EarBuds Gesture Recognition with TinyML Using Novel Short-Range Radar Sensors

Michele Magno, Andrea Ronco

D-ITET Center for Project-based Learning and Embedded Systems, ETH Zurich

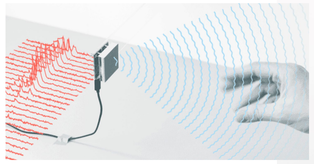
Gesture Recognition Based on Short-Range Radar

This work proposes a low-power high-accuracy embedded hand-gesture recognition targeting battery-operated wearable devices using low power short-range radar sensors. A 2D Convolutional Neural Network (CNN) using range-frequency Doppler features is combined with a Temporal Convolutional Neural Network (TCN) for time sequence prediction. The final algorithm has a model size of only 19855 parameters, yielding a memory footprint of only 39.5kB. A dataset with 11 challenging gestures plus an idle class has been recorded, containing 50 samples per gesture acquired by 19 people, and a single-user dataset of 700 samples per gesture from a single person. On these datasets an accuracy of 92% (20 users) has been achieved. The network has been integrated in an embedded system based on ARM Cortex-M4 with 270mm², suitable for wearable devices. The system can perform real-time gesture recognition with an active power consumption of ~17mA and of ~400uA in idle, which opens the door to prolonged activity in battery operated, small wearable devices.

Increasing research on radar for gesture recognition[1,2,3,4]

Google developed micro-radar for gesture recognition

Good results on difficult hand-gestures: **92%** accuracy on 11 gestures and 20 people



[1]Soli: Ubiquitous Gesture Sensing with Millimeter Wave Radar, 2016

[2]Interacting with Soli: Exploring Fine-Grained Dynamic Gesture Recognition in the Radio-Frequency Spectrum. 2016

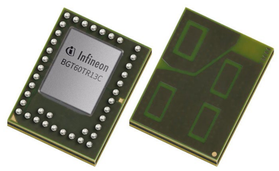
[3]Sparsity-based Dynamic Hand Gesture Recognition Using Micro-Doppler Signatures, 2017

[4]A Hand Gesture Recognition Sensor Using Reflected Impulses. 2017

Contribution of This Work

Implementing Radar-Based Hand Gesture Recognition in Embedded Systems

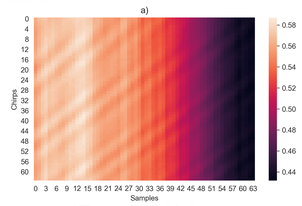
- Created a dataset with challenging fine-grained gestures (1250 samples per class, 20 users)
- Real-Time algorithm that runs on well-know hardware (ARM Cortex-M4)
- Accuracy similar to what seen in Interacting With Soli (92% with 20 users)
- Network at least 700x smaller than I.w.Soli (40kB), capable of real-time prediction



Technical Background

Short Range Radar

- Radar technology: FMCW
  RX signal mixed with TX to obtain *Intermediate Frequency* signal. Distance of the target encoded in the frequency of IF signal.
- Aviation Radar: 1-12GHz (λ=2.5-30cm)
- Infineon's Radar: 60GHz (λ=5mm)
- Sensor data: samples of IF signal over one frame
  - Frame: 32 chirps spaced 700µs, 16 samples per chirp
  - 13 FPS (frames per second)

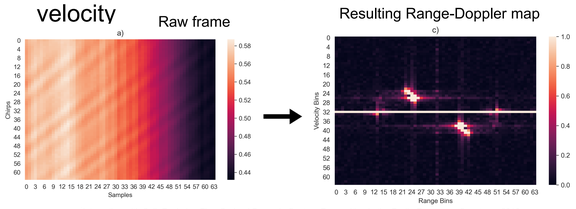


*Example with larger scale for demonstration

Range-Doppler Maps

- Strong feature based on research[1,2]
- Efficient computation based on CMSIS-DSP

1. FFT on the sample axis (Fast Time) → Resolve Range
2. FFT on the chirps axis (Slow Time) → Resolve velocity

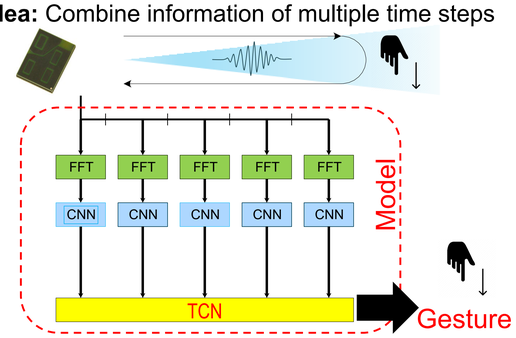


[1]Interacting with Soli: Exploring Fine-Grained Dynamic Gesture Recognition in the Radio-Frequency Spectrum. 2016

[2]Short-Range FMCW Monopulse Radar for Hand-Gesture Sensing, 2015

Embedded And Energy Efficient Algorithm CNN+TCN [1]

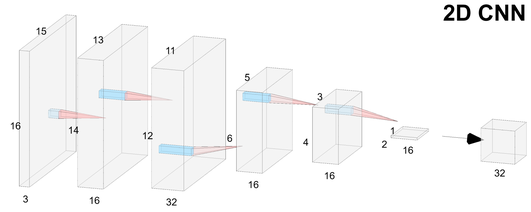
Idea: Combine information of multiple time steps



