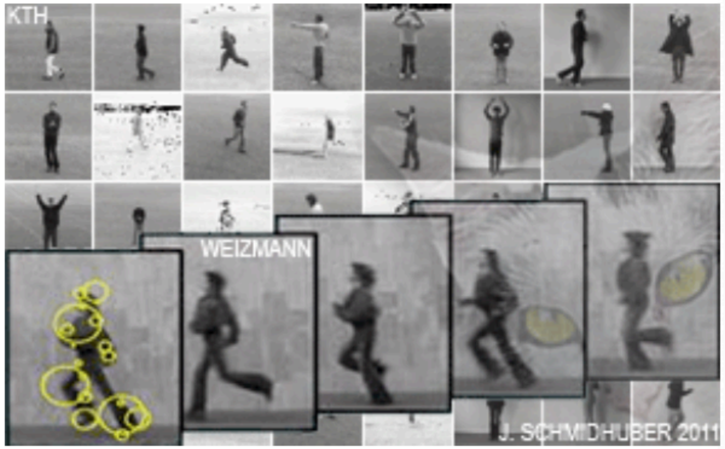
UNIVERSITY OF CAMBRIDGE

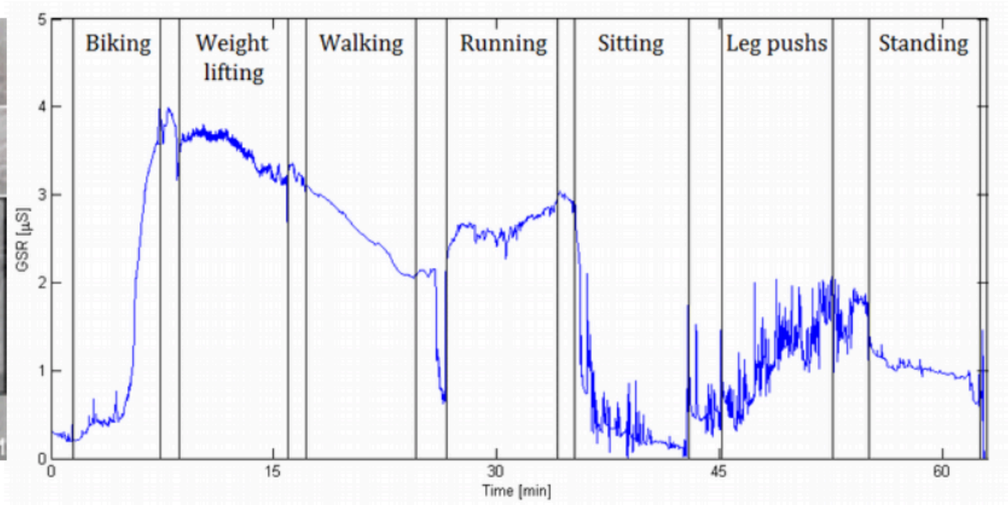
Activity Recognition from Sensor Data

- Activity recognition aims to recognize the **actions of an individual** from a **series of observations** on the individual's actions and the **environmental conditions.**

