“Constrained Object Detection on Microcontrollers with FOMO”

Shawn Hymel - Edge Impulse

April 5, 2022
tinyML Talks Strategic Partners

Additional Sponsorships available – contact Olga@tinyML.org for info
Executive Strategic Partners
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
EDGEE IMPULSE The leading edge ML platform

www.edgeimpulse.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
Neural Decision Processors
- At-Memory Compute
- Sustained High MAC Utilization
- Native Neural Network Processing

ML Training Pipeline
- Enables Production Quality Deep Learning Deployments

Data Platform
- Reduces Data Collection Time and Cost
- Increases Model Performance

End-to-End Deep Learning Solutions for TinyML & Edge AI

partners@syntiant.com
www.syntiant.com
Platinum Strategic Partners
Fastest Video Analytics Solutions on Arm CPUs
Add Advanced Sensing to your Product with Edge AI / TinyML

Pre-built Edge AI sensing modules, plus tools to build your own

Reality AI solutions
- Pre-built sound recognition models for indoor and outdoor use cases
- Solution for industrial anomaly detection
- Pre-built automotive solution that lets cars “see with sound”

Reality AI Tools® software
- Build prototypes, then turn them into real products
- Explain ML models and relate the function to the physics
- Optimize the hardware, including sensor selection and placement

https://reality.ai  info@reality.ai  @SensorAI  Reality AI
# BROAD AND SCALABLE EDGE COMPUTING PORTFOLIO

## Microcontrollers & Microprocessors

<table>
<thead>
<tr>
<th>Arm® Core</th>
<th>Renesas Core</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Arm® Cortex®-M 32-bit MCUs</strong>&lt;br&gt;Arm ecosystem, Advanced security, Intelligent IoT</td>
<td><strong>Ultra-low Energy 8 &amp; 16-bit MCUs</strong>&lt;br&gt;Bluetooth® Low Energy, SubGHz, LoRa®-based Solutions</td>
</tr>
<tr>
<td><strong>Arm®-based High-end 32 &amp; 64-bit MPUs</strong>&lt;br&gt;High-resolution HMI, Industrial network &amp; real-time control</td>
<td><strong>High Power Efficiently 32-bit MCUs</strong>&lt;br&gt;Motor control, Capacitive touch, Functional safety, GUI</td>
</tr>
<tr>
<td><strong>Arm® Cortex®-M0+ Ultra-low Power 32-bit MCUs</strong>&lt;br&gt;Innovative process tech (SOTB), Energy harvesting</td>
<td><strong>40nm/28nm process Automotive 32-bit MCUs</strong>&lt;br&gt;Rich functional safety and embedded security features</td>
</tr>
</tbody>
</table>

**Renesas Synergy™**<br>Arm®-based 32-bit MCUs for Qualified Platform<br>Qualified software and tools

## Core technologies

### AI
A broad set of high-power and energy-efficient embedded processors

### Security & Safety
Comprehensive technology and support that meet the industry’s stringent standards

### Digital & Analog & Power Solution
Winning Combinations that combine our complementary product portfolios

### Cloud Native
Cross-platforms working with partners in different verticals and organizations
Gold Strategic Partners
The new MAX78000 implements AI inferences at low energy levels, enabling complex audio and video inferencing to run on small batteries. Now the edge can see and hear like never before.

www.maximintegrated.com/MAX78000

Large (3MB flash + 1MB SRAM) and small (256KB flash + 96KB SRAM, 1.6mm x 1.6mm) Cortex M4 microcontrollers enable algorithms and neural networks to run at wearable power levels.

www.maximintegrated.com/microcontrollers

Health sensors measure PPG and ECG signals critical to understanding vital signs. Signal chain products enable measuring even the most sensitive signals.

www.maximintegrated.com/sensors
Build Smart IoT Sensor Devices From Data

SensiML pioneered TinyML software tools that auto generate AI code for the intelligent edge.

- End-to-end AI workflow
- Multi-user auto-labeling of time-series data
- Code transparency and customization at each step in the pipeline

We enable the creation of production-grade smart sensor devices.

sensiml.com
SynSense builds sensing and inference hardware for ultra-low-power (sub-mW) 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
tinyML Summit 2022
Miniature dreams can come true...
March 28-30, 2022
Hyatt Regency San Francisco Airport
https://www.tinyml.org/event/summit-2022/

The Best Product of the Year and the Best Innovation of the Year awards are open for nominations between November 15 and March 14.

tinyML Research Symposium 2022
March 28, 2022
https://www.tinyml.org/event/research-symposium-2022

More sponsorships are available: sponsorships@tinyML.org
Our next tinyML Trailblazers Series
Success Stories with Eric Pan
(Founder, Seeed Studio and Chaihuo Makerspace)

LIVE ONLINE April 6th, 2022 at 8 am PST

Register now!
Join Growing tinyML Communities:

tinyML - Enabling ultra-low Power ML at the Edge

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

8.8k members in 45 Groups in 35 Countries

2.7k members & 5.7k followers
Subscribe to tinyML YouTube Channel for updates and notifications (including this video)
www.youtube.com/tinyML
# Next tinyML Talks

