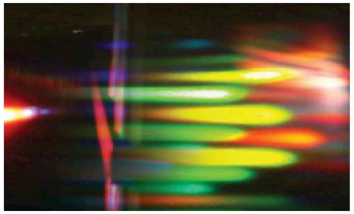
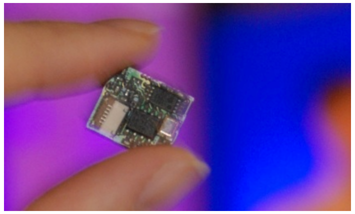
Imperial College London





Deep Learning for Human Activity Recognition: A Resource Efficient Implementation on Low-Power Devices

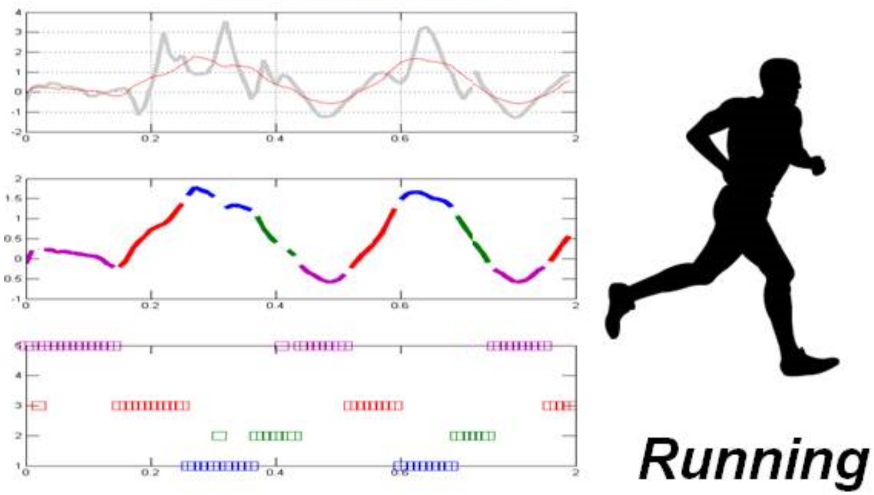
Daniele Ravì, Charence Wong, Benny Lo and Guang-Zhong Yang

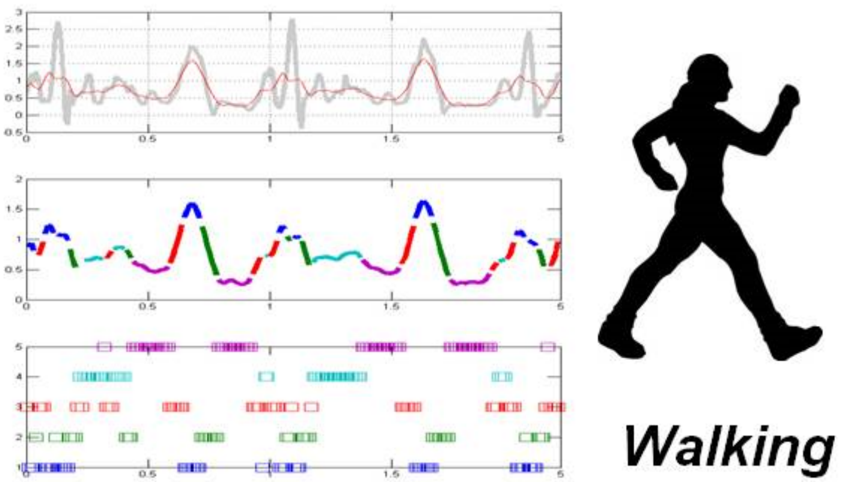
To appear in the proceedings of the IEEE EMBS 13th Annual International Body Sensor Networks (BSN) Conference, June 2016

The Hamlyn Centre
The Institute of Global Health Innovation

Activity Recognition

- Main Goal
  - Identify **human activities** (such as walking, jogging, cycling, etc.)