UNIVERSITY OF CAMBRIDGE

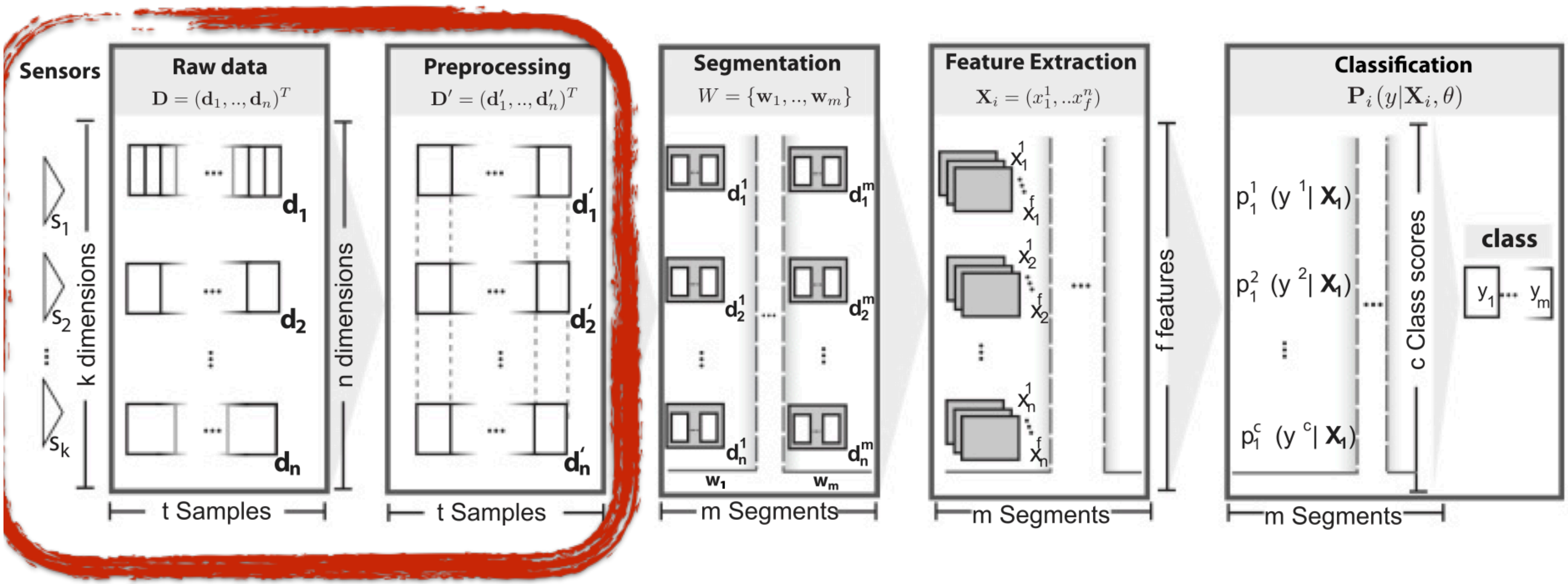
Activity Recognition





- Wearables and mobiles produce sequential data

Sensor Inference Pipeline



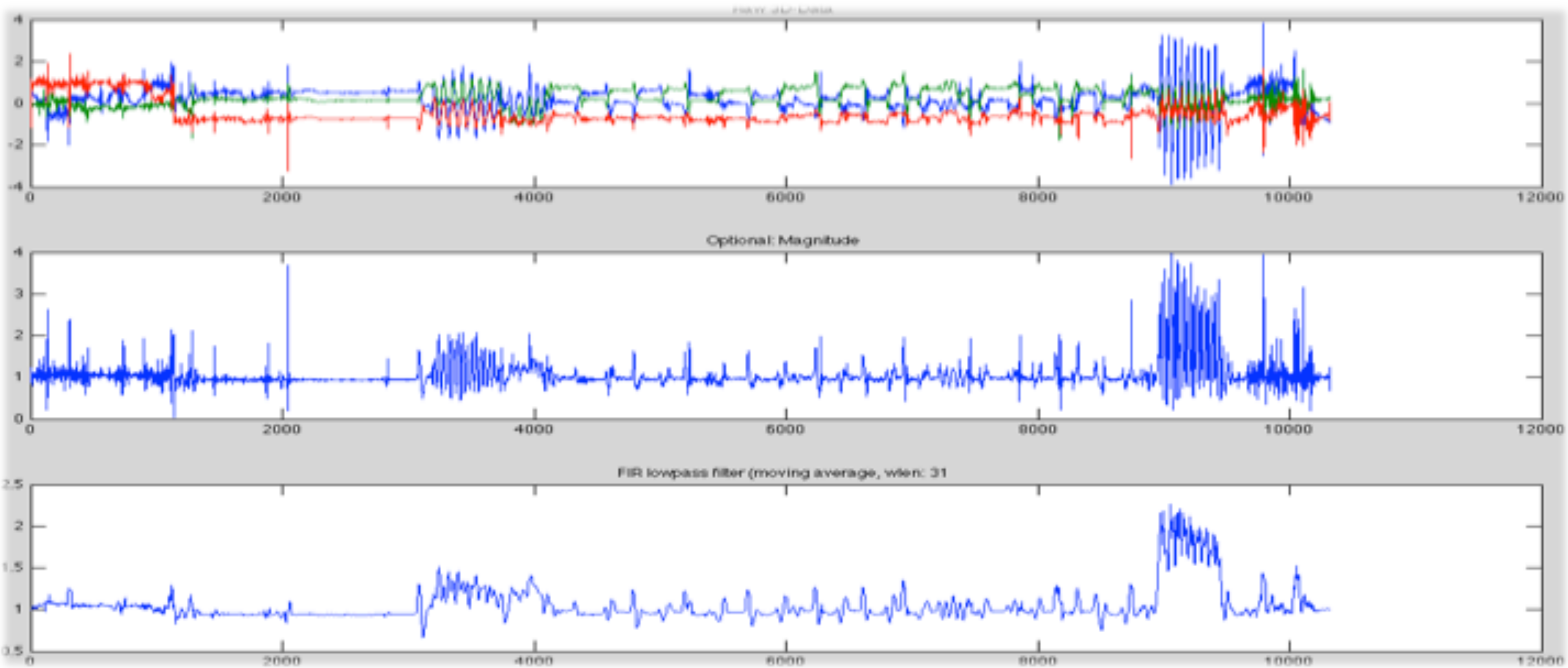
Sensor Raw Data and Preprocessing

- Acquisition can happen
  - from different sensors (at different locations on the body or orientation, or from different sensors, acceleration or GPS…)
  - At different sampling rate (eg for energy reasons)
- Sensor data can be corrupted or contain errors
- Preprocessing synchronizes and removes artifacts (calibration, unit conversion, normalization, resampling, synchronization..)

Accelerometer Preprocessing

- Magnitude

$$\forall i : m_i = \sqrt{x^2 + y^2 + z^2}$$



Sampling Rate

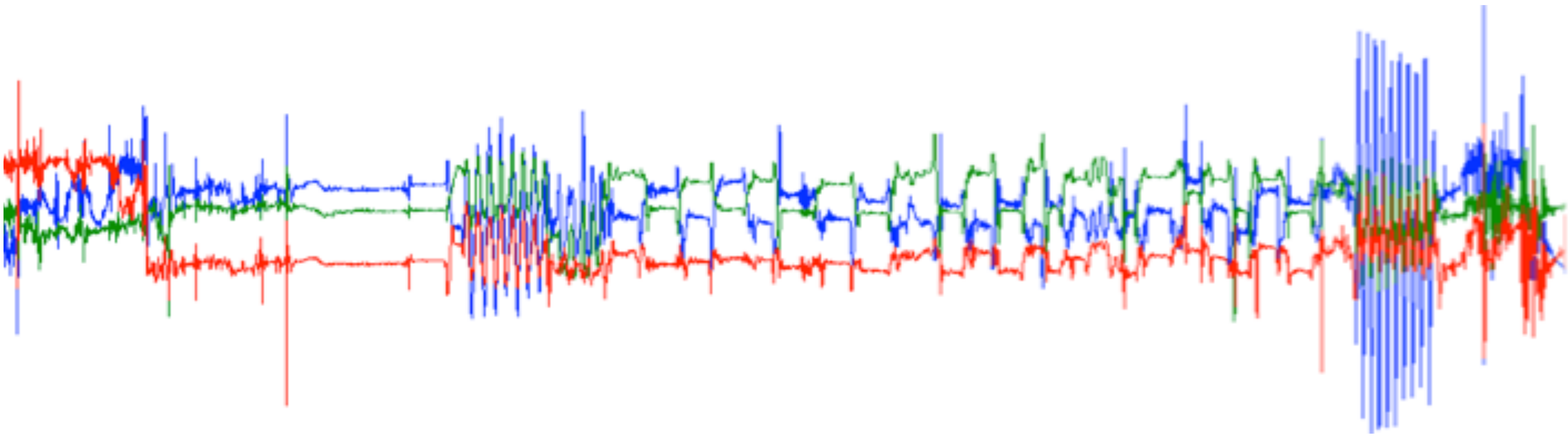
Dataset	#Classes	Q	q (S=0.99)	Δ
Skoda	11	96Hz	12Hz	-87.5%
PAMAP2-Hand	13	100Hz	32Hz	-68%
PAMAP2-Chest			33Hz	-67%
PAMAP2-Ankle			42Hz	-58%
USC-HAD	12	100Hz	17Hz	-83%
PHealth	10	100Hz	15Hz	-85%
Walk8	4	250Hz	18Hz	-92.8%

UNIVERSITY OF CAMBRIDGE

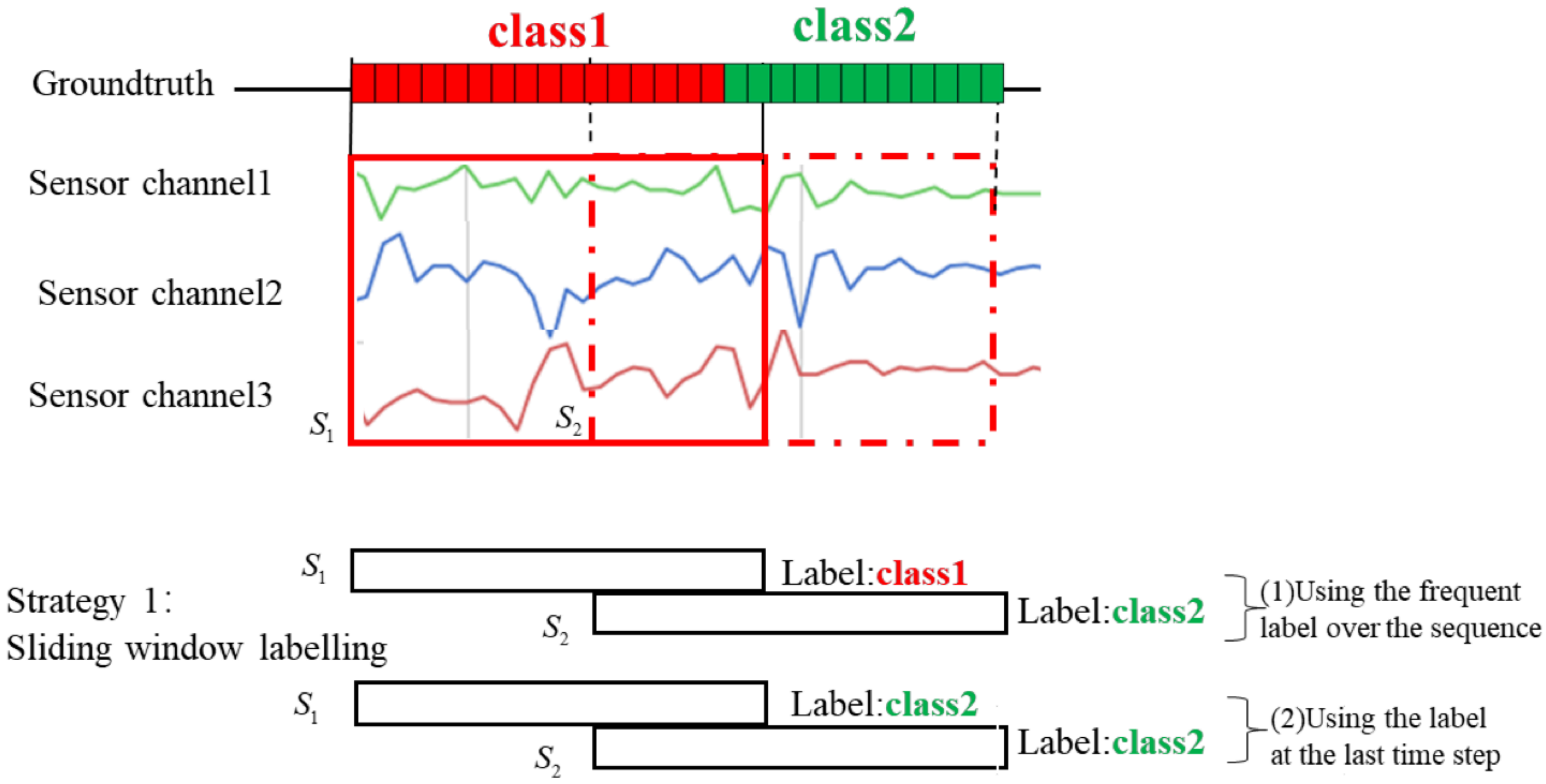
[Khan et a 2016]

