

tinyML[®] EMEA

Enabling Ultra-low Power Machine Learning at the Edge

tinyML EMEA Technical Forum 2021 Proceedings

June 7 – 10, 2021

Virtual Event



www.tinyML.org



tinyML EMEA Technical Forum 2021

June 7-10, 2021

Mini-NAS: A Neural Architecture Search Framework for Image Classification Applications at Edge

Presented by: Shahid Siddiqui, P.hD. Student, KIOS Center of Excellence, University of Cyprus, Cyprus

June 10, 2021



tinyML EMEA Technical Forum 2021

June 7-10, 2021

- Neural Architecture Search (NAS) for benchmark image classification datasets e.g. CIFAR-10.
- What about binary classification applications for microcontrollers?
- Gradient NAS struggles, offers partial network discovery and unsuitable for constrained deployment.
- Search in existing discrete spaces e.g. with RL or evolution is expensive.

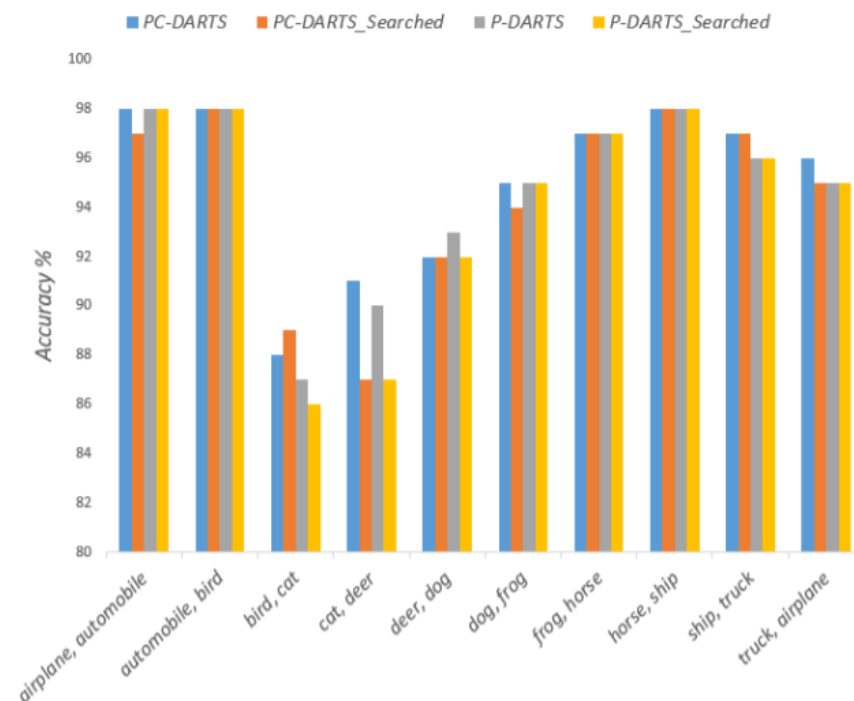


Figure 1: Gradient-based search in modular search space struggles with binary datasets. Cells searched on binary datasets are no better than cells discovered on CIFAR.



tinyML EMEA Technical Forum 2021

June 7-10, 2021

Table 1: On CIFAR-10 benchmark, Mini-NAS discovers the smallest model among methods exploring in respective global search spaces, with on par performance and search efficiency.

NAS Method	Search Network Parameters								CIFAR-10 Error (%)	Parameters (Millions)	GPU Days
	Depth (Layers)	Width (Channels)	Operations per Layer	Convolutional Kernel	Strides	Pooling Layers	Fully Connected Layers	Skip Connections			
NAS-RL [25]	✓	✓		✓	✓	✓		✓	3.65	37.4	22400
Meta-QNN [1]	✓	✓		✓	✓	✓	✓		6.92	11.18	100
Large-scale Evolution [14]	✓	✓		✓	✓			✓	5.40	5.4	2600
EAS [2]	✓	✓		✓	✓	✓	✓		4.23	23.4	10
Genetic Programming CNN [18]		✓		✓				✓	5.98	1.7	14.9
NASH-Net [7]	✓	✓		✓				✓	5.2	19.7	1
NASBOT [10]	✓	✓		✓	✓	✓	✓	✓	8.69	N/A	1.7
Mini-NAS (Ours)	✓	✓	✓	✓					5.27	1.44	2.4