Applications

Wellbeing

Healthcare

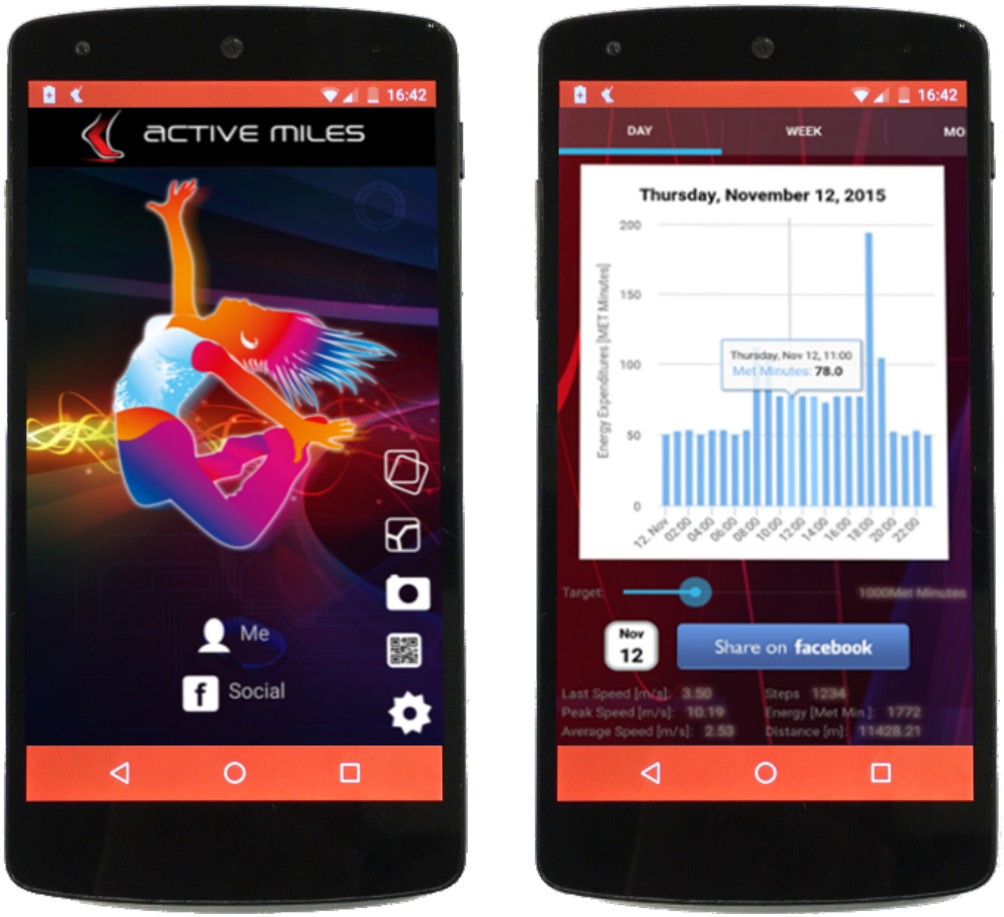
Sports

How is it recorded?

- Inertial sensors
  - Accelerometer
  - Gyroscope
  - Magnetometer
- Camera
- GPS
- Audio
- Proximity
- Barometer



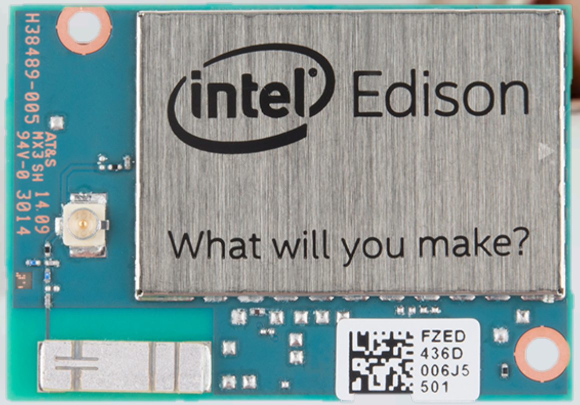
Can we use a phone?



