“Making ML work in the real world”

Dominic Binks - Audio Analytic

UK Area Group – February 23, 2021

www.tinyML.org
tinyML Talks Sponsors

arm

tinyML Strategic Partner

EDGE IMPULSE

maxim integrated™

Qeexo

Reality AI®

SynSense

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

1. Connect to high-level frameworks
   - Profiling and debugging tooling such as Arm Keil MDK

2. Supported by end-to-end tooling
   - Optimized models for embedded
     - Runtime (e.g., TensorFlow Lite Micro)
     - Optimized low-level NN libraries (i.e., CMSIS-NN)

3. Connect to Runtime
   - RTOS such as Mbed OS
   - Arm Cortex-M CPUs and microNPUs

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

Get your free account at http://edgeimpulse.com
Health sensors measure PPG and ECG signals critical to understanding vital signs. Signal chain products enable measuring even the most sensitive signals.

The biggest (3MB flash and 1MB SRAM) and the smallest (256KB flash and 96KB SRAM) Cortex M4 microcontrollers enable algorithms and neural networks to run at wearable power levels.

The new MAX78000 implements AI inferences at over 100x lower energy than other embedded options. Now the edge can see and hear like never before.
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!
is for building products

https://reality.ai  info@reality.ai  @SensorAI  Reality AI

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
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 2</td>
<td>Eben Upton founder of the Raspberry Pi Foundation</td>
<td>Inference with Raspberry Pi Pico and RP2040</td>
</tr>
<tr>
<td>tinyML UK</td>
<td><strong>Arduino</strong></td>
<td>Talk on ML on Arduino platforms</td>
</tr>
<tr>
<td>Tuesday, April 20</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Webcast start time is 8 am Pacific time

Please contact talks@tinylm.org if you are interested in presenting
Announcement

www.tinyML.org/summit2021

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><strong>Attendees</strong></td>
<td>160</td>
<td>400+</td>
<td>3000+</td>
</tr>
<tr>
<td><strong>Companies</strong></td>
<td>90</td>
<td>172</td>
<td>300+ (?)</td>
</tr>
<tr>
<td><strong>Linkedin members</strong></td>
<td>0</td>
<td>798</td>
<td>~ 2000</td>
</tr>
<tr>
<td><strong>Meetups members</strong></td>
<td>0</td>
<td>1140</td>
<td>~ 5000</td>
</tr>
<tr>
<td><strong>YouTube subscribers</strong></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)
Alessandro Grande  
Developer advocate & ecosystem manager, Arm

Dominic Binks  
VP Technology, Audio Analytic

Gian Marco Iodice  
ML Techlead, Arm

Neil Cooper  
VP Marketing, Audio Analytic
Reminders

Slides & Videos will be posted tomorrow

tinyml.org/forums  youtube.com/tinyml

Please use the Q&A window for your questions
Dr Dominic Binks was previously a Staff Engineer at Qualcomm working in a variety of different software roles prior to joining Audio Analytic. At Qualcomm in Cambridge, he worked on mShop, a BREW-based shopping application and Vuforia, Qualcomm’s cross-platform augmented reality SDK. In addition Dominic spent time in San Diego working on Qualcomm’s core Android porting team with responsibility for the build and release team. Prior to Qualcomm, Dominic worked in technical presales at SavaJe, Android’s forerunner. Before joining SavaJe Dominic worked as a technical consultant at Scientific Generics (now Sagentia) and prior to this he worked on pre-paid calling platforms deployed to a number of mobile operators worldwide. Dominic’s PhD investigated techniques for automating fault finding (debugging) in pieces of software.
Giving machines a sense of hearing
“Like a Shazam for real-world sounds”

Bloomberg

as featured in

THE TIMES  The Economist  WIRED  FASTCOMPANY  Forbes

BBC  IEEE SPECTRUM  engadget  techradar  Gartner

© Audio Analytic Ltd, 2021
Machine Learning in the real world is hard
Google’s medical AI was super accurate in a lab. Real life was a different story.

If AI is really going to make a difference to patients we need to know how it works when real humans get their hands on it, in real situations.


Why ML is hard in the real world – some causes

• Unobserved differences
  • Images from one smartphone camera being of very different quality to those from another

• Improbable inputs
  • Share price changes very rapidly

• Unforeseen inputs
  • Parrots in France?

• Combinations
  • Leaves from a tree moving in a camera image

• Over simplification
  • Alexa (the person) verses Alexa (the digital assistant)

• Essentially all examples of under specification
“We say that an ML pipeline is *underspecified* if there are many predictors $f$ that a pipeline could return with similar predictive risk”
How do we mitigate underspecification?
Understanding use case reduces under specification
A product is defined by how it is to be *used*
Smoke alarm detection – a use case approach

- What constitutes a correct detection?
- Where will it work?
- When will it work?
- What does it mean to work correctly?
- What’s the budget for processing?
- What’s the desirable latency?
- Who will use it?
- How will a user interact with it?
How could we go about building a smoke alarm detector?
Sound Recognising as an ML Task

Unlike speech, general sound has no language model – there is no existing corpus of data with which to train how sound occurs in time.

- **Temporal modelling**: Predominantly occurrence of frequency components over longer time window (e.g. sequences of pitch elements, repetitions etc.)
- **Acoustic modelling**: Predominantly frequency composition of a short time window
When is a smoke alarm sounding?
How many smoke alarms do you need?
Acceptance
Data Wrangling

Use case

- Slicing
- Labelling
- Test datasets
- Training datasets
Training
Real World Evaluation

What and why?

Evaluation metric

Use case

Alexandria™
“There are known knowns, things we know that we know; and there are known unknowns, things that we know we don't know. But there are also unknown unknowns, things we do not know we don't know.”

Donald Rumsfeld, 2002
Thank you

info@audioanalytic.com
audioanalytic.com

UK headquarters
2 Quayside
Cambridge
CB5 8AB, UK

Sound Labs
11b Nuffield Road
Cambridge
CB4 1TF, UK

US office
44 Montgomery Street
San Francisco
CA 94104, USA
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