- A minimal search space with multi-level adaptability to different tasks.
- A computationally affordable search algorithm.
- Use just sufficient parameters for a given dataset.
- Competitive results on CIFAR-10 dataset.



tinyML EMEA Technical Forum 2021

June 7-10, 2021

Table 2: Mini-NAS shows negligible performance drops with significant parameter efficiency as compared to MobileNetV2.

Dataset	airplane, auto	auto, bird	bird, cat	cat, deer	deer, dog	dog, frog	frog, horse	horse, ship	ship, truck	truck, airplane	airplane, auto, bird	bird, cat, deer	deer, dog, frog	frog, horse, ship	ship, truck, airplane	airplane, auto, bird, cat
Accuracy Gain	+0.15	-0.45	-0.25	-0.8	0	0	-0.7	-0.45	-0.25	-0.2	-0.27	-1.23	-0.34	-0.06	+0.43	+0.05
Parameter Efficiency	3.1x	17.6x	8.1x	14.9x	4.4x	12.5x	103.9x	20.1x	16.2x	52x	1.8x	12.5x	12.8x	9.2x	10.6x	0.6x
Dataset	cat, deer, dog, frog	frog, horse, ship, truck	airplane, auto, bird, cat, deer	dog, frog, horse, ship, truck	CIFAR-10 All	volleyball, basketball	espresso, icecream	soda bottle, beer bottle	bucket, barrell	seashore, lakeside	egyptian cat, persian cat	limo, sports	bullfrog, tailedfrog	imagenet dogs	imagenet vehicles	
Accuracy Gain	-0.57	-0.68	-1.14	+0.05	+0.29	-1	-2	+2	+1	+1	+2	-2	+3	-3	-3.07	
Parameter Efficiency	5.9x	2.8x	2.1x	3.9x	1.6x	11.9x	8.9x	15.1x	44.4x	5.9x	5.5x	11.9x	10.1x	4.9x	8.7x	

- Experiments on many smaller datasets.
- Complete network discovery.
- Negligible accuracy drop, fewer parameters → low memory footprint.
- Pruning and quantization further broaden deployment options.
- Future research directions: MCU latency aware search, training hyper-parameter search, and real-world use cases.

