On Device Learning Forum

Enabling Ultra-low Power Machine Learning at the Edge

“NeuroMem®, Ultra Low Power hardwired incremental learning and parallel pattern recognition”

Guy Paillet – Co-founder and Chairman, General Vision Holdings

May 16, 2023

www.tinyML.org
The goal of On Device Learning (ODL) is to make edge devices “smarter” and more efficient by observing changes in the data collected and self-adjusting/reconfiguring the device’s operating model. Optionally the “knowledge” gained by the device is shared with other deployed devices.

Danilo Pau, Elias Fallon, Evgeni Gousev, Davis Sawyer, Ira Feldman, Christopher B. Rogers
tinyML On Device Learning Forum
8/31 – 9/1, 2022 Online

Accademia on 8/31/2022
- On-Device Learning Under 256KB Memory, Song HAN, Assistant Professor, MIT EECS
- Neural Network ODL for Wireless Sensor Nodes, Hiroki MATSUTANI, Professor, Keio University
- Scalable, Heterogeneity-Aware and Trustworthy Federated Learning, Yiran CHEN, Professor, Duke University
- On-Device Learning For Natural Language Processing with BERT, Warren J. GROSS, Professor, McGill University
- Is on-device learning the next “big thing” in TinyML? Manuel ROVERI, Associate Professor, Politecnico di Milano
- ODL Professors Panel

Industry on 9/1/2022
- TinyML ODL in industrial IoT, Haoyu REN, PhD Student, Technical University of Munich/Siemens
- NeuroMem® wearable, hardwired sub milliwatt real time machine learning with wholly parallel access to “neuron memories” fully explainable, Guy PAILLET, Co-founder, General Vision
- Using Coral Dev Board Micro for ODL innovations, Bill LUAN, Senior Program Manager, Google
- Platform for Next Generation Analog AI Hardware Acceleration, Kaoutar EL MAGHRAOUI, Principal Research Scientist, IBM T.J Watson Research Center
- Enabling on-device learning at scale, Joseph SORIAGA, Sr. Director of Technology, Qualcomm
- Training models on tiny edge devices, Valeria TOMASELLI, Senior Engineer, STMicroelectronics

https://www.tinyml.org/event/on-device-learning/
• **A framework of algorithms and associated tool for on-device tiny learning**, Danilo PAU, Technical Director, IEEE and ST Fellow, STMicroelectronics

• **In Sensor and On-device Tiny Learning for Next Generation of Smart Sensors** Michele MAGNO, Head of the Project-based learning Center, ETH Zurich, D-ITET

• **Continual On-device Learning on Multi- Core RISC-V MicroControllers** Manuele RUSCI, Embedded Machine Learning Engineer, Greenwaves

• **On-device continuous event-driven deep learning to avoid model drift**, Bijan MOHAMMADI, CSO, Bondzai
2023 Gartner Emerging Technologies and Trends Impact Radar

gartner.com

Note: Range measures number of years it will take the technology/trend to cross over from early adopter to early majority adoption. Mass indicates how substantial the impact of the technology or trend will be on existing products and markets.

Source: Gartner
© 2023 Gartner, Inc. All rights reserved.
On Device Learning Forum 2023, May 16 2023

- 8:00 - 8:10 Opening remarks by Danilo Pau
- 8:10 - 8:40 Charlotte Frenkel "Merging insights from artificial and biological neural networks for neuromorphic edge intelligence"
- 8:40 - 9:40 Giorgia Dellafererra "Forward Learning with Top-Down Feedback: Solving the Credit Assignment Problem without a Backward Pass"
- 9:40 - 10:10 Guy Paillet "NeuroMem®, Ultra Low Power hardwired incremental learning and parallel pattern recognition"
- 10:10 - 10:40 Aida Todri-Sanial "On-Chip Learning and Implementation Challenges with Oscillatory Neural Networks"
- 10:40 - 11:10 Eduardo S. Pereira “Online Learning TinyML for Anomaly Detection Based on Extreme Values Theory”
- 11:10 - 11:15 Closing remarks by Danilo Pau
Thank you, tinyML Strategic Partners, for committing to take tinyML to the next Level, together
EDGE IMPULSE

The Leading Development Platform for Edge ML

dgeimpulse.com
Advancing AI research to make efficient AI ubiquitous

Power efficiency
- Model design, compression, quantization, algorithms, efficient hardware, software tool

Personalization
- Continuous learning, contextual, always-on, privacy-preserved, distributed learning

Efficient learning
- Robust learning through minimal data, unsupervised learning, on-device learning

Perception
- Object detection, speech recognition, contextual fusion

Reasoning
- Scene understanding, language understanding, behavior prediction

Action
- Reinforcement learning for decision making

A platform to scale AI across the industry
Accelerate Your Edge Compute

SYNTIANT

Making Edge AI A Reality

www.syntiant.com
Platinum Strategic Partners
Renesas is enabling the next generation of AI-powered solutions that will revolutionize every industry sector.
DEPLOY VISION AI AT THE EDGE AT SCALE
Gold Strategic Partners
Witness potential made possible at analog.com.
Easily deploy your tinyML solutions with Arduino Pro

arduino.cc/pro

Made In Italy
Arm AI Virtual Tech Talks

The latest in AI trends, technologies & best practices from Arm and our Ecosystem Partners.

Demos, code examples, workshops, panel sessions and much more!

Fortnightly Tuesday @ 4pm GMT/8am PT

Find out more: www.arm.com/techtalks
Driving decarbonization and digitalization. Together.

