“Inference with Raspberry Pi Pico and RP2040”

Eben Upton - Raspberry Pi Foundation

March 4, 2021

www.tinyML.org
tinyML Talks Sponsors

arm
Deeplite
EDGE IMPULSE
maxim integrated
Qeexo
Reality AI
SynSense

Additional Sponsorships available – contact sponsorships@tinyML.org for info
Arm: The Software and Hardware Foundation for tinyML

1. Connect to high-level frameworks
2. Supported by end-to-end tooling
3. Connect to Runtime

Profiling and debugging tooling such as Arm Keil MDK

Application

Optimized models for embedded

Runtime (e.g. TensorFlow Lite Micro)

Optimized low-level NN libraries (i.e. CMSIS-NN)

RTOS such as Mbed OS

Arm Cortex-M CPUs and microNPUs

AI Ecosystem Partners

Stay Connected

@ArmSoftwareDevelopers

@ArmSoftwareDev

Resources: developer.arm.com/solutions/machine-learning-on-arm
WE USE AI TO MAKE OTHER AI FASTER, SMALLER AND MORE POWER EFFICIENT

Automatically compress SOTA models like MobileNet to <200KB with little to no drop in accuracy for inference on resource-limited MCUs

Reduce model optimization trial & error from weeks to days using Deeplite's design space exploration

Deploy more models to your device without sacrificing performance or battery life with our easy-to-use software

BECOME BETA USER bit.ly/testdeeplite
TinyML for all developers

**Dataset**
- Acquire valuable training data securely
- Enrich data and train ML algorithms

**Edge Device**
- Real sensors in real time
- Open source SDK

**Impulse**
- Embedded and edge compute deployment options
- Test impulse with real-time device data flows

**Test**
- Get your free account at http://edgeimpulse.com

Copyright © EdgeImpulse Inc.
健康传感器测量PPG和ECG信号，对理解关键生命体征至关重要。信号链产品能够测量甚至最敏感的信号。

低功耗Cortex M4微控制器
最大的（3MB闪存和1MB SRAM）和最小的（256KB闪存和96KB SRAM）Cortex M4微控制器使算法和神经网络能够在可穿戴电源水平下运行。

高级AI加速
新的MAX78000在嵌入式选项中实现了AI推断的100多倍低能效。现在，边缘可以像以前从未有过的那样看到和听到。

健康传感器测量PPG和ECG信号，对理解关键生命体征至关重要。信号链产品能够测量甚至最敏感的信号。

低功耗Cortex M4微控制器
最大的（3MB闪存和1MB SRAM）和最小的（256KB闪存和96KB SRAM）Cortex M4微控制器使算法和神经网络能够在可穿戴电源水平下运行。

高级AI加速
新的MAX78000在嵌入式选项中实现了AI推断的100多倍低能效。现在，边缘可以像以前从未有过的那样看到和听到。
Qeexo AutoML for Embedded AI
Automated Machine Learning Platform that builds tinyML solutions for the Edge using sensor data

**Key Features**

- Wide range of ML methods: GBM, XGBoost, Random Forest, Logistic Regression, Decision Tree, SVM, CNN, RNN, CRNN, ANN, Local Outlier Factor, and Isolation Forest
- Easy-to-use interface for labeling, recording, validating, and visualizing time-series sensor data
- On-device inference optimized for low latency, low power consumption, and a small memory footprint
- Supports Arm® Cortex™- M0 to M4 class MCUs
- Automates complex and labor-intensive processes of a typical ML workflow – no coding or ML expertise required!

**Target Markets/Applications**

- Industrial Predictive Maintenance
- Smart Home
- Wearables
- Automotive
- Mobile
- IoT

QEEXO AUTOML: END-TO-END MACHINE LEARNING PLATFORM

For a limited time, sign up to use Qeexo AutoML at automl.qeexo.com for FREE to bring intelligence to your devices!
Reality AI Tools® software

- Automated Feature Exploration and Model Generation
- Bill-of-Materials Optimization
- Automated Data Assessment
- Edge AI / TinyML code for the smallest MCUs

Reality AI solutions

- Automotive sound recognition & localization
- Indoor/outdoor sound event recognition
- RealityCheck™ voice anti-spoofing

https://reality.ai    info@reality.ai    @SensorAI    Reality AI
SynSense builds ultra-low-power (sub-mW) sensing and inference hardware for embedded, mobile and edge devices. We design systems for real-time always-on smart sensing, for audio, vision, IMUs, bio-signals and more.

https://SynSense.ai
## Next tinyML Talks

<table>
<thead>
<tr>
<th>Date</th>
<th>Presenter</th>
<th>Topic / Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday, March 16</td>
<td>Vijay Janapa Reddi</td>
<td>tinyMLPerf: Deep Learning Benchmarks for Embedded Devices</td>
</tr>
<tr>
<td></td>
<td>Associate Professor, Harvard University</td>
<td></td>
</tr>
</tbody>
</table>