Premier Sponsor



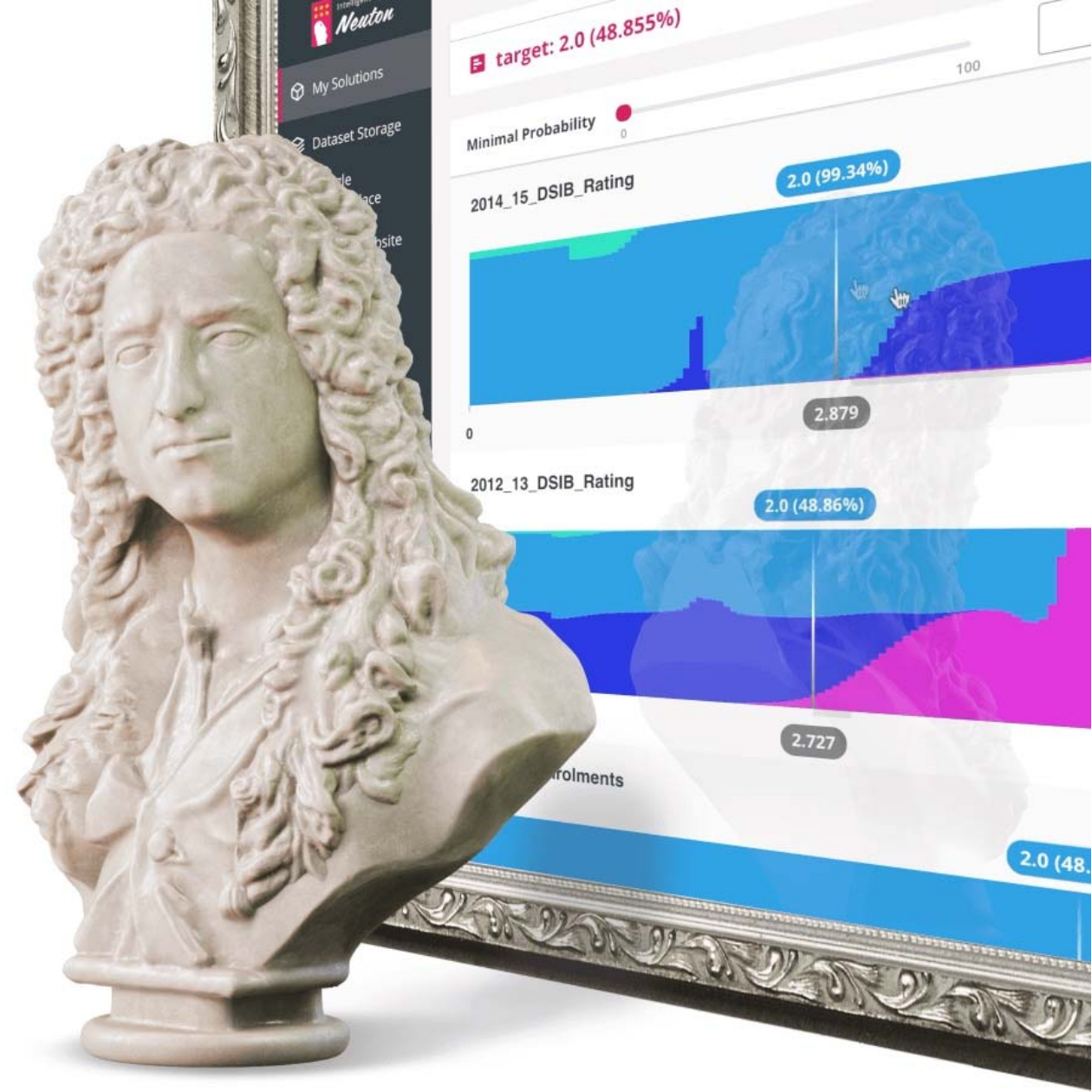
Automated TinyML

Zero-code SaaS solution

**Create tiny models, ready for embedding,
in just a few clicks!**