Infineon serving all target markets as Leader in Power Systems and IoT

www.infineon.com
NEUROMORPHIC INTELLIGENCE FOR THE SENSOR-EDGE
The Right Edge AI Tools Can Make or Break Your Next Smart IoT Product

Analytics Toolkit Suite

AutoML

Data Collection

Tool & Validation

Data Labeling

Code Generation

Model Eliciting

Version Control and Model Management

sensiml.com/tinyML
STMicroelectronics provides extensive solutions to make tiny Machine Learning easy
We engineer exceptional experiences for consumers in the home, at work, in the car, or on the go.

www.synaptics.com
Join Growing tinyML Communities:

**Meetup**

- tinyML - Enabling ultra-low Power ML at the Edge

**LinkedIn**

- The tinyML Community
  - https://www.linkedin.com/groups/13694488/

- 14.7k members in 47 Groups in 39 Countries
- 4k members & 11.6k followers
Subscribe to tinyML YouTube Channel for updates and notifications (including this video)
www.youtube.com/tinyML
tinyML EMEA Innovation Forum

June 26 -28, 2023
Amsterdam

EMEA 2023
https://www.tinyml.org/event/emea-2023

More sponsorships are available: sponsorships@tinyml.org
Reminders

Slides & Videos will be posted tomorrow

Please use the Q&A window for your questions

tinyml.org/forums  youtube.com/tinyml
Guy Paillet

Guy’s background is hardware design since 1976 starting with Motorola MC6800 as application engineer. He has been innovating on high performance Tiny Machine Learning since 1993 while inventing the ZISC36 with IBM Paris, Guy and family moved from France in 1996 and co-founder General Vision in 2000. Since, General has licensed its NeuroMem ZISC technology giving birth to 4 additional successful Neuromorphic AISC from 2007 to 2022, including the Intel Curie for “NeuroMEMS.”
BRAINCARD® the smallest wearable device with on chip learning in the World

ST SensorTile:
- Accelero, gyro, magneto, pressure
- Audio MEM
- Bluetooth Low Energy

STM32L476 microcontroller

NeuroMem NM500

Lattice FPGA XO3

Optional 2nd NeuroMem NM500

Reserved

14 GPIO lines (SPI, I2C and digital lines)

Back
- Battery-charging circuitry
- MicroSD card

BRAINCARD® is a registered trademark of General Vision
ZYNQ +NM500 industrial application steel factory condition monitoring
And more at https://general-vision.com/category/post_applications/
The missing brain

Dependency, Latency
Power consumption

MEMS sensors

NeuroMem chip
Parallel pattern recognition

MEMS actuators

Direct link, High speed, Low Power

General Vision Inc.
History of classifiers

1960: Threshold method
Too simple and not non-linear

1980: Perceptron method
Too complex, and time consuming

1990: Radial Basis Function
Map spaces of any shape with the relevant training set
The power of parallelism

CPU/DSP

1 processor for many memory entries
Recognition in N cycles per entry UNTIL the right one is found

NeuroMem

Many memories with their own processing unit
Recognition in 16 cycles ∀ number of entries
Biologically inspired...

The brain is a repository of experiences associating stimuli to actuators

massive parallel wiring!

10 Billions neurons
100 Hz
25 Watts

Patented time domain multiplexing

NeuroMem is a parallel neural network associating input patterns to categories

256 8-bit inputs
262,144 Synapses/chip
(256 * 1024)
32,766 outputs

1024 neurons
27 Mhz / 0.5 Watts
The power of a non-linear classifier

Radial Basis Function to associate and discriminate complex and large datasets

Embedded regions (A and B into C)

Overlapping regions (A and B)

Disjoint regions
Decision = Non linear classification

Input parameters =
- Voltage,
- Torque,
- Temperature,
- Vibration
- and more

Class = Operation modes such as
- Warm up
- Light load
- Medium load
- Heavy load
- Cool down

Anomaly!
The power of small footprint and low power

Example: Predictive Maintenance for Ball Bearing
Semi-supervised learning
Logging of abnormal vibrations
Novelties can be reviewed at later time and added to knowledge
Consumer and mobile devices

- 3D Gesture human machine interface
- Activity monitoring
- Blood pressure monitoring
- Gaze tracking
Fig. 1: PPG pulse.
Photoplethysmogram features associated with hypertension
January 2010 - January 2019
NeuroMem® chips are available TODAY in large volume
ANM5500 by our licensee Alfaplus (Taiwan)
NM500 by General Vision (Manufactured by Nepes Korea)

Smallest size chip with
on chip real time incremental
Learning capability hardwired
Rich development tools ecosystem
Still a long way to go before reaching this level of autonomy with almost no energy!
Let's be a little realistic about so called AI achievement!
Thanks to the internet, the storage and CPU's for the "hyper-communication"

Courtesy of General Vision

Silicon Valley / San Francisco!
After 28 days in the eggs and five week outside
No software, no GPU, minimum energy
They learned real time, incrementally how to walk, swim, and follow Mom and Dad by looking at them ...
In a few weeks they will fly miles!!!
Copyright Notice

This multimedia file is copyright © 2023 by tinyML Foundation. All rights reserved. It may not be duplicated or distributed in any form without prior written approval.

tinyML® is a registered trademark of the tinyML Foundation.

www.tinyml.org
Copyright Notice

This presentation in this publication was presented as a tinyML® Talks webcast. The content reflects the opinion of the author(s) and their respective companies. The inclusion of presentations in this publication does not constitute an endorsement by tinyML Foundation or the sponsors.

There is no copyright protection claimed by this publication. However, each presentation is the work of the authors and their respective companies and may contain copyrighted material. As such, it is strongly encouraged that any use reflect proper acknowledgement to the appropriate source. Any questions regarding the use of any materials presented should be directed to the author(s) or their companies.

tinyML is a registered trademark of the tinyML Foundation.

www.tinyml.org