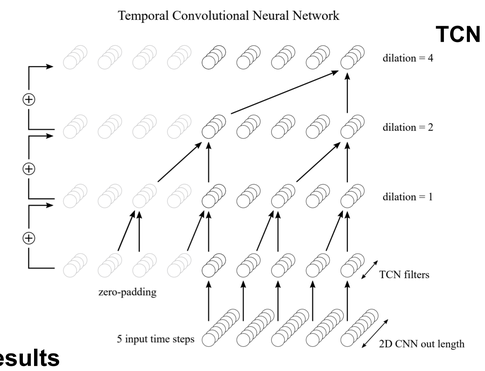
1. **2D Convolutional Neural Network (CNN)** scaling down the range frequency Doppler input map of size 15x16x3 to a representation vector of length 32.
2. **Temporal Convolutional Neural Network (TCN)** using as input a time sequence of stacked representation vectors of length 32, leverage temporal information for more accurate predictions.
3. **Dense Classifier:** The output representations of length 24 produced by the TCN are then fed into two fully connected layers, which create the probability distribution to classify the observed gesture.

[1] Scherer, M., Magno, M., Erb, J., Mayer, P., Eggimann, M., & Benini, L. (2021). Tinyradarnn: Combining spatial and temporal convolutional neural networks for embedded gesture recognition with short range radars. *IEEE Internet of Things Journal*, *8*(13), 10336-10346.

2D CNN

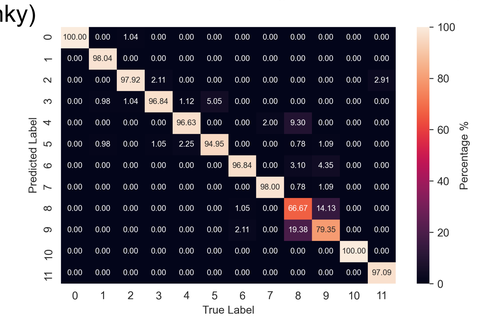


TCN



Results

Achieved an accuracy 92%, only two very similar gestures are confused often (pinch index and pinch pinky)



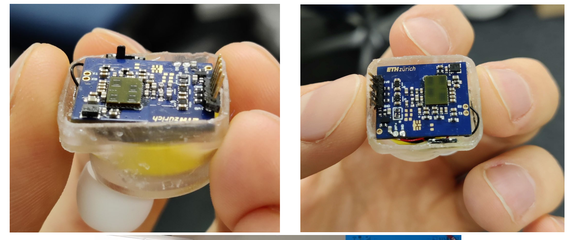
In Field Ev. and Comparison with Google Soli [1]

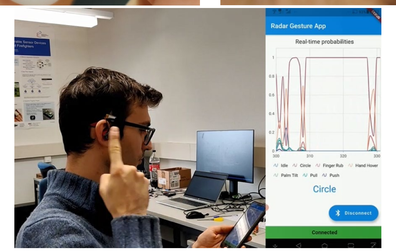
Other Properties	Soli	This Work
Model Size	689MB	**39.5kB**
Dataset: Total Instances per Gesture	500	**1650**
Dataset: People	10	**20**
Embedded Implementation	No	**Yes**
Power Consumption	-	**~25mA (run)**

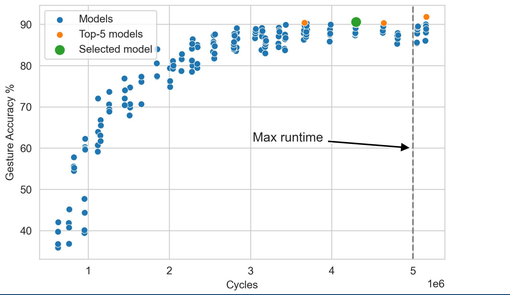
Step	Time[ms] @ 80MHz
Data RX + Preprocessing	5.13
2D CNN	26.28
TCN	27.07
Dense/Fully Connected	0.38
Total	**58.86**

The inference takes 91% of the total time per frame elaboration

At 80MHz the system can achieve real-time inference up to 16FPS.

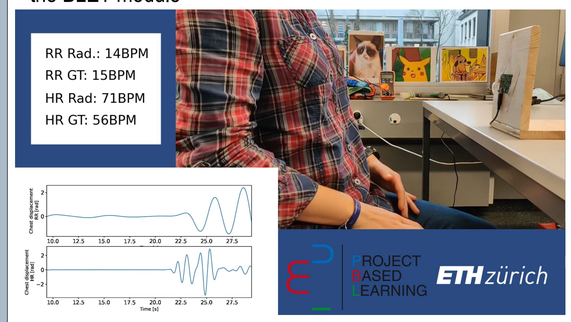






Vital Sign Recognition Preliminary results

We investigated on the vital signs monitoring exploiting the same era-buds. Signal processing algorithms for vital signs implemented on the ARM-Cortex-M4 microcontrollers insed the BLET module



Conclusion

This work presented a high-accuracy and low-power hand-gesture recognition system based on short-range radar. Two large datasets with 11 challenging hand-gestures performed by 20 different people containing a total of 20210 gesture instances are recorded, on which the final algorithm reaches an up to 92%. The model size is below 40kB and the hardware implementation shows that the application is feasible in a Cortex-M4 based, battery operated wearable device. The Earbud can be used also as vital sign monitoring