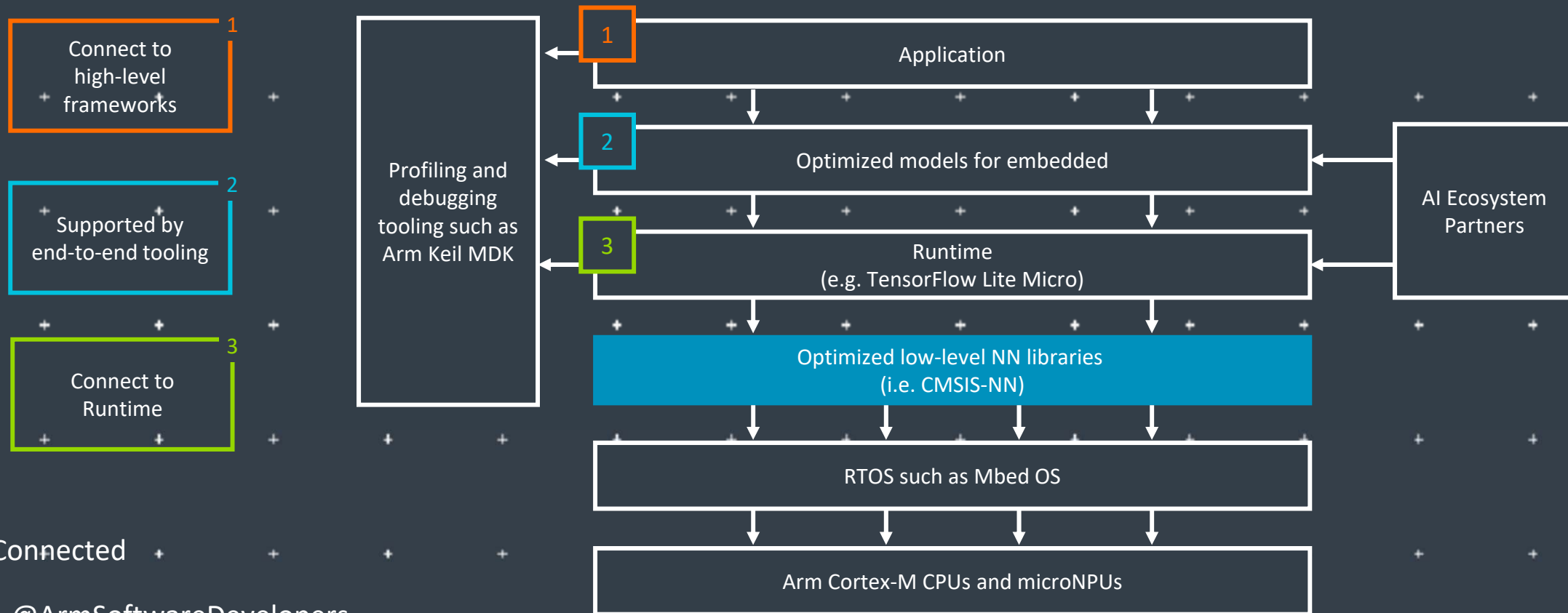
Compare the benchmarks of our compact models to those of TensorFlow and other leading neural network frameworks.

Build Fast. Build Once. Never Compromise.