What about wearables?

Intel Edison

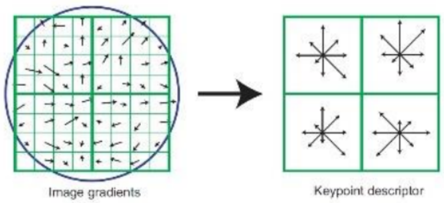
- Intel Atom dual-core CPU @ 500Hz
- 1GB DDR3 RAM
- 4GB eMMC Flash
- Dimensions: 35mm × 25mm



Deep Learning

- For many years, activity recognition approaches have been designed using *shallow* features
  - Those methods are *task dependent* and *limited*

Shallow features



SIFT [Loewe IJCV 04]

- With **Deep Learning**, features are extracted from the training data instead of being handcrafted for a specific application

TABLE I
FEATURE TABLE

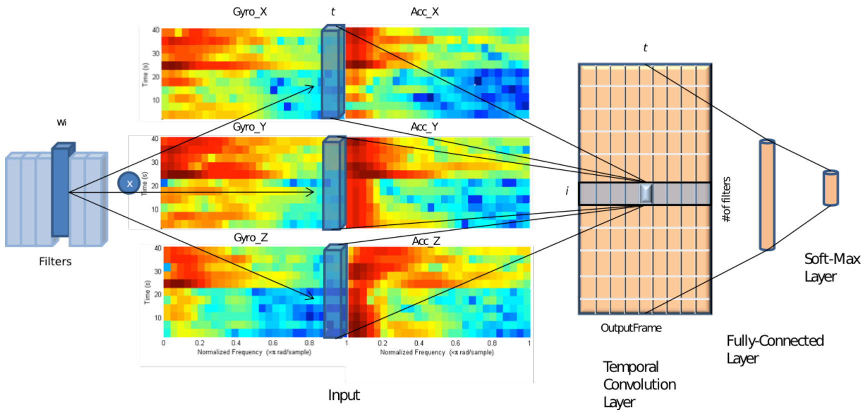
Mean	Standard Deviation	Mean Derivatives
Median	Pairwise Correlation	Interquartile Range
Skewness	Root Mean Square	Zero Crossing Rate
Variance	Mean Crossing Rate	Kurtosis
Peak-to-Peak	Max/Min Value	Amplitude

