



On-device subject recognition in UWB-radar data with Tiny Machine Learning

Massimo Pavan

- **Phd Student**

Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB)

Politecnico di Milano, Italy

Email: massimo.pavan@polimi.it

Web: <https://www.deib.polimi.it/ita/personale/dettagli/1115470>

- **Research interests:** machine/deep learning, TinyML, IoT and edge computing, robotics.



Prof. Manuel Roveri, PhD

- **Associate Professor**

Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB)

Politecnico di Milano, Italy

Email: manuel.roveri@polimi.it

Web: <http://roveri.faculty.polimi.it>

- **Research interests:** machine/deep learning, TinyML, IoT and edge computing, learning in nonstationary environments, privacy preserving computation.



Armando Caltabiano

- **CEO @ Truesense s.r.l.**

Email: armando.caltabiano@truesense.it

What is tiny Machine Learning, anyway?

Tiny machine learning is the field of **machine learning** technologies including **hardware**, **algorithms** and **software** capable of performing on-device analytics at extremely **low power** ($\approx$ mW), enabling a variety of **always-on** use-cases and targeting battery operated devices.

-“tinymml.org”

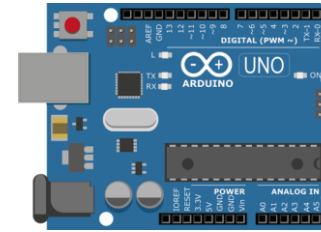


Move (intelligent) processing as close as possible to data generation units ...

- Increase autonomy
- Reduce decision-making latency
- Reduce transmission bandwidth
- Increase energy-efficiency
- Security and Privacy

- Low computing ability Constraints on energy
- Constraints on memory (RAM/FLASH)
- Complexity in design and development
- Strong connection between HW, SW and ML

Arduino



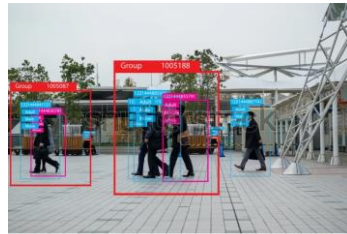
STM32



Domain	STM32 L1 Series	STM32F4 Series
	Ultra Low-Power	High-Performance
Flash Memory (kB)	32 to 512	64 to 2048
RAM Memory (kB)	4 to 80	32 to 320
CPU	ARM® Cortex®-M3	ARM® Cortex®-M4
Frequency (MHz)	32	84 to 180
Supply Voltage (V)	1.65 to 3.6	1.71 to 3.6
Supply Current (μA)	0.28 (0.28) to 230	1.1 (140) to 282