Data Segmentation

- Localize temporal patterns of interest
- But you do not know what/where these are…
- Sliding window approach
  - Issues: window length, stride, window label choice…



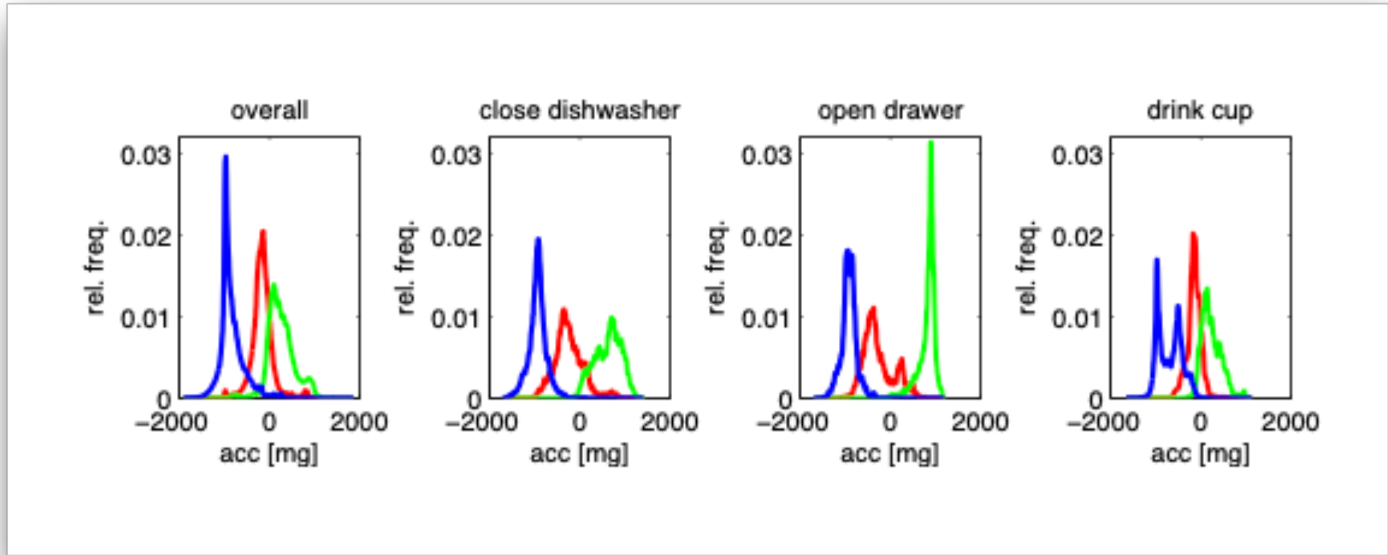
Data Segmentation Example



UNIVERSITY OF CAMBRIDGE

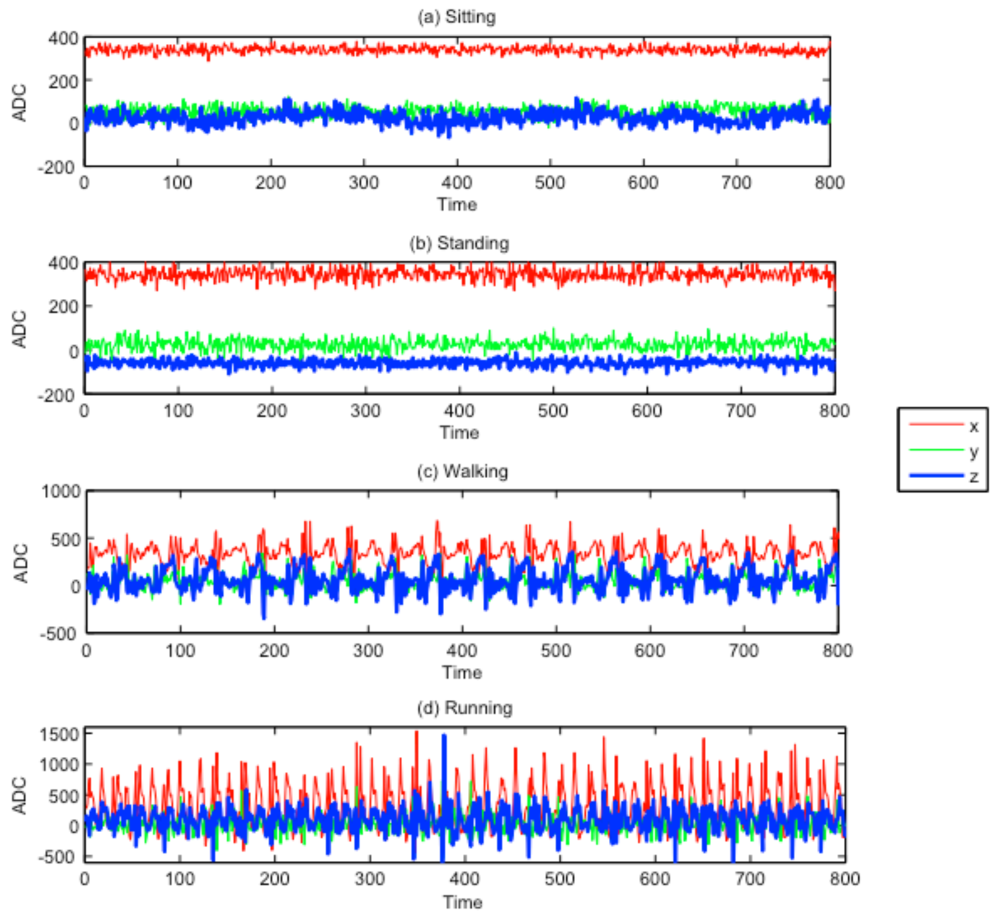
https://arxiv.org/pdf/1809.08113.pdf

Feature Extraction
Example: Activity Recognition



Physical Activity using Accelerometer

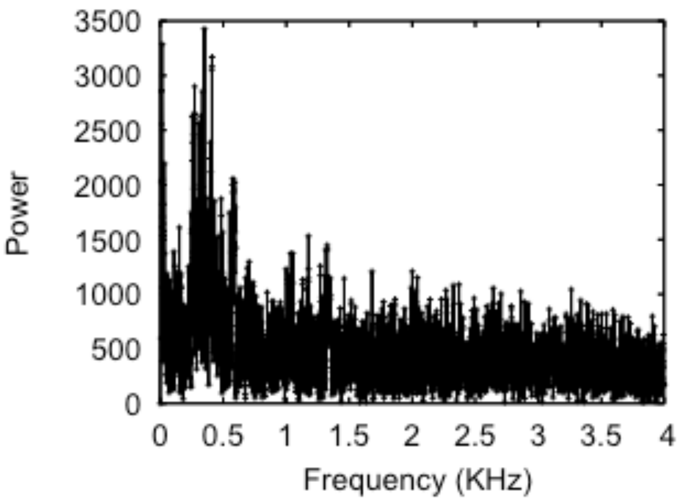
- Sensor: accelerometer
- Activities: sitting, standing, waking, running
- Features:
  - Mean (can help distinguish between standing and sitting).
  - Standard deviation
  - Number of peaks (can help distinguish between waking and running).