Executive Sponsors

Arm: The Software and Hardware Foundation for tinyML



Stay Connected



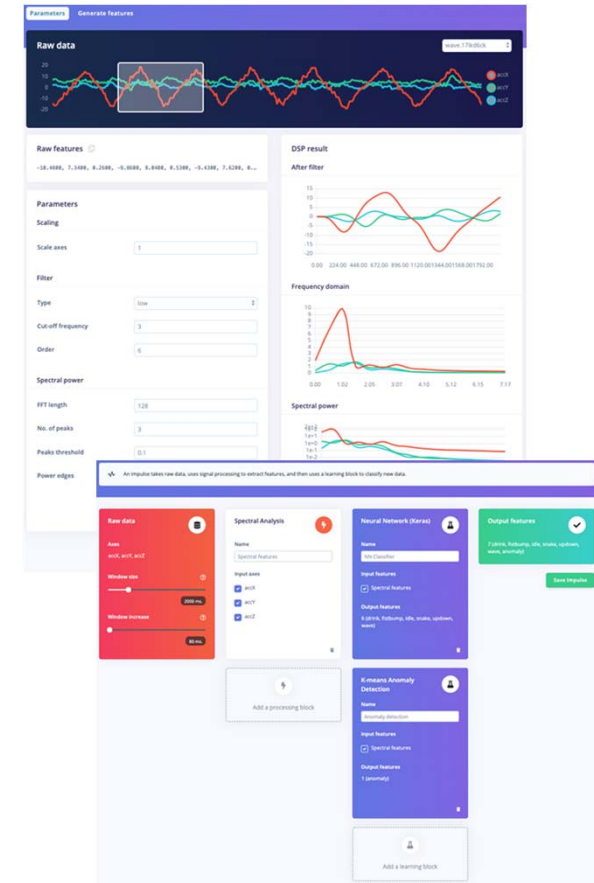
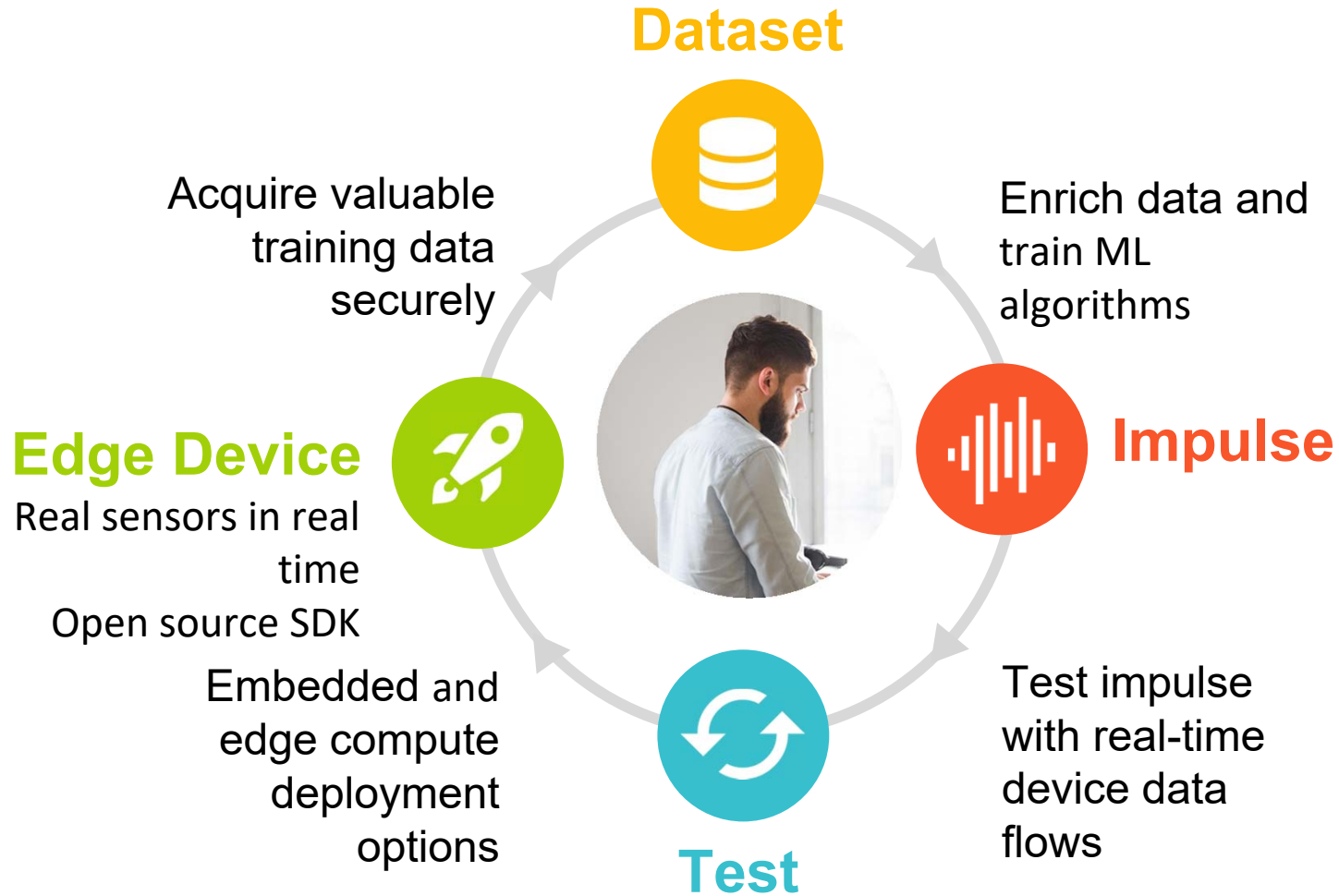
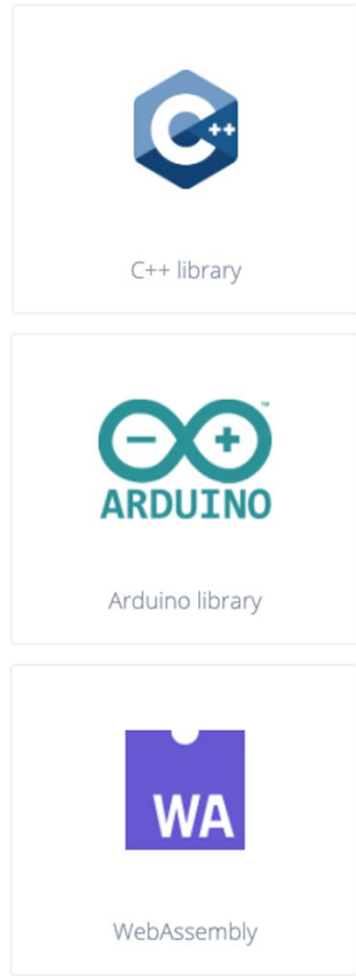
@ArmSoftwareDevelopers