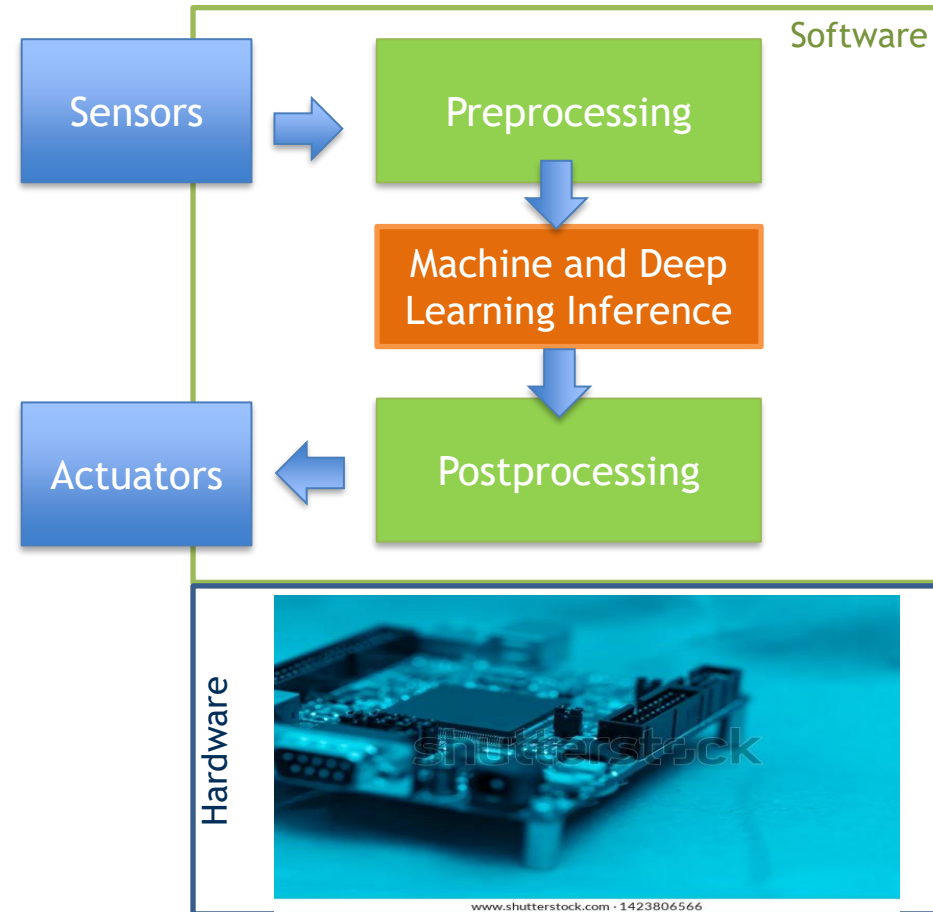
Wake-word detection



Person detection

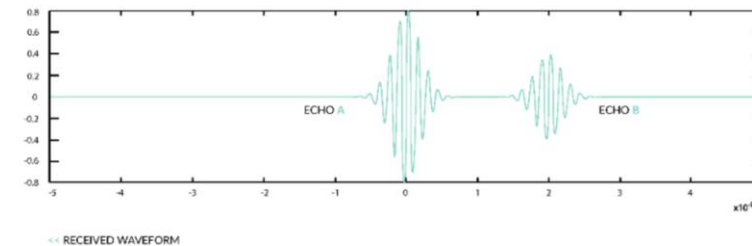
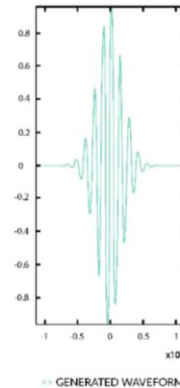
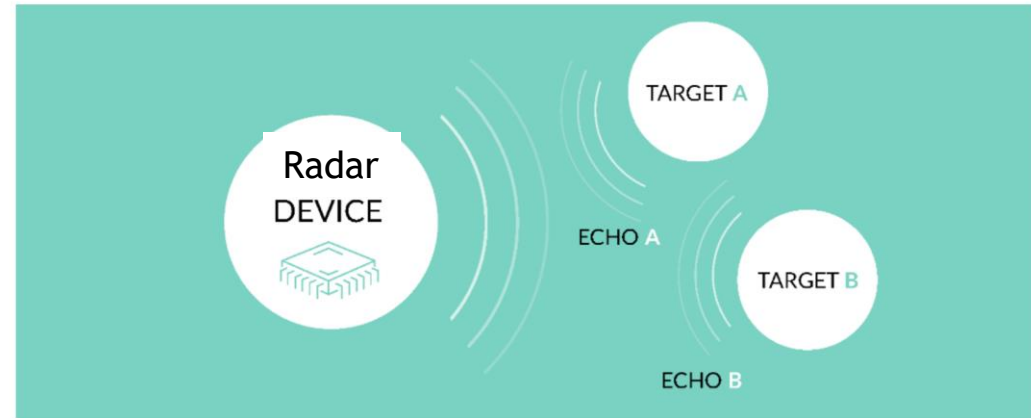