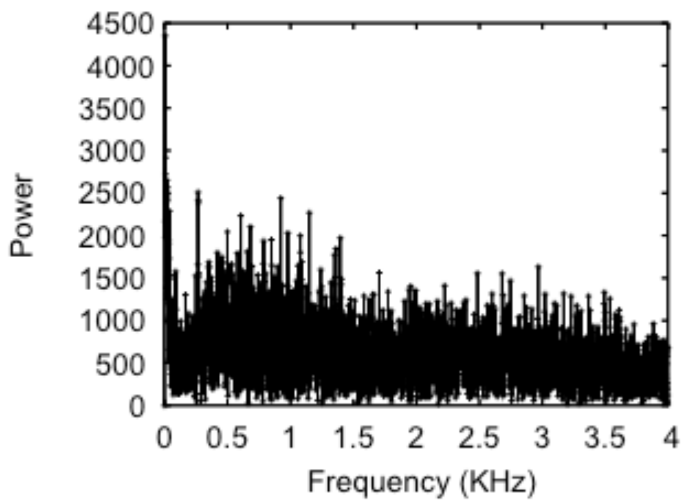
Feature Extraction: Conversation Detection

- FFT (Fast Fourier Transform) of audio (from microphone)

Human voice



Noisy environment

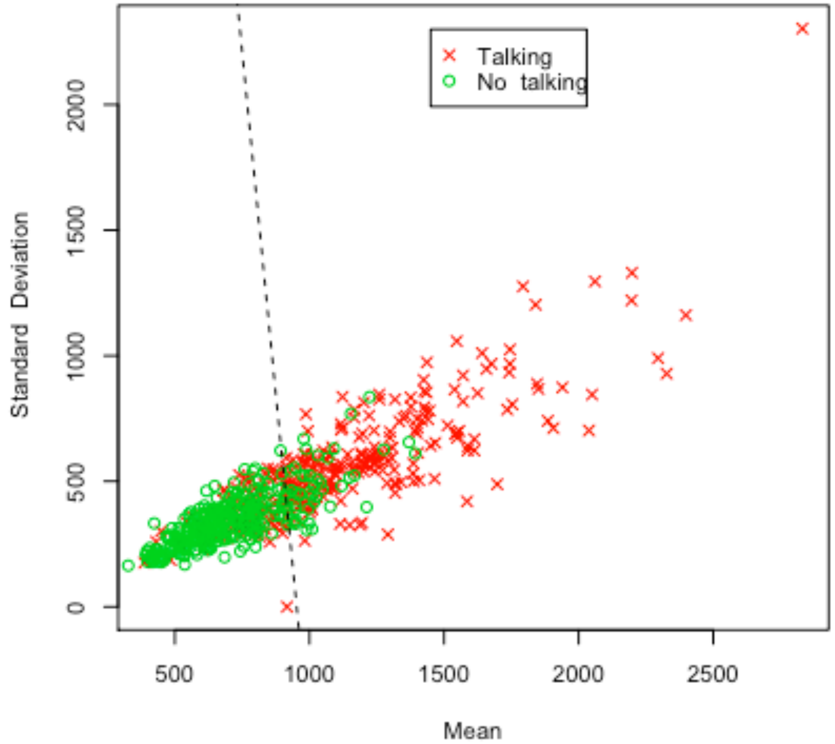


- Sound samples of human voice present most of their energy within the 0-4 KHz spectrum.

UNIVERSITY OF CAMBRIDGE

Feature Extraction: Conversation Detection

- Selecting as Features the mean and standard deviation of the FFT power

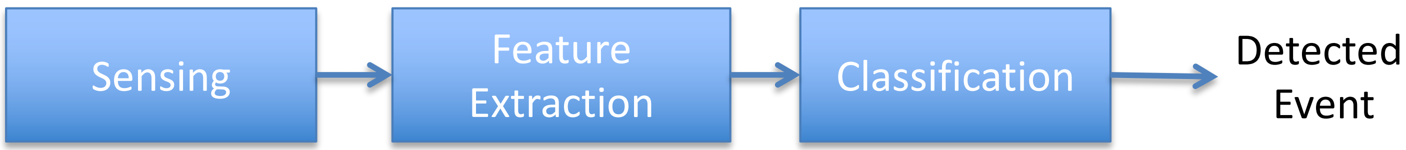


- Using a simple threshold line, could give a relatively accurate detection (with a high number of false positives, however)

UNIVERSITY OF CAMBRIDGE

Inference

- The process of mapping raw sensor data to meaningful high-level events. Inference Pipeline:



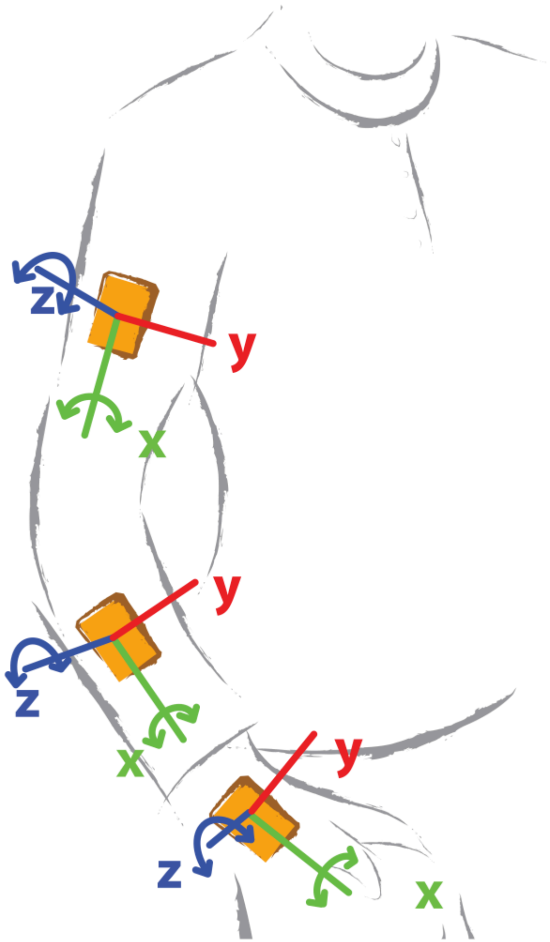
- Designing an Inference Engine:
  - Collecting raw sensor data, typically labelled with ground truth information.
  - Data set should also cover states we are not trying to detect but look similar (e.g. detect *walking* : we need data also for *running* and *standing*).
  - Train the inference engine with the collected data.
  - Applying the inference engine to the target application.

UNIVERSITY OF CAMBRIDGE

Classification

- Feature extraction produces a feature vector.
- The classification matches the feature vector to a pre-defined set of high-level classes.
- The classification engine is typically based on machine-learning techniques and is trained using labelled training data.
- Common classification algorithms include:
  – K Nearest Neighbour.
  – Naive Bayes classifier.
  – Decision Trees.
  – Hidden Markov Models.

Activity Recognition Classification



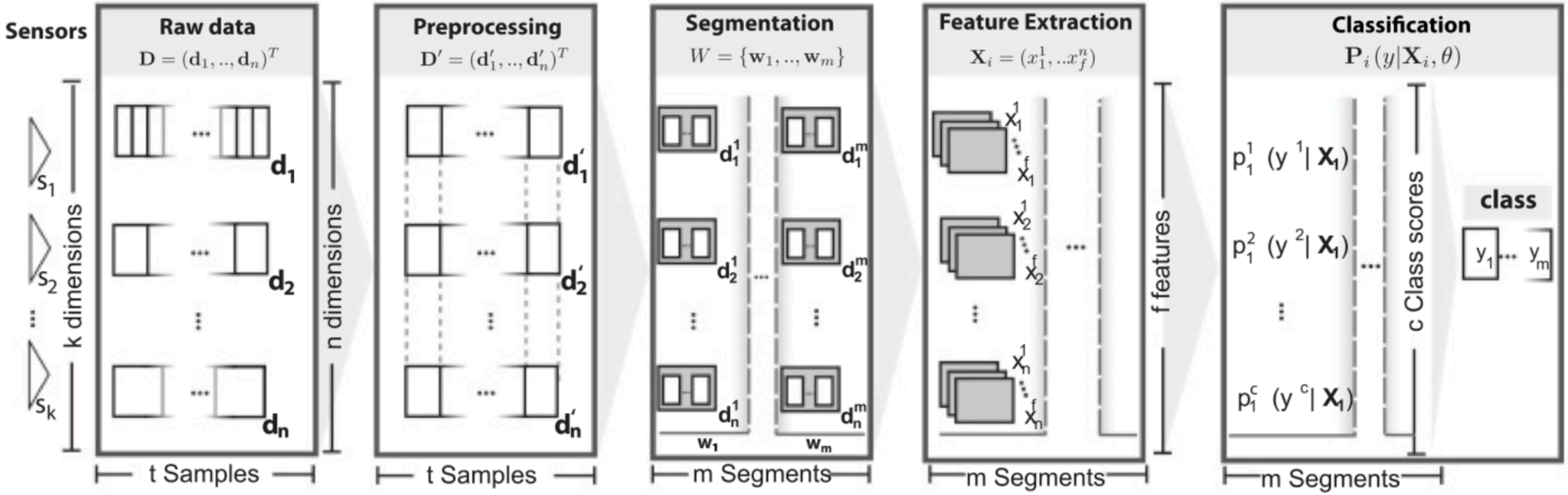
Activities

opening a window
closing a window
watering a plant
turning book pages

drinking from a bottle
cutting with a knife
chopping with a knife
stirring in a bowl