Webcast start time is 8 am Pacific time

Please contact [talks@tinyml.org](mailto:talks@tinyml.org) if you are interested in presenting
Announcement

https://www.tinyml.org/event/summit-2021/

Highlights:
- Keywords: Premier Quality, Interactive, LIVE ... and FREE
- 5 days, 50+ presentations
- 4 Tutorials
- 2 Panel discussions: (i) VC and (ii) tinyML toolchains
- tinyML Research Symposium
- Late Breaking News
- 3 Best tinyML Awards (Paper, Product, Innovation)
- 10+ Breakout sessions on various topics
- tinyML Partner sessions
- tinyAI for (Good) Life
- LIVE coverage, starting at 8am Pacific time

What should I do about it:
- Check out the program – you will be impressed
- Register on-line (takes 5 min)
- If interested: Submit nominations for Best Awards and/or Late News – February 28 deadline
- Block out your calendar: March 22-26
- Become a sponsor (sponsorships@tinyML.org)
- Actively participate at the Summit
- Provide your feedback – we listen!
- Don’t worry about missing some talks – all videos will be posted on YouTube.com/tinyML
tinyML is growing fast

<table>
<thead>
<tr>
<th></th>
<th>2019 Summit (March 2019)</th>
<th>2020 Summit (Feb 2020)</th>
<th>2021 Summit (March 2021), expected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attendees</td>
<td>160</td>
<td>400+</td>
<td>3000+</td>
</tr>
<tr>
<td>Companies</td>
<td>90</td>
<td>172</td>
<td>300+ (?)</td>
</tr>
<tr>
<td>LinkedIn members</td>
<td>0</td>
<td>798</td>
<td>~ 2000</td>
</tr>
<tr>
<td>Meetups members</td>
<td>0</td>
<td>1140</td>
<td>~ 5000</td>
</tr>
<tr>
<td>YouTube subscribers</td>
<td>0</td>
<td>0</td>
<td>~ 3000</td>
</tr>
</tbody>
</table>

Also started in Asia: tinyML WeChat and BiliBili
Summit Sponsors
(as of Feb 15, 2021)

Contact: sponsorships@tinyML.org

multiple levels and benefits available
(also check www.tinyML.org)
Reminders

Slides & Videos will be posted tomorrow

tinyml.org/forums  youtube.com/tinyml

Please use the Q&A window for your questions
Dr Eben Upton CBE FREng DFBCS HonFIET is a founder of the Raspberry Pi Foundation, a former Distinguished Engineer with fabless semiconductor manufacturer Broadcom Inc, and founder and former CTO of mobile games middleware developer Idea works 3d Ltd. He holds a BA in Physics and Engineering, a PhD in Computer Science, and an MBA, from the University of Cambridge.
Raspberry Pi Pico

• Raspberry Pi Pico is a new $4 board
• Built on our RP2040 microcontroller
  – “Just” a break-out board
  – But with a nice power chain...
  – ...and 2MB of QSPI Flash
• Showing promise as an ML platform
• Our first in-house silicon design
• Dual Cortex-M0+ @ 133MHz
• 264KB on-die SRAM
• “Flashless” architecture
• Simple, deterministic bus fabric
• Rich peripheral set
  – UART, SPI, I2C
  – USB 1.1
  – Programmable I/O (PIO)
• Third-party boards also available
RP2040 detail
Programmable I/O

- IRQ Masking
- FIFOs
- State Machines: 0, 1, 2, 3
- Instruction Memory: 32 instructions, 4 Read Ports
- GPIO Input (x32)
- GPIO Output (x32)
- GPIO Output Enable (x32)
- From TX FIFO: Out Shift, Scratch X, To RX FIFO: In Shift, Scratch Y
- To instruction memory: PC, Clock Div, Control Logic
- From instruction memory (or bus)
- IRQ Set, Clear, Status
Early ML work

- **The good**
  - High clock rate
  - Dual core
  - Large on-chip SRAM

- **The bad**
  - No SIMD
  - No single-cycle MAC
  - Currently limited sensor choice

- **Initial TensorFlow Lite port**
  - Stock clocks (2×)
  - Single-core (2×)
  - Model parameters in SPI Flash (2.7×)

<table>
<thead>
<tr>
<th>Device</th>
<th>keyword</th>
<th>person detect</th>
</tr>
</thead>
<tbody>
<tr>
<td>SparkFun Edge (Cortex-M4 @ 48MHz)</td>
<td></td>
<td>800ms</td>
</tr>
<tr>
<td>SparkFun Edge (Cortex-M4 @ 96MHz)</td>
<td></td>
<td>400ms</td>
</tr>
<tr>
<td>Arduino BLE Sense Nano (Cortex-M4 @ 64MHz)</td>
<td></td>
<td>600ms</td>
</tr>
<tr>
<td>Raspberry Pi Pico (Cortex-M0+ @ 125MHz; model in Flash)</td>
<td>10.2ms</td>
<td>2200ms</td>
</tr>
<tr>
<td>Raspberry Pi Pico (Cortex-M0+ @ 125MHz; model in RAM)</td>
<td>3.8ms</td>
<td></td>
</tr>
</tbody>
</table>
Future directions

• ML-focused third-party boards
  – SparkFun MicroMod RP2040
  – Arduino Nano RP2040 Connect
  – ArduCam Pico4ML

• Optimised TensorFlow Lite
  – 1.2V operating point
  – Dual-core support
  – Streaming model parameters via DMA

• Other frameworks

• Future silicon
  – Lightweight (4-8MACs/clock) accelerators
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