Gesture recognition

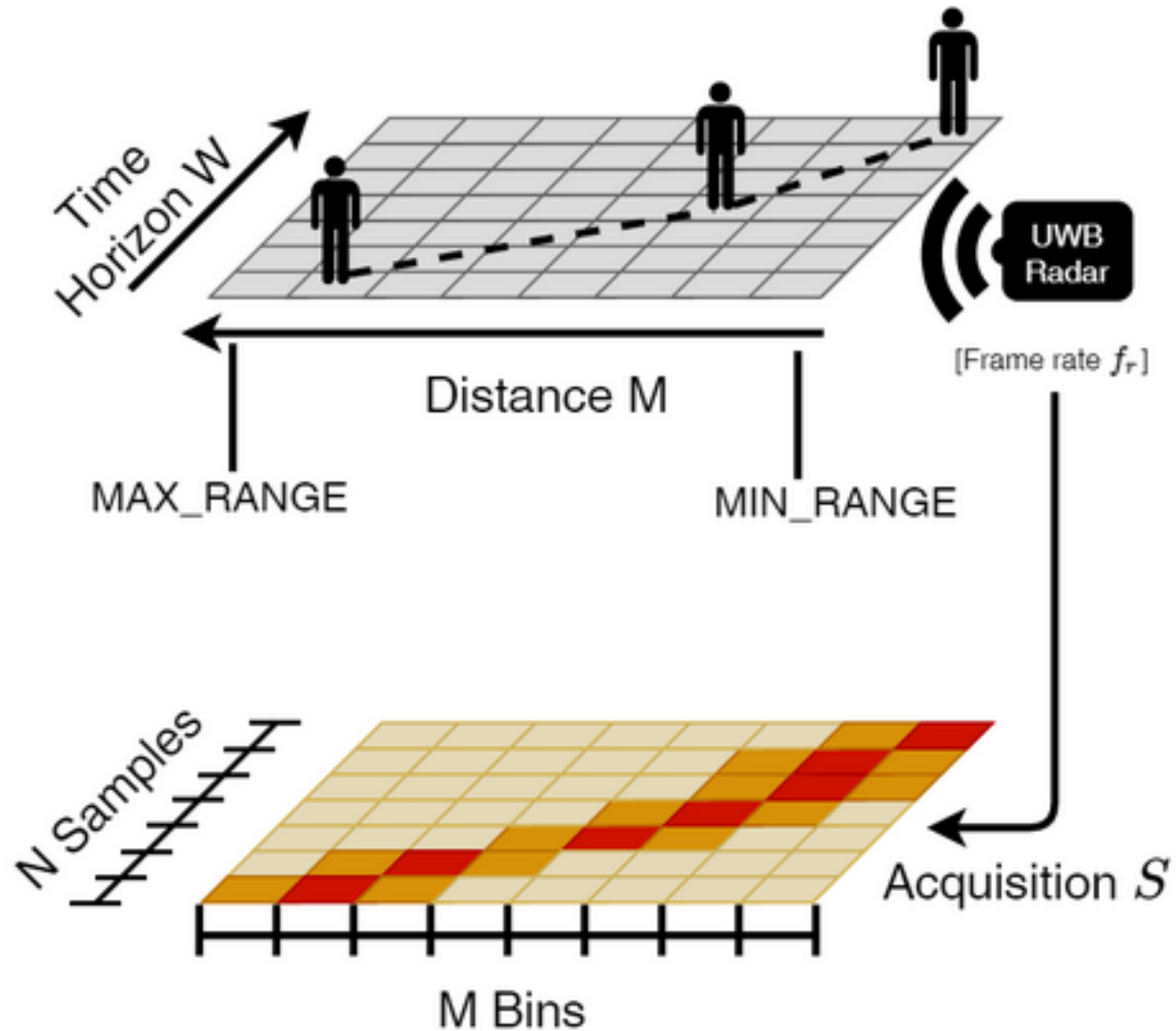


UltraWideBand (UWB) Radar

- **UWB** is a radio technology used for short-range, high-bandwidth communications.
- UWB radar advantages:
 1. **Privacy** preservation
 2. high **precision** data
 3. Fastness and **low energy** consumption.



From the world to the data



Related Literature

- TinyML literature
 - Hardware Design
 - Approximated ML and DL solutions:
 - Quantization
 - Pruning
 - Gate Classification
 - Presence Detection
 - Video PD (visual wake word detection)
 - Audio PD (keyword spotting)
- UWB-Radar literature
 - Presence Detection
 - Human activity Recognition

No example present in the literature about the use of UWB-Radar data in TinyML environment