forehand
backhand
and smash

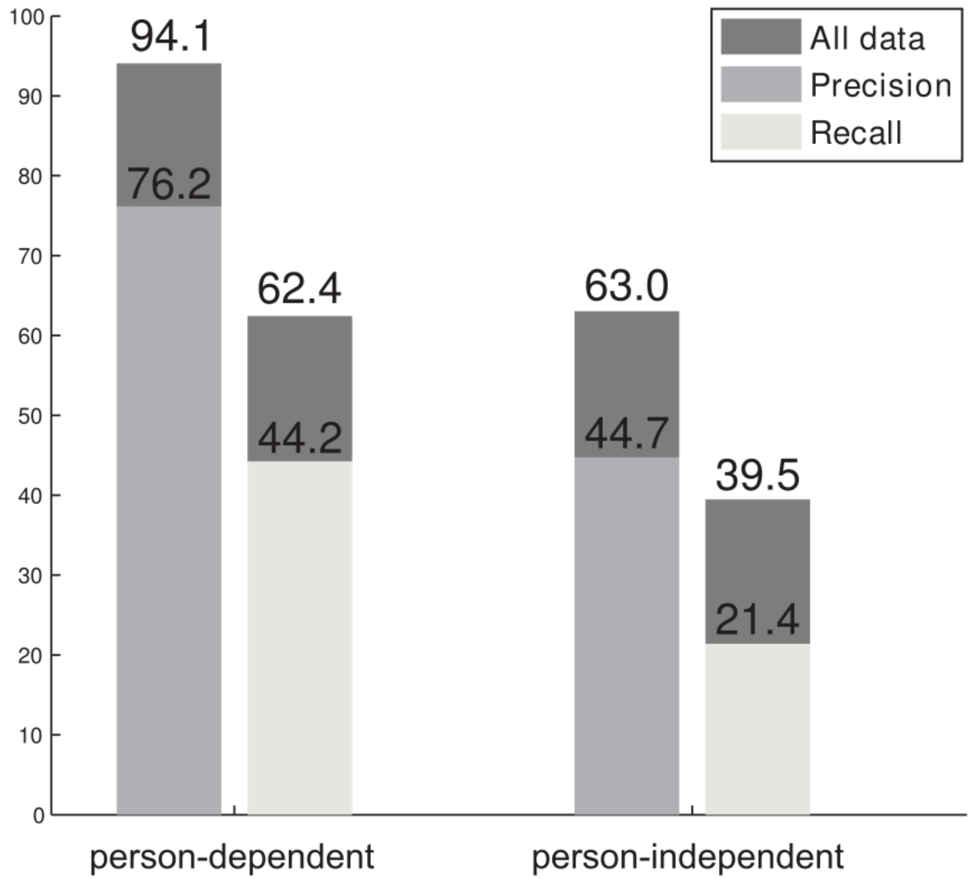
UNIVERSITY OF CAMBRIDGE



Data Acquisition	Segmentation	Feature Calculation	Modeling & Inference	Classification
ACC_{Wrist} →	Fixed-Size (1s) Sliding Window →	Mean, Variance →	Nearest Class Centroid (NCC) →	$\arg\max_{y \in \mathcal{Y}} \; p_i(y|\mathbf{X}_i, \theta)$

UNIVERSITY OF CAMBRIDGE

Classification Results: Person Dependence and Multiple Sensors



UNIVERSITY OF CAMBRIDGE

Confusion Matrix on Activities

classification

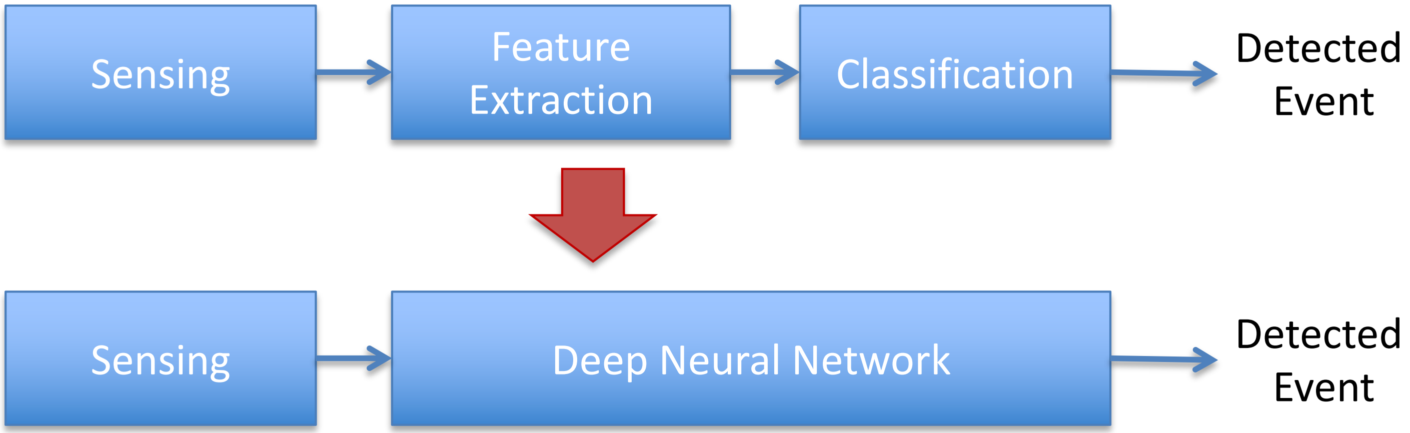
groundtruth

	NULL	Open window	Drink	Water plant	Close window	Cut	Chop	Stir	Book	Forehand	Backhand	Smash	*recall*
NULL	**24267**	216	444	3228	48	24	60	75	45		3		**85.42**
Open window	3849	**1938**	453	291	48	12	9		24				**29.26**
Drink	3984	927	**3780**	321	3	9							**41.89**
Water plant	3984	726	774	**3735**	21	57	15						**40.11**
Close window	3891	381	1173	945	**1533**								**19.35**
Cut	2940		264	450		**6585**	456		3				**61.55**
Chop	2895	168	435	153		909	**5742**		126				**55.06**
Stir	4947	39	135	42	21	474	561	**4392**	207				**40.60**
Book	4560	27	144	951		354	1725	60	**6687**				**46.09**
Forehand	3195	330		144	609	9	66		3	**969**	6	3	**18.17**
Backhand	3003	207	21		21	3	6	24	33		**1302**		**28.18**
Smash	1860	57		78	185		42	45		1567	137	**230**	**5.47**
precision	**38.29**	**38.64**	**49.59**	**36.13**	**61.59**	**78.06**	**66.14**	**95.56**	**93.81**	**38.21**	**89.92**	**98.71**	