Our approach

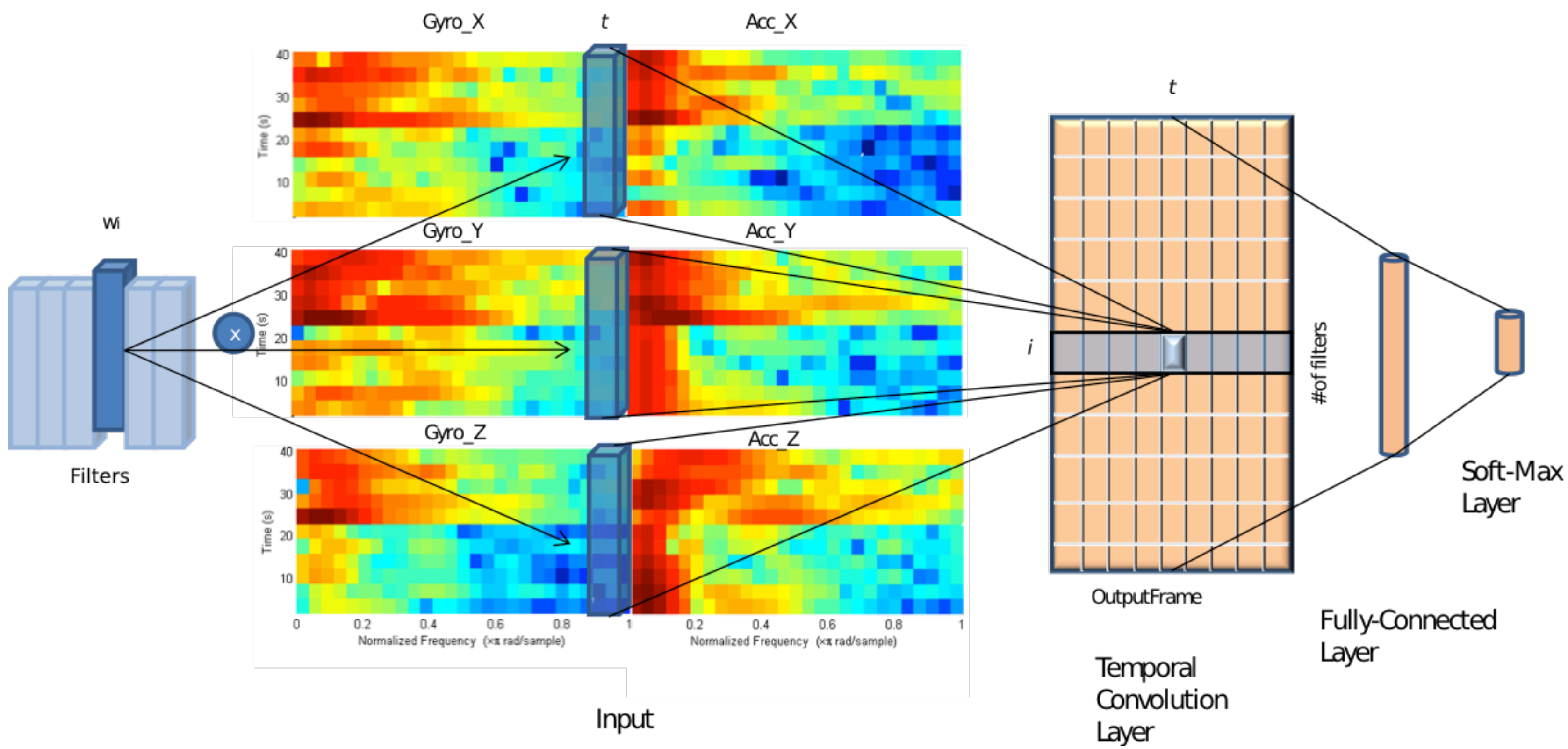
- We propose an approach that combines a *descriptive input domain* with a *deep learning* method

- The method must be *efficient*

- It must also be *robust* against transformations and variations in sensor properties