Breathing rates per age

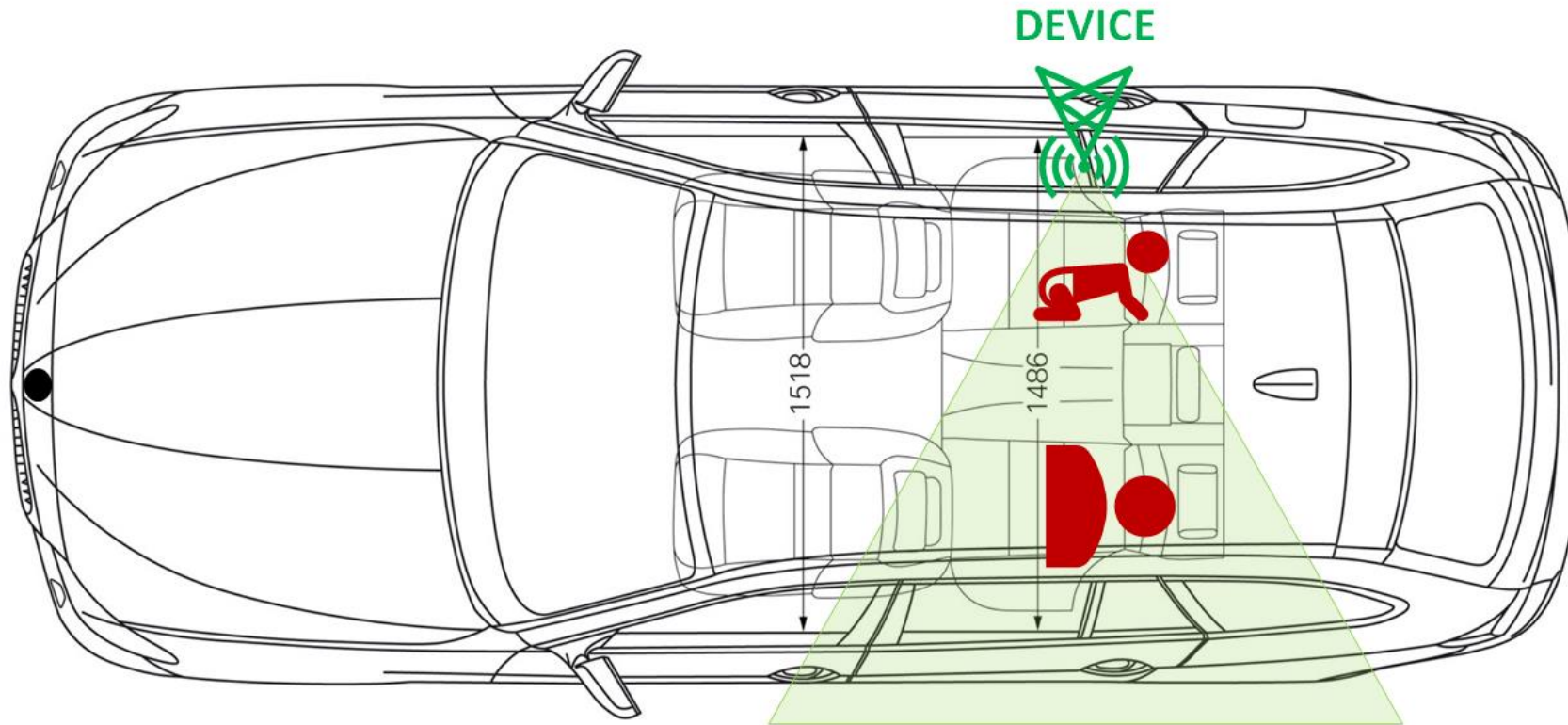
Class	Avg. respiratory rate at rest
birth to 6 weeks	30-40 breaths per minute
6 months	25-40 breaths per minute
3 years	20-30 breaths per minute
Adults	15-18 breaths per minute
Elderly ≥ 65 years old	12-28 breaths per minute

Table 1. Average respiratory rates

Class	Avg. respiratory rate at rest	Hz	Expected bins
Children	20-40 breaths per minute	1/3 - 2/3 Hz	8,33 - 16, 66
Adults	15-18 breaths per minute	1/4 - 0.3 Hz	6,25 - 8

Table 2. Average respiratory rates and expected bins

The application: in-car person recognition



Goal:

- **Detect the presence of children or pets on the back seats.**

Used device

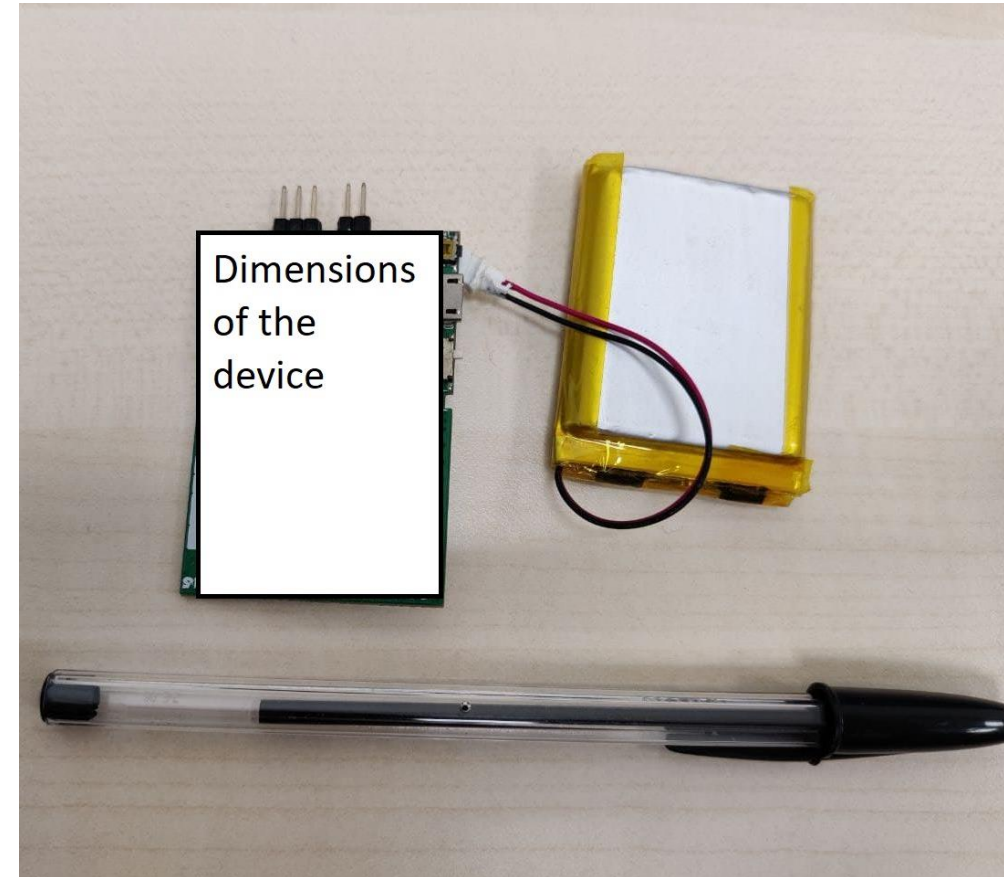
- Based on an **ESP32** Microcontroller unit (MCU).

Equipped with:

- Ultra Wide Band radar module, with **only one pair of antennas**.

And with available for the algorithm:

- Flash memory of **3MB**.
- RAM memory of **100KB**.
- Battery of **1800 mAh**, or 6/7 months of normal use.