UNIVERSITY OF CAMBRIDGE

Beyond Features: Deep Learning

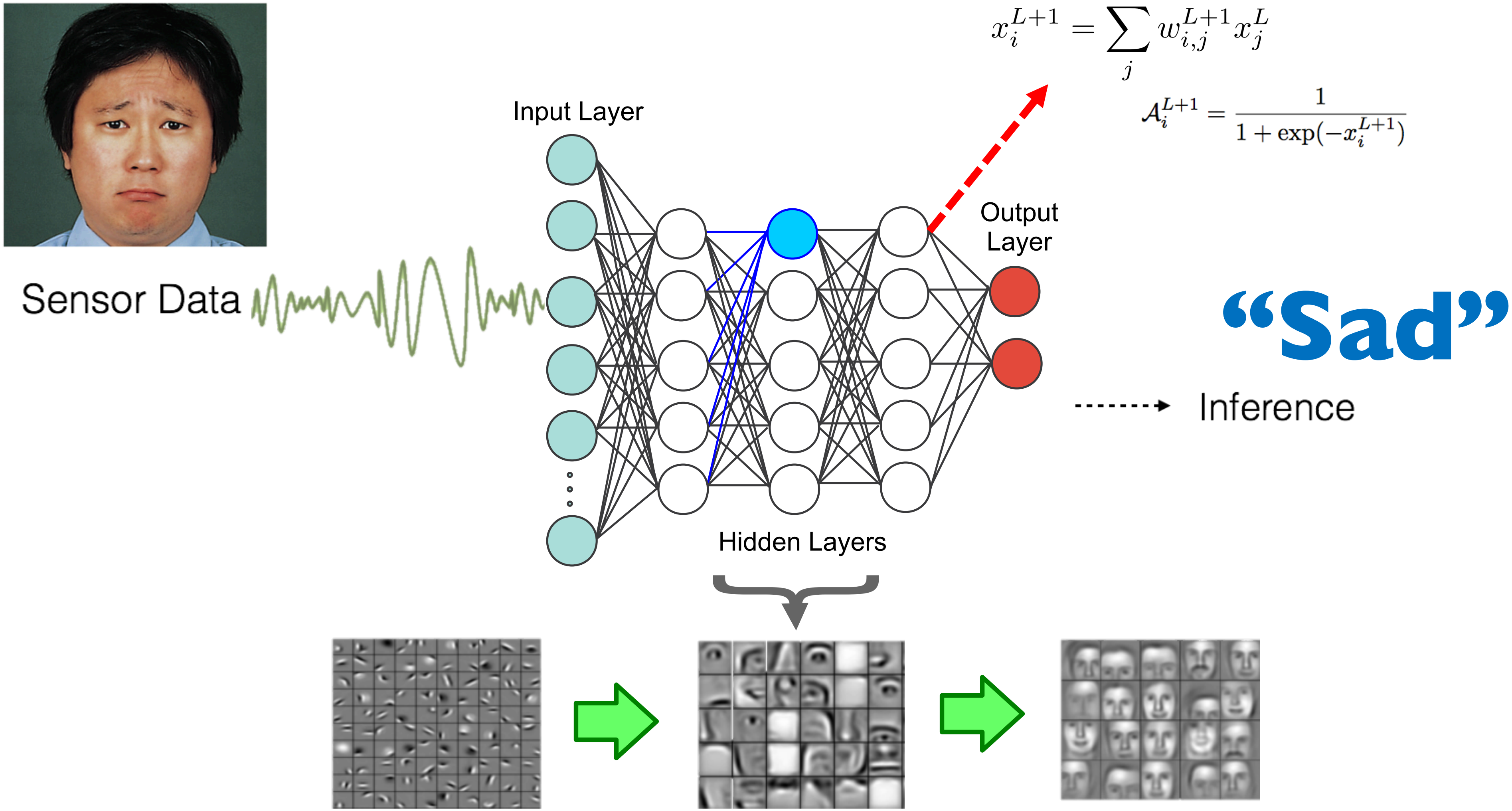
- Movement away from hand-crafted experimentally driven features towards models that combine feature and classification phases



- Paradigm of learning discriminative representations (“feature representation learning”) directly from large amounts of relatively raw data (“end-to-end learning”)
- Modelling techniques (e.g., training algorithms, network architecture) are less tied to specific domains and tasks.

UNIVERSITY OF CAMBRIDGE

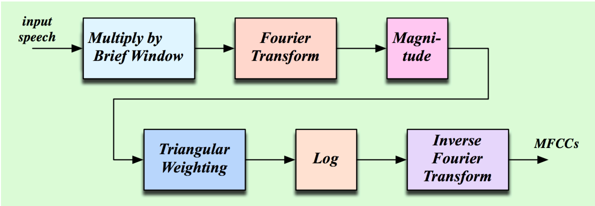
Representative Deep Neural Network



UNIVERSITY OF CAMBRIDGE

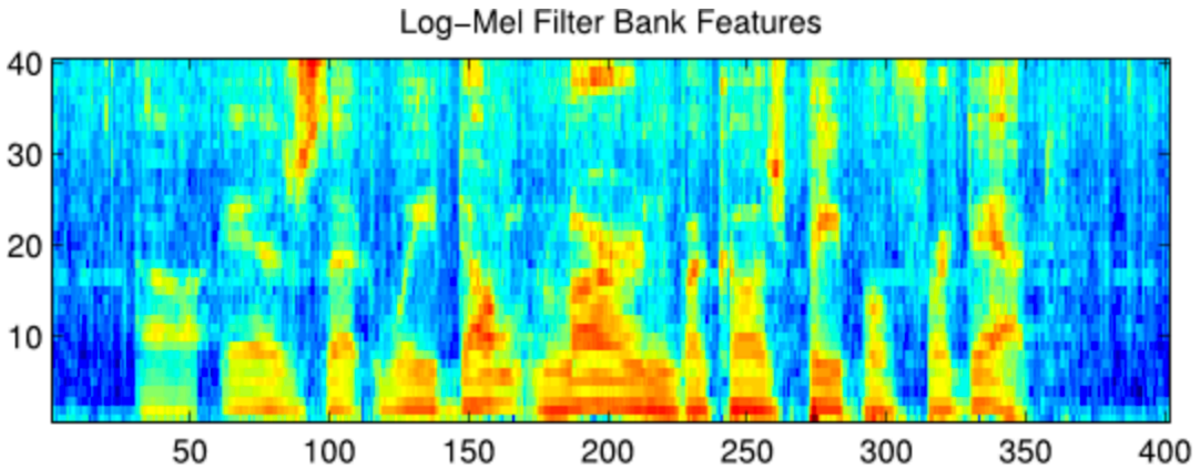
- Just one of dozens of types of deep learning that exist (CNNs, RNNs, etc.)

Example of
Feature Representation Learning

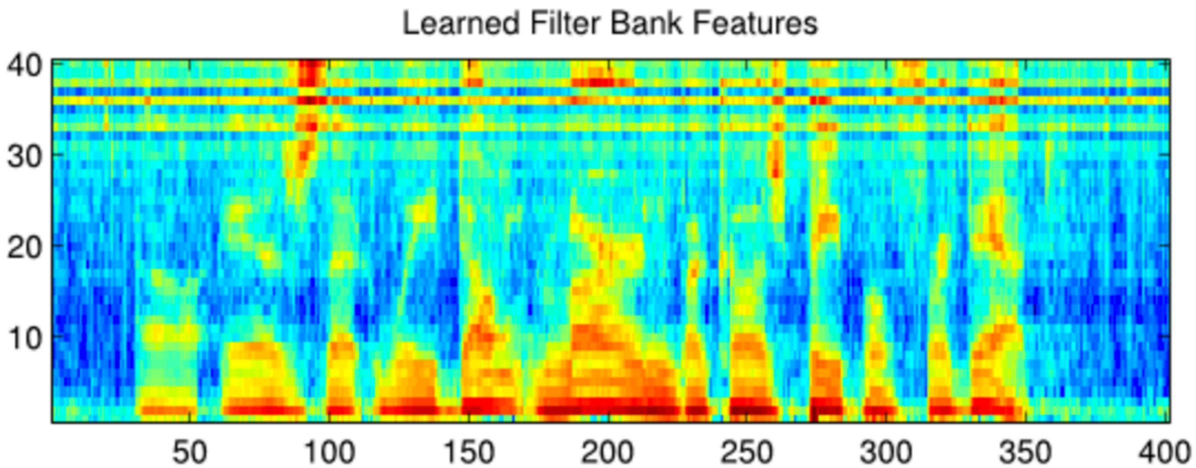


(1) Mel frequency cepstral coefficients

- Result of decades of research into audio
- Dominant general purpose audio representation