@ArmSoftwareDev

Resources: developer.arm.com/solutions/machine-learning-on-arm

TinyML for all developers



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

A platform to scale AI across the industry



Perception

Object detection, speech recognition, contextual fusion



Reasoning

Scene understanding, language understanding, behavior prediction



Action

Reinforcement learning for decision making



Edge cloud



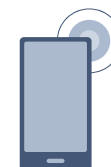
Cloud



IoT/IIoT



Automotive



Mobile

SYNTIANT

[Syntiant Corp.](#) is moving artificial intelligence and machine learning from the cloud to edge devices. Syntiant's chip solutions merge deep learning with semiconductor design to produce ultra-low-power, high performance, deep neural network processors. These network processors enable always-on applications in battery-powered devices, such as smartphones, smart speakers, earbuds, hearing aids, and laptops. Syntiant's Neural Decision Processors™ offer wake word, command word, and event detection in a chip for always-on voice and sensor applications.

Founded in 2017 and headquartered in Irvine, California, the company is backed by Amazon, Applied Materials, Atlantic Bridge Capital, Bosch, Intel Capital, Microsoft, Motorola, and others. Syntiant was recently named a [CES® 2021 Best of Innovation Awards Honoree](#), [shipped over 10M units worldwide](#), and [unveiled the NDP120](#) part of the NDP10x family of inference engines for low-power applications.

www.syntiant.com



@Syntiantcorp

Platinum Sponsors



Part of your life. Part of tomorrow.

www.infineon.com



Reality AI[®]

Add Advanced Sensing to your Product with Edge AI / TinyML

<https://reality.ai>



info@reality.ai



[@SensorAI](https://twitter.com/SensorAI)



[Reality AI](#)

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

Reality AI solutions

Prebuilt 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

Gold Sponsors



LatentAI

Adaptive AI for the Intelligent Edge

[Latentai.com](https://latent.ai)



Build Smart IoT Sensor Devices From Data

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

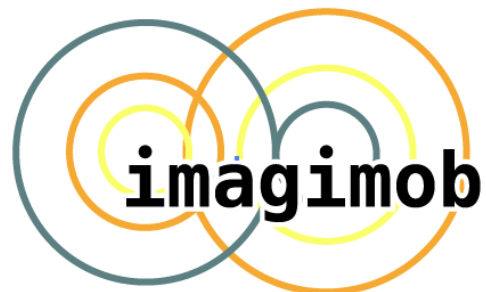
- End-to-end AI workflow
- Multi-user auto-labeling of time-series data
- Code transparency and customization at each step in the pipeline

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



sensiml.com

Silver Sponsors



Copyright Notice

The presentation(s) in this publication comprise the proceedings of tinyML® EMEA Technical Forum 2021. The content reflects the opinion of the authors and their respective companies. This version of the presentation may differ from the version that was presented at tinyML EMEA. 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