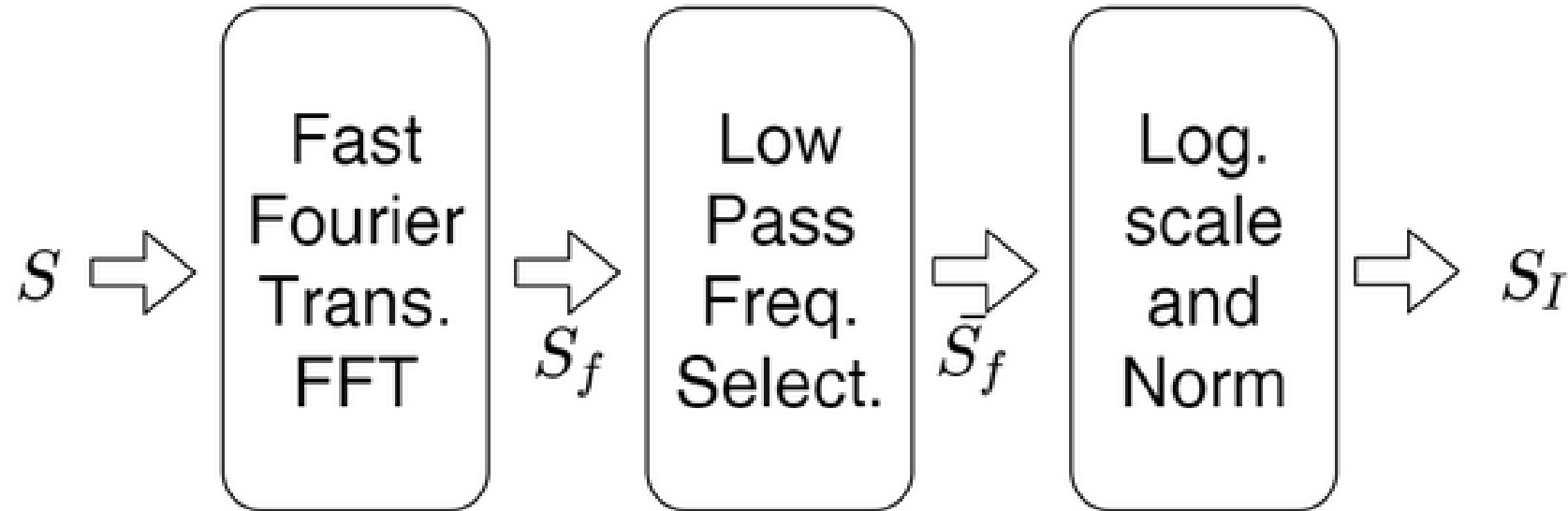


The solution

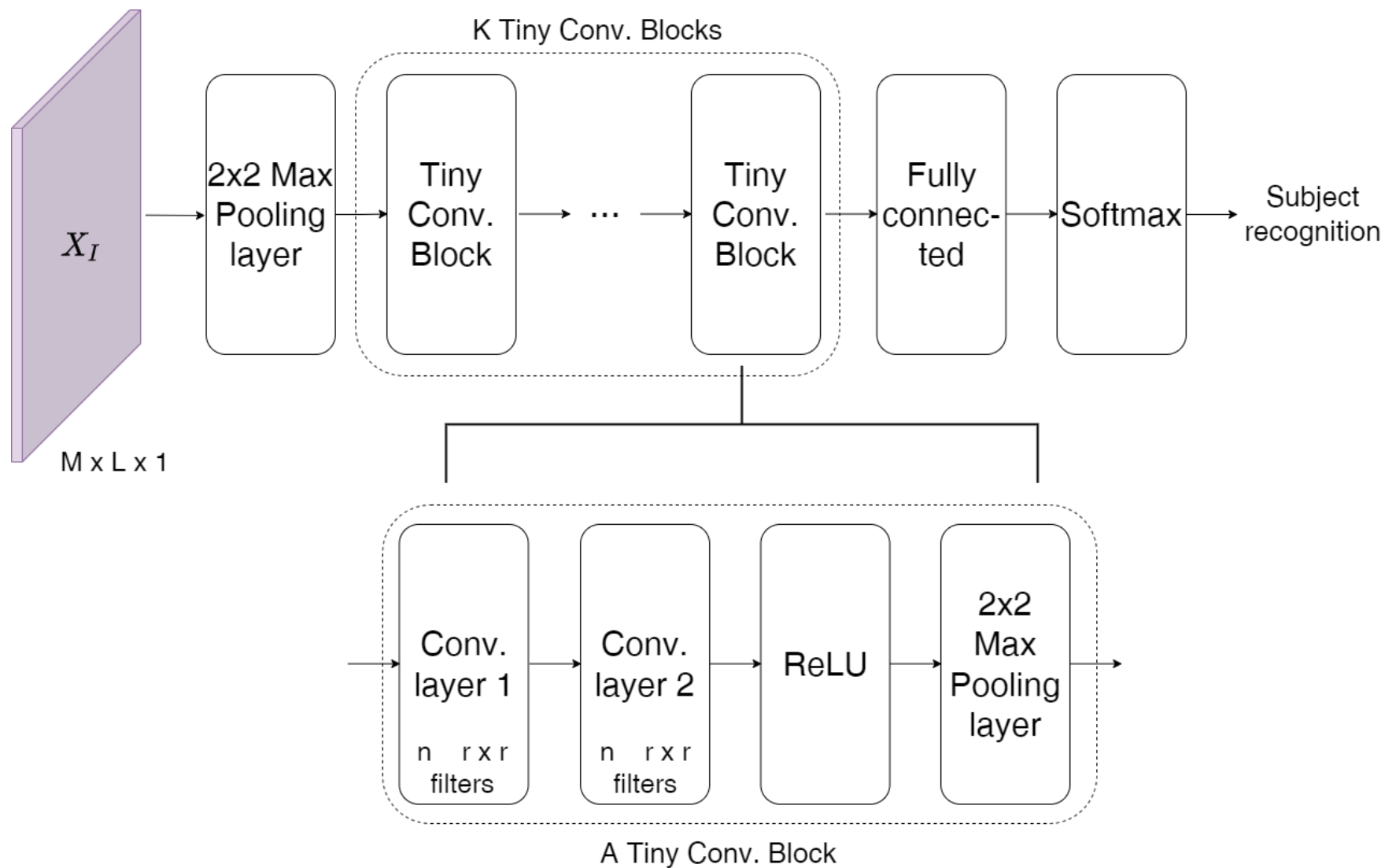
Proposed solution:

- A preprocessing module
- TyCNN-C, a design for tiny CNN designed considering hardware limitations.
 - Max Pooling for dimensionality reduction
 - K Tiny Conv Blocks
 - Fully connected classification head
 - Post-training Quantization of weights, activation and inputs

The preprocessing module



TyCNN-C architecture



Results

Table 4. Comparison of the results of the TyCNN-C and the baseline algorithm.

Network	Accuracy	m (kB)	c (10^6)
naive baseline	0.513	/	/
TyCNN	0.783 ± 0.076	47.79	8.57

true child	none	82.72%	15.45%	1.83%
	child	19.01%	69.32%	11.67%
	adult	0.0%	15.22%	84.78%
		empty	child predicted	adult

Memory occupations

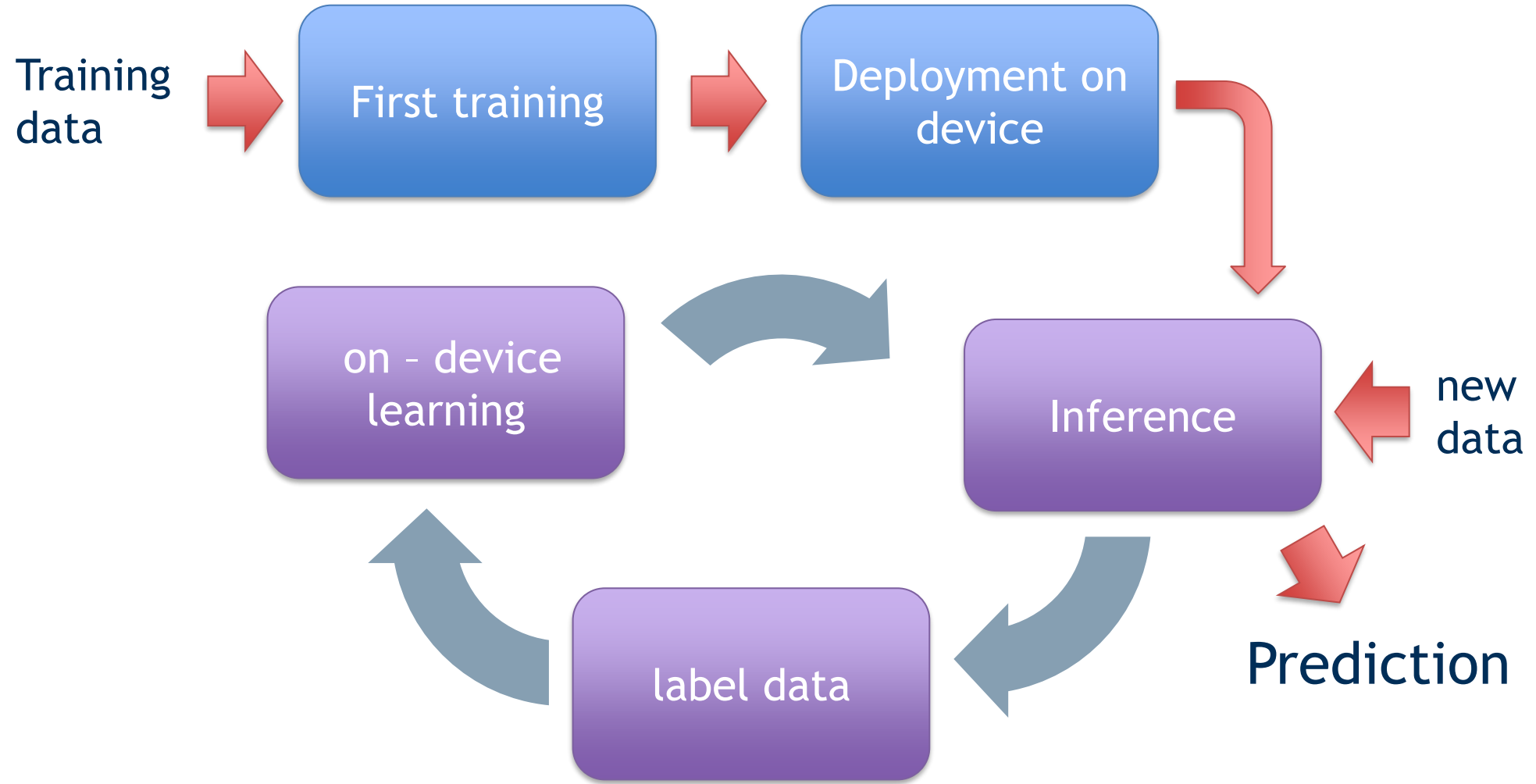
Memory occupations (B)