(2) Representation learned directly from data



UNIVERSITY OF CAMBRIDGE

DL and AR

- Unlike other domains sensor data lacks large scale labelled datasets
  – Difficult to collect large scale ground truth
- Can lead to overfitting! (DNN with many parameters will memorize small data)
- Solutions:
  – Transfer learning
  – Classifier ensembles

UNIVERSITY OF CAMBRIDGE

No best architecture for DNN in AR

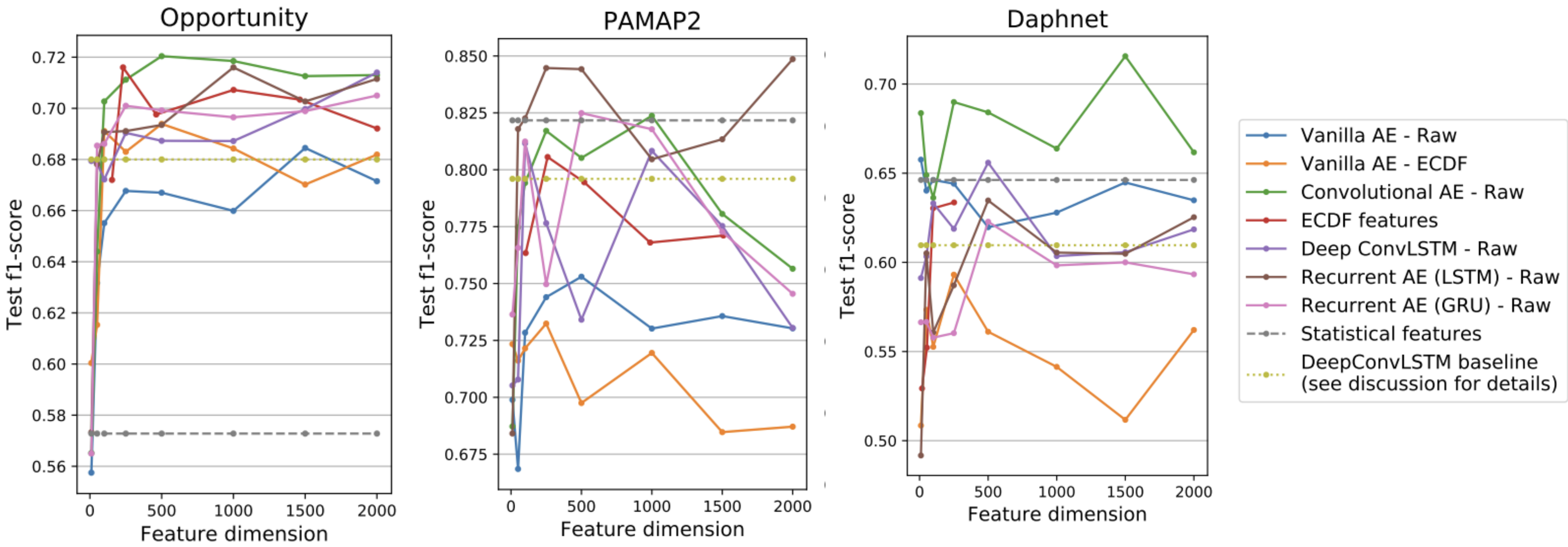


Figure 3: Classification performance (F1) results.

Opportunity Dataset: wide range of activities. Convolutional autoencoder best

Statistical features perform well and have good performance in resource constrained conditions. Next lesson!

[Haresamudram 19]

UNIVERSITY OF CAMBRIDGE

[Hammerla et al 2016]

DNN in Activity Recognition

	PAMAP2	DG	OPP	
Performance	F_m	F_1	F_m	F_w
DNN	0.904	0.633	0.575	0.888
CNN	**0.937**	0.684	0.591	0.894
LSTM-F	0.929	0.673	0.672	0.908
LSTM-S	0.882	**0.760**	0.698	0.912
b-LSTM-S	0.868	0.741	**0.745**	**0.927**
CNN	[Yang *et al.*, 2015]		–	0.851
CNN	[Ordóñez and Roggen, 2016]		0.535	0.883
DeepConvLSTM	[Ordóñez and Roggen, 2016]		0.704	0.917
Delta from median	ΔF_m	ΔF_1	ΔF_m	mean
DNN	0.129	0.149	0.357	0.221
CNN	**0.071**	**0.122**	0.120	**0.104**
LSTM-F	0.10	0.281	0.085	0.156
LSTM-S	0.128	0.297	**0.079**	0.168
b-LSTM-S	0.087	0.221	0.205	0.172

Bidirectional LSTMs considering temporal nature of the data worked very well on a large dataset (OPP). CNN work well on short term movement patterns. RNN better than CNN on short activities with temporal ordering: RNN contextualizes over longer timescales.

Semi Supervised Approaches for AR

- Semi-supervised learning methods can achieve similar performance to fully supervised ones using only a fraction of the labels

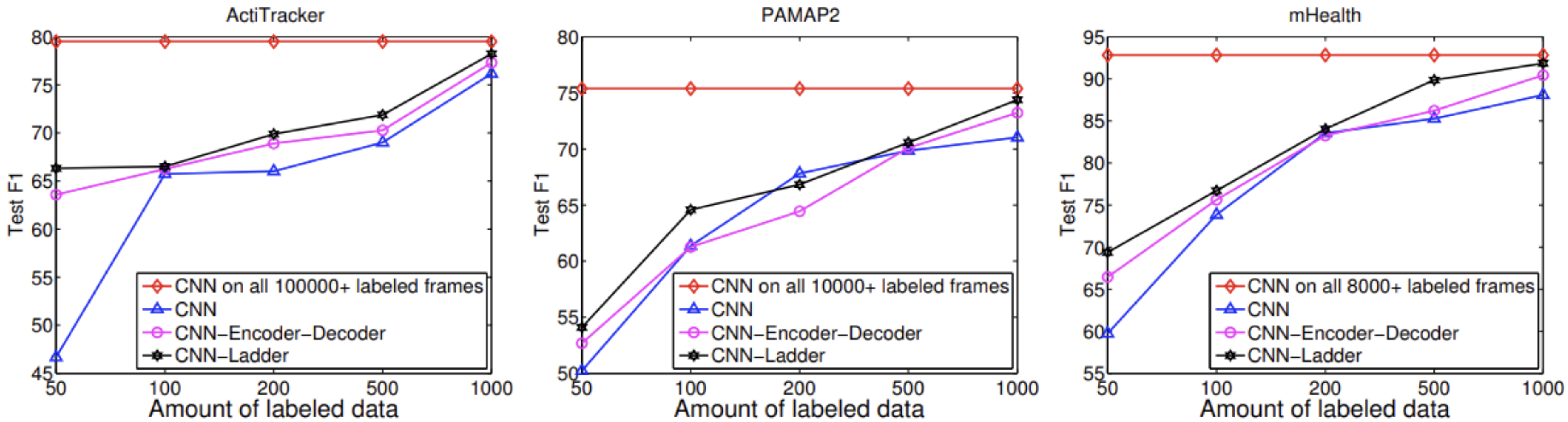
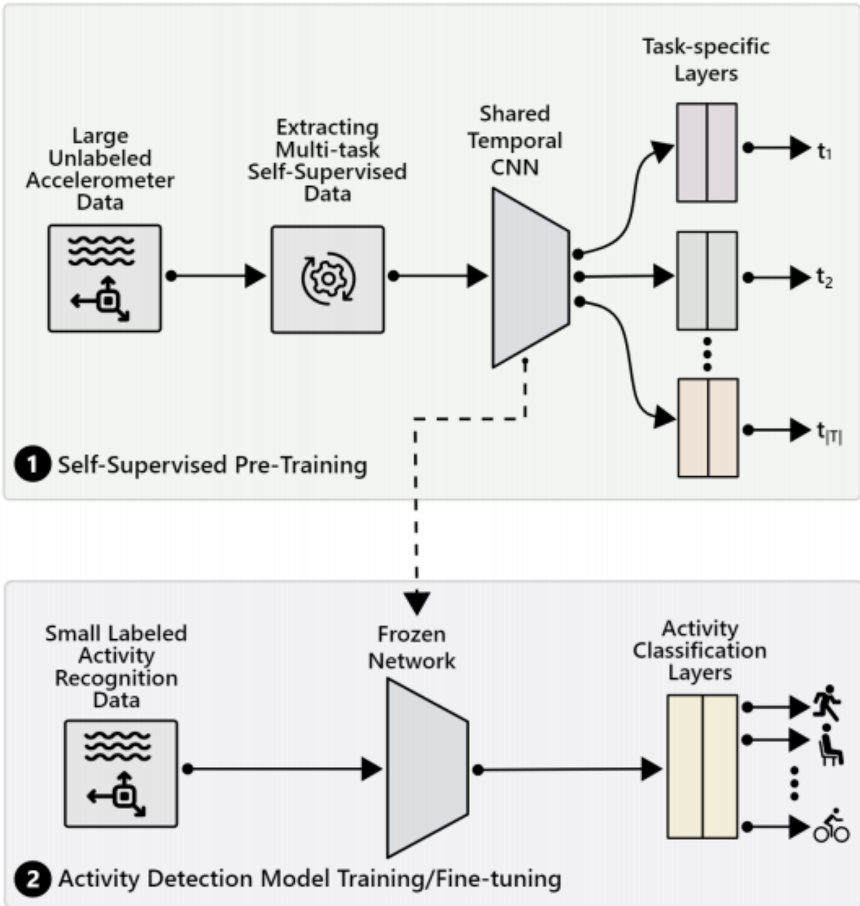