Our approach

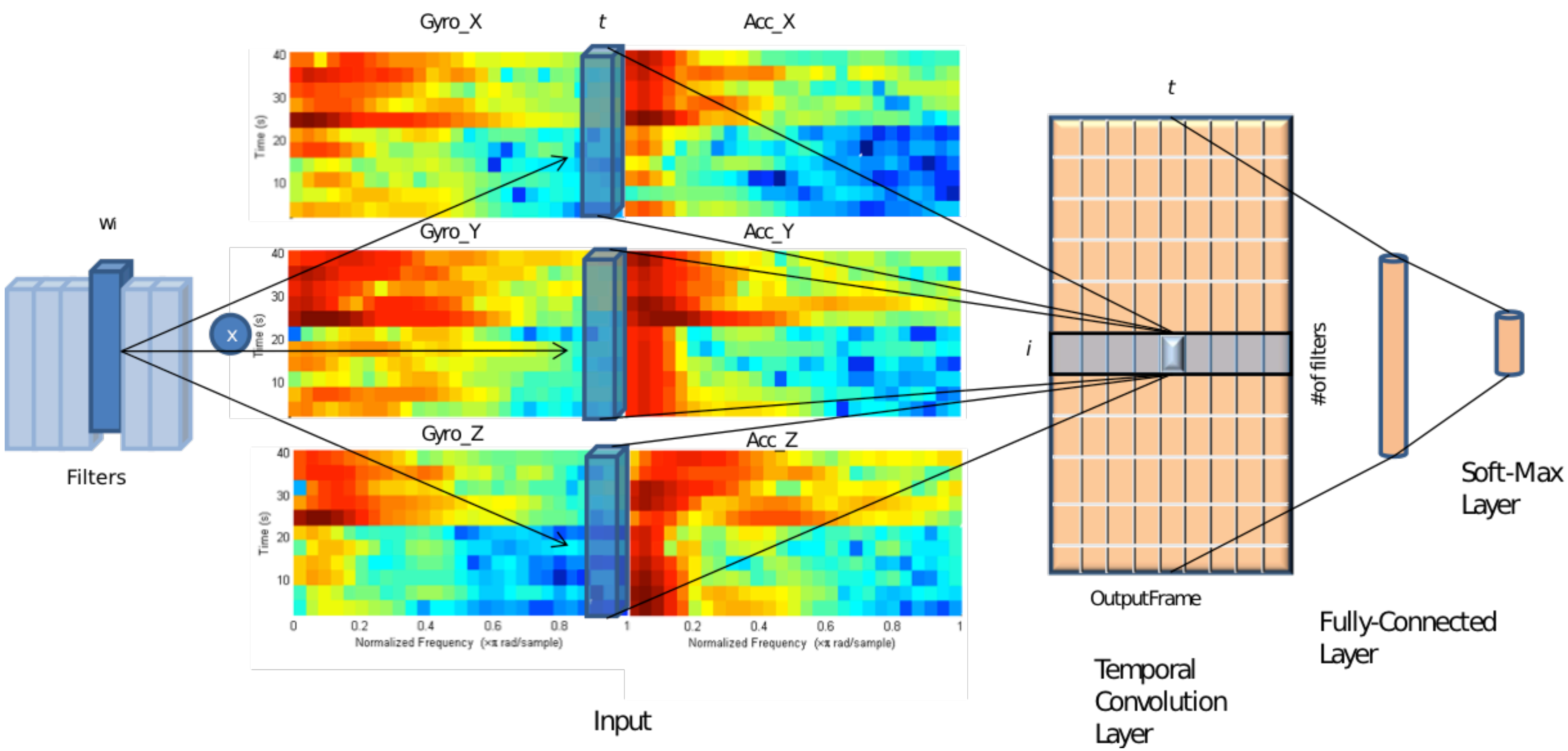


Our approach

Spectrogram

- A spectrogram of an inertial signal is a new representation of the signal as a function of *frequency* and *time*
- The procedure for computing the spectrogram is to divide a longer time signal into shorter segments of equal length and then compute the Fourier transform separately on each shorter segment
- This results in a representation that describes the changing spectra as a function of time
- The representation captures the inertial input signal whilst providing a form of temporal and sample rate invariance

Our approach



Our approach

Temporal convolution layer

- Each filter w_i is applied to the spectrogram vertically and the weighted sum of the convolved signal is computed as follows:

$$o[t][i] = \sum_{j=1}^{st} \sum_{k=1}^{kw} w[i][j][k] * input[dw * (t-1) + k][j]$$

- These temporal convolutions produce an output layer of learned features with a small size for real-time processing
- This provides orientation invariance to the input signal
- The last two layers are used to finally classify the features

Training

- We use the error value in the backward propagation routine to update each weight of the network through the Stochastic Gradient Descent (SGD) approach
- To improve the training procedure of the weights, we have used 3 regularisations:
    - Weight decay – causes the weights to exponentially decay to zero if no other update is scheduled to avoid over-fitting
    - Momentum – accelerates gradient descent to move the global minimum of the function
    - Dropout – removes units randomly from the neural network to lower generalisation error