Input	4 558
Weights	17 629
Peak Activations	31 304
Total	48 933

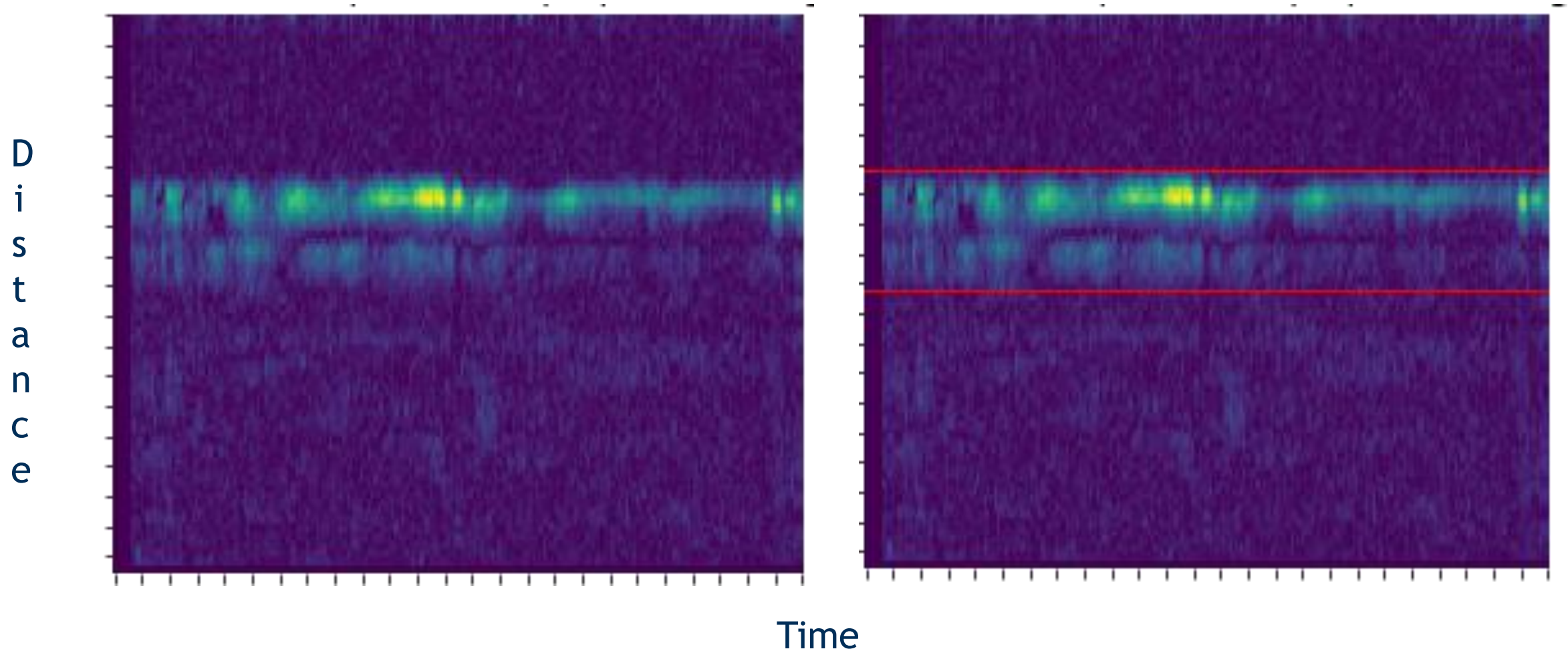
Number of operations

Total	8 571 216
-------	-----------

The next challenge: incremental on-device learning



The next next challenge: «Radar Object Detection» for people tracking



Thanks for your
attention!
Any question?