<table>
<thead>
<tr>
<th>Date</th>
<th>Presenter</th>
<th>Topic / Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday, April 5</td>
<td>Daniel Konegen, Embedded AI and data science engineer, Hahn-Schickard</td>
<td>AutoFlow - an open source Framework to automatically implement neural networks on embedded devices</td>
</tr>
<tr>
<td></td>
<td>Marcus Rüb, Research assistant, Hahn-Schickard</td>
<td></td>
</tr>
</tbody>
</table>

Webcast start time is 9:30 am Pacific time

Please contact talks@tinyml.org if you are interested in presenting
Reminders

Slides & Videos will be posted tomorrow

Please use the Q&A window for your questions

tinyml.org/forums  youtube.com/tinyml
Shawn Hymel

Shawn is a machine learning DevRel engineer, instructor, and university program manager at Edge Impulse. He creates compelling technical videos, courses, and blog posts around edge machine learning and embedded systems that inspire and teach engineers of all skill levels. Shawn is an advocate for enriching education through STEM and believes that the best marketing comes from teaching. He can be found giving talks, running workshops, and swing dancing in his free time.
Constrained Object Detection on Microcontrollers with FOMO
Agenda

1. What is Edge Impulse?
2. Object Detection and Image Segmentation
3. Constrained Object Detection
4. Use Cases and Limitations
5. Live Demo
Edge Impulse
Go to market faster with confidence
Deploy to any edge device with ease

- The largest silicon ecosystem
- Award-winning compiler
- Access to device source code
- Full firmware integration for a number of devices
Object Detection + Image Segmentation
Object Detection

Bounding box

cat 0.98

dog 0.83
R-CNN

Region Proposal → CNN → SVM

Class predictions:
- background
- ball
- dog
- toy

Regressor

\[ d_x, d_y, d_w, d_h \]
Single Shot MultiBox Detector (SSD)

CNN backbone

Non-Maximum Suppression (NMS)

Class predictions + boundary boxes

e.g. VGG19, MobileNet
Image Segmentation
Constrained Object Detection
MobileNet V2

Convolutional layer(s) → Bottleneck residual block 1 → Bottleneck residual block 2 → Bottleneck residual block 3 → Bottleneck residual block 4 → Bottleneck residual block 17 → Convolutional layer(s) → Fully Connected layer → Softmax

Class predictions for image

240x240
Faster Objects, More Objects (FOMO)

Height and width are each divided by 8 (default)

240x240

Weights in residual blocks are pre-trained from ImageNet

16 feature maps, each 30x30 cells

2D convolution with 1x1 kernel is used

Fully Connected layer

Fully Connected layer

Fully Connected layer

Softmax

Class predictions per cell

Looks like segmentation of feature maps
Faster Objects, More Objects (FOMO)

Each cell is given scores:
- $P(\text{background})$
- $P(\text{ball})$
- $P(\text{dog})$
- $P(\text{toy})$
Faster Objects, More Objects (FOMO)
Faster Objects, More Objects (FOMO)

Example: screws
  - Grayscale
  - Image: 96x96
  - Feature maps: 12x12
Faster Objects, More Objects (FOMO)

Example: screws
- Grayscale
- Image: 96x96
- Feature maps: 12x12
Faster Objects, More Objects (FOMO)

Example: screws
- Grayscale
- Image: 96x96
- Feature maps: 12x12

Neighboring cells with same class are removed (leaving highest scores)
FOMO Ground Truth

Example: screws
  - Grayscale
  - Image: 96x96
  - Feature maps: 12x12

User draws bounding boxes, tool picks cell with centroid of bounding box
FOMO Ground Truth

Example: screws
- Grayscale
- Image: 96x96
- Feature maps: 12x12

User draws bounding boxes, tool picks cell with centroid of bounding box

Those cells are now representatives of that class
FOMO

Uses

+ Limitations
Use Cases

Want to know **where** and **how many** objects there are

**Recommendations for success:**

- Objects are same size
- Objects are square/round
- Objects take up 1 cell

**Very fast!**

- Cortex-M7 at 480 MHz
- 240x240 image input
- 30 fps
- 245K RAM

https://matpalm.com/blog/counting_bees/
Limitations

- Each cell has its own classifier
- Small objects may be missed
- Neighboring objects may get lumped together
- Ends of oblong objects may be ignored
- Lots of objects/classes: use YOLOv5
Getting Started

docs.edgeimpulse.com/docs/

- Tutorials > Counting objects using FOMO
- Various supported dev boards
Please take the 5 question poll and continue the conversation @ tinyML.org/forums
tinyML Talks Strategic Partners

Additional Sponsorships available – contact Olga@tinyML.org for info
## Next tinyML Talks

<table>
<thead>
<tr>
<th>Date</th>
<th>Presenter</th>
<th>Topic / Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tuesday, April 5</td>
<td>Daniel Konegen, Embedded AI and data science engineer, Hahn-Schickard</td>
<td>AutoFlow - an open source Framework to automatically implement neural networks on embedded devices</td>
</tr>
<tr>
<td></td>
<td>Marcus Rüb, Research assistant, Hahn-Schickard</td>
<td></td>
</tr>
</tbody>
</table>

Webcast start time is 9:30 am Pacific time

Please contact talks@tinyml.org if you are interested in presenting
Copyright Notice

This multimedia file is copyright © 2022 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