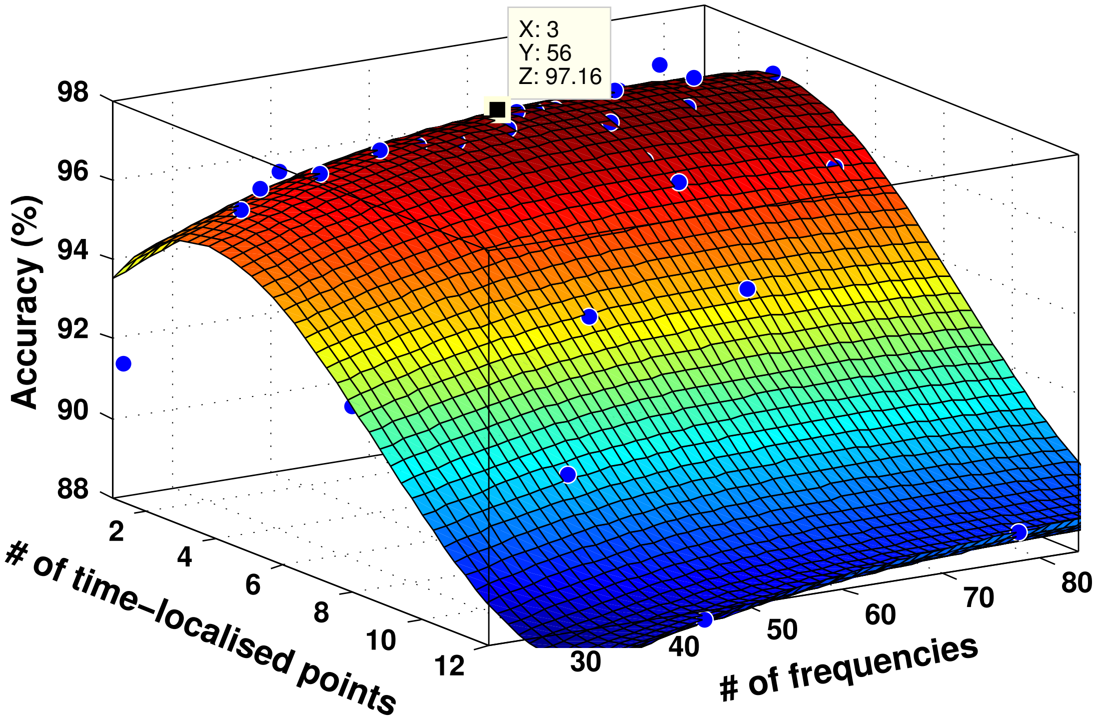
Datasets

- Four datasets are considered in our analysis
- We also release a new dataset called ActiveMiles
  - It contains unconstrained real world data from 10 subjects
  - It is one of the largest dataset (around 30 hours of labelled raw data)
  - It is the first database that contains data captured using different devices

Dataset	Number of Activities	Subjects	Sensor Placement	Sensors Accel.	Sensors Gyro.	Sampling Rate	Samples
ActiveMiles	7	10	Any placement (unconstrained)	✓	✓	50 – 200Hz	4,390,726
WISDM v1.1	6	29	Front trouser pocket (thigh)	✓	✗	20Hz	1,098,207
Daphnet FoG	2	10	Trunk, thigh, ankle	✓	✗	64Hz	1,917,887
Skoda	10	1	Arms (20 positions)	✓	✗	98Hz	$\sim 701,440$