Figure 2: The F_m scores of CNN, CNN-Encoder-Decoder, and CNN-Ladder, with varying number of labeled examples. The F_m scores of supervised CNN on all labeled training examples are also shown as red lines.

Zheng17

UNIVERSITY OF CAMBRIDGE

Self Supervision using Transformation Recognition



Saeed19

UNIVERSITY OF CAMBRIDGE

Fig. 1. Illustration of the proposed multi-task self-supervised approach for feature learning. We train a temporal convolutional network for transformation recognition as a pretext task as shown in Step 1. The learned features are utilized by (or transferred to) the activity recognition model (Step 2) for improved detection rate with a small labeled dataset.

Results

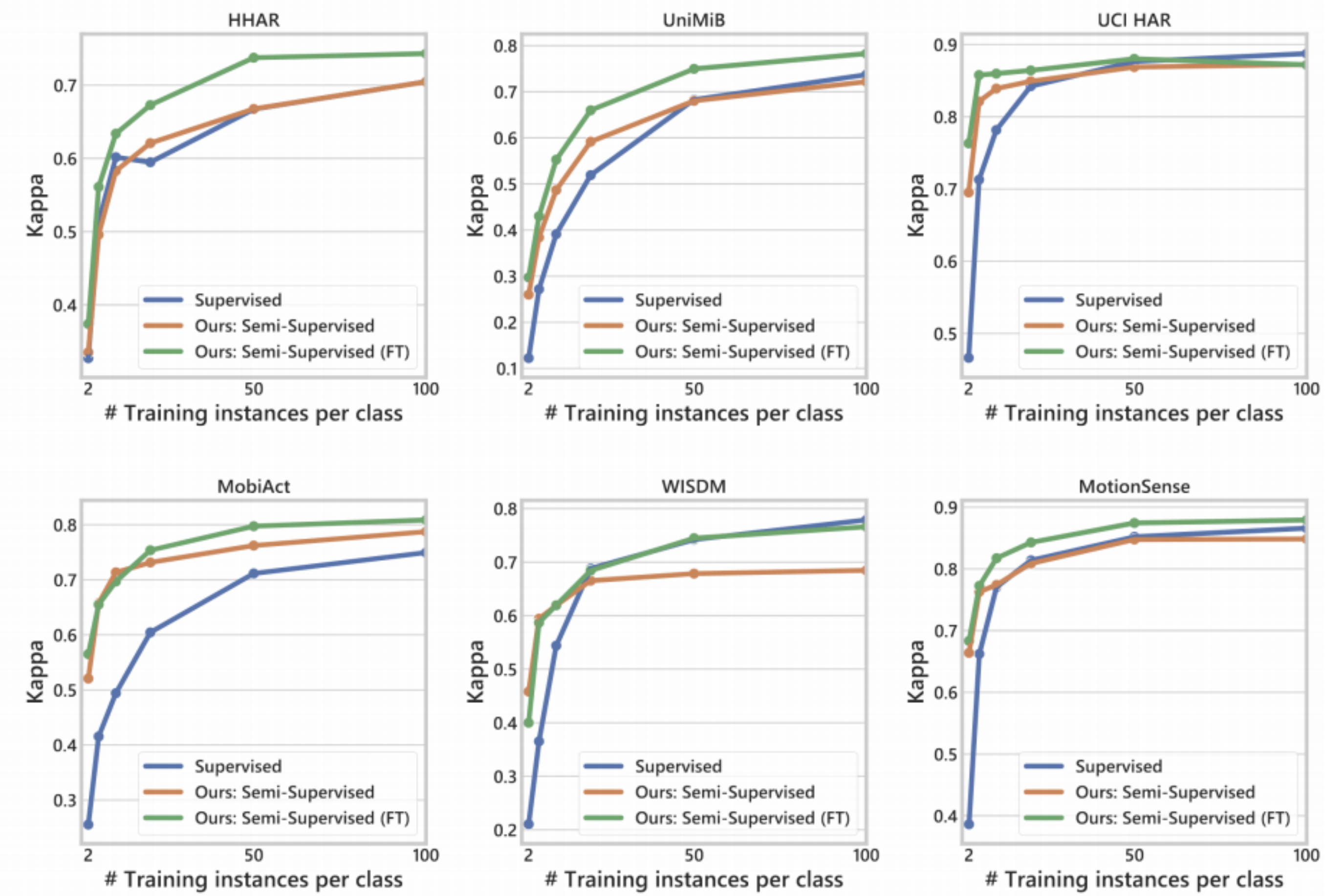


Fig. 6. Generalization of the self-supervised learned features under semi-supervised setting. The TPN is pre-trained on an entire set of unlabeled data in a self-supervised manner and the activity classifier is trained from scratch on 2, 5, 10, 20, 50, and 100 labeled instances per class. The blue curve (baseline) depicts the performance when an entire network is trained in a standard supervised way while the orange curve shows performance when we keep the transferred layers frozen. The green curve illustrates the kappa score when the last layer is fine-tuned along with the training of a classifier on the available set of labeled instances. The reported results are averaged over 10 independent runs for each of the evaluated approaches. The results with weighted f-score are provided in Figure 12 of the Appendix.

UNIVERSITY OF CAMBRIDGE

References

- N.D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, A. Campbell. A survey of mobile phone sensing. IEEE Computer Magazine. Vol. 48. No 9. September 2010.
- A. Bulling, U. Blanke, and B. Schiele, "A Tutorial on Human Activity Recognition Using Body-worn Inertial Sensors," ACM Computing Surveys, vol. 46, no. 3, Jan. 2013.
- A. Khan, N. Hammerla, S. Mellor, and T. Ploetz, "Optimising sampling rates for accelerometer-based human activity recognition," Pattern Recognition Letters, 2016.
- Nils Y. Hammerla, Shane Halloran, and Thomas Plötz. 2016. Deep, convolutional, and recurrent models for human activity recognition using wearables. In Procs of Int. Joint Conference on Artificial Intelligence (IJCAI'16.) AAAI Press.
- Deep learning for human activity recognition in mobile computing. T Plötz, Y Guan - Computer, 2018
- Haresamudram, H., Anderson, D. V., & Plötz, T. (2019, September). On the role of features in human activity recognition. In Proceedings of the 23rd International Symposium on Wearable Computers (pp. 78-88).
- Zeng, M., Yu, T., Wang, X., Nguyen, L. T., Mengshoel, O. J., & Lane, I. (2017, December). Semi-supervised convolutional neural networks for human activity recognition. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 522-529). IEEE.
- Saeed, A., Ozcelebi, T., & Lukkien, J. (2019). Multi-task self-supervised learning for human activity detection. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 3(2), 1-30.

UNIVERSITY OF CAMBRIDGE

[ack to Prof T. Plötz and Ian Tang for some material]