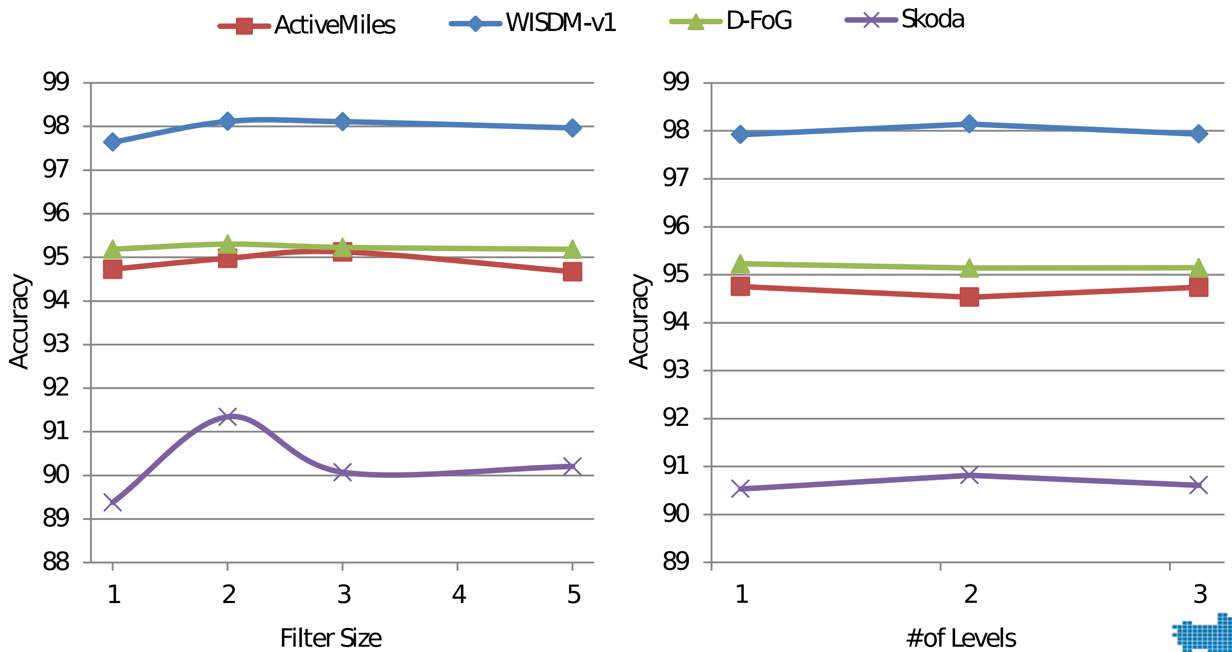
Experimental setup

- Classification accuracy changes when the spectrogram generation parameters are modified



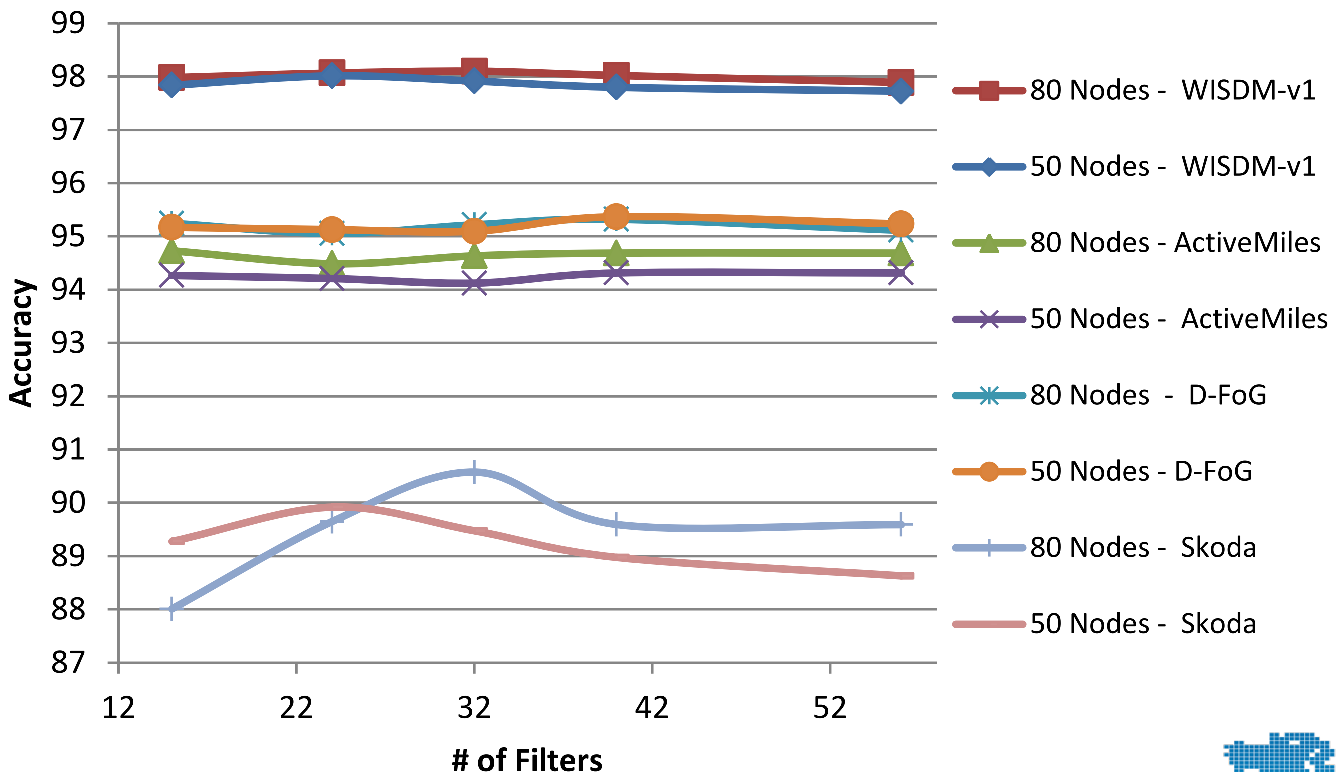
Experimental setup

- The optimal size of the temporal convolution kernel is two or three, depending on the data being classified
- The proposed approach requires few levels in order to obtain good results



Experimental setup

- 15 filters and just 80 nodes in the fully convolution layer are sufficient for a good classification



Results

- A comparison of HAR results using four baselines, existing methods, and the considered datasets are shown in Table III
- The accuracy of the proposed method is typically better in comparison to the other methods

Dataset	Approach	Window	Accuracy (%)	
ActiveMiles	Cs1	10s	81.0	
	Cs2		89.3	
	Cs3		94.7	
	Cs4		95.0	
	Ours		**95.1**	
WISDM v1.1	Cs1	10s	85.1	
	Cs2		91.3	
	Cs3		96.7	
	Cs4		97.4	
	[13]		**98.2**	
	[7]		94.3	
	Ours		**98.2**	
Skoda (Node 16)	Cs1	4s	86.2	
	Cs2		89.2	
	Cs3		94.0	
	Cs4		**95.9**	
	[21]		86.0	
	[13]		89.4	
	Ours		91.7	
Daphnet FoG			Sensitivity	Specificity
	Cs3	4s	62.2	96.9
	Cs4		66.3	**97.7**
	[20]		73.1	81.6
	[13]		**91.5**	91.5
	Ours		71.9	96.7

Performance

- A comparison of the computation times required to classify activities on different devices
  - All resulting times are within the requirements for real-time processing

Device	Spectrogram	Deep Learning
LG Nexus 5	5.4ms	5.7ms
Samsung Galaxy S5	20ms	8ms
Intel Edison	13ms – 42ms	14.9ms

Conclusion

- The proposed system generates discriminative features that are generally more powerful than handcrafted features
- The accuracy of the proposed approach is better or comparable against existing state-of-the-art methods
- The ability of the proposed method to generalise across different classification tasks is demonstrated using a variety of human activity datasets
- The computation times obtained from low-power devices are consistent with real-time processing

Thank you for your attention!