tinyML® Talks

Enabling Ultra-low Power Machine Learning at the Edge

“AIfES - an open-source standalone AI framework for almost any hardware”

Pierre Gembaczka - Fraunhofer IMS

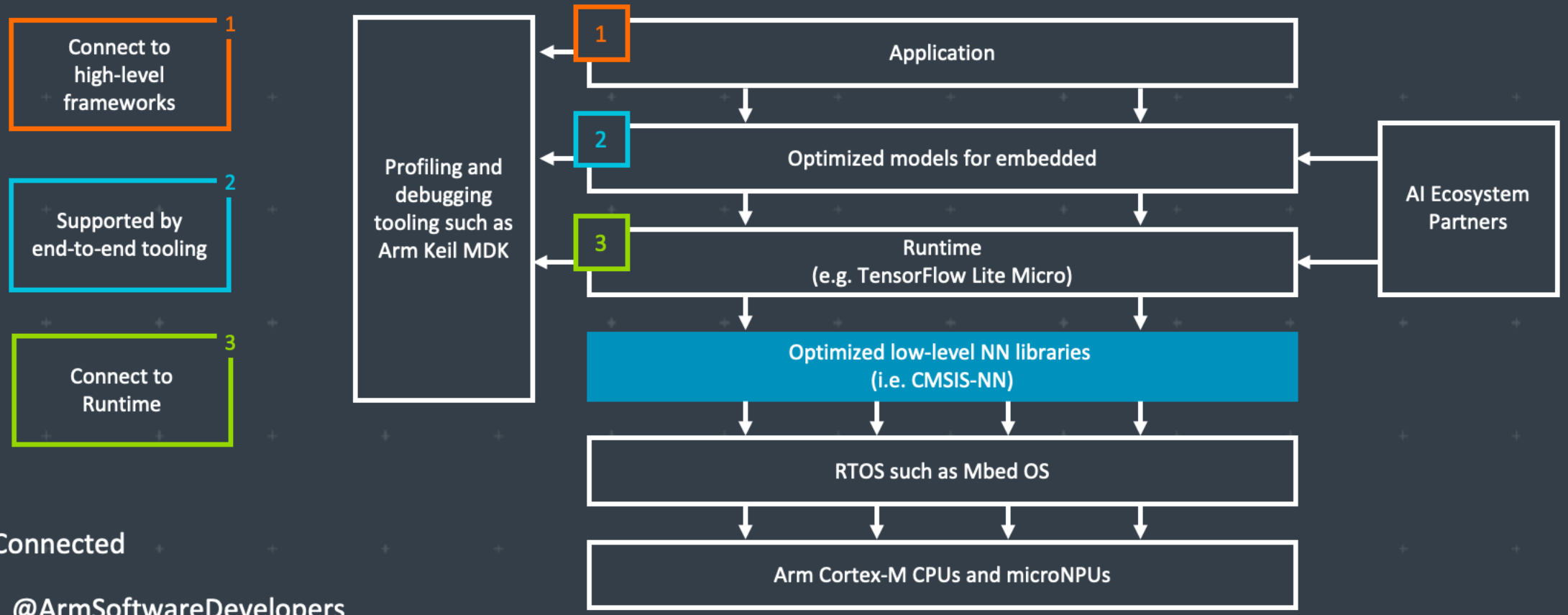
December 1, 2021

www.tinyML.org

tinyML Talks Strategic Partners

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

Arm: The Software and Hardware Foundation for tinyML



Stay Connected

@ArmSoftwareDevelopers

@ArmSoftwareDev

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

arm

Deeplite

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**

mobilityXlab arm AI 100 CBINSIGHTS

TinyML for all developers

C++ library

Arduino library

WebAssembly

Dataset

Acquire valuable training data securely

Enrich data and train ML algorithms

Edge Device

Real sensors in real time

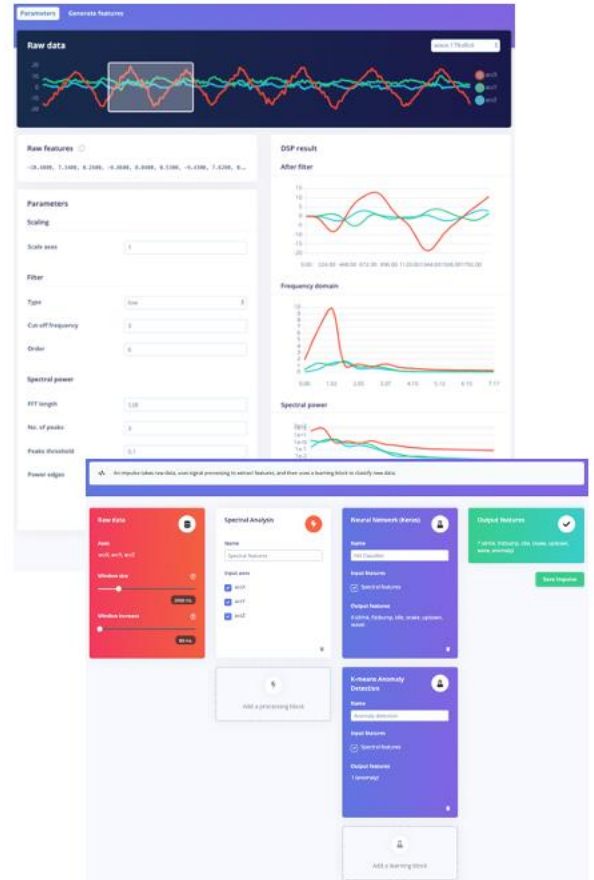
Open source SDK

Impulse

Embedded and edge compute deployment options

Test impulse with real-time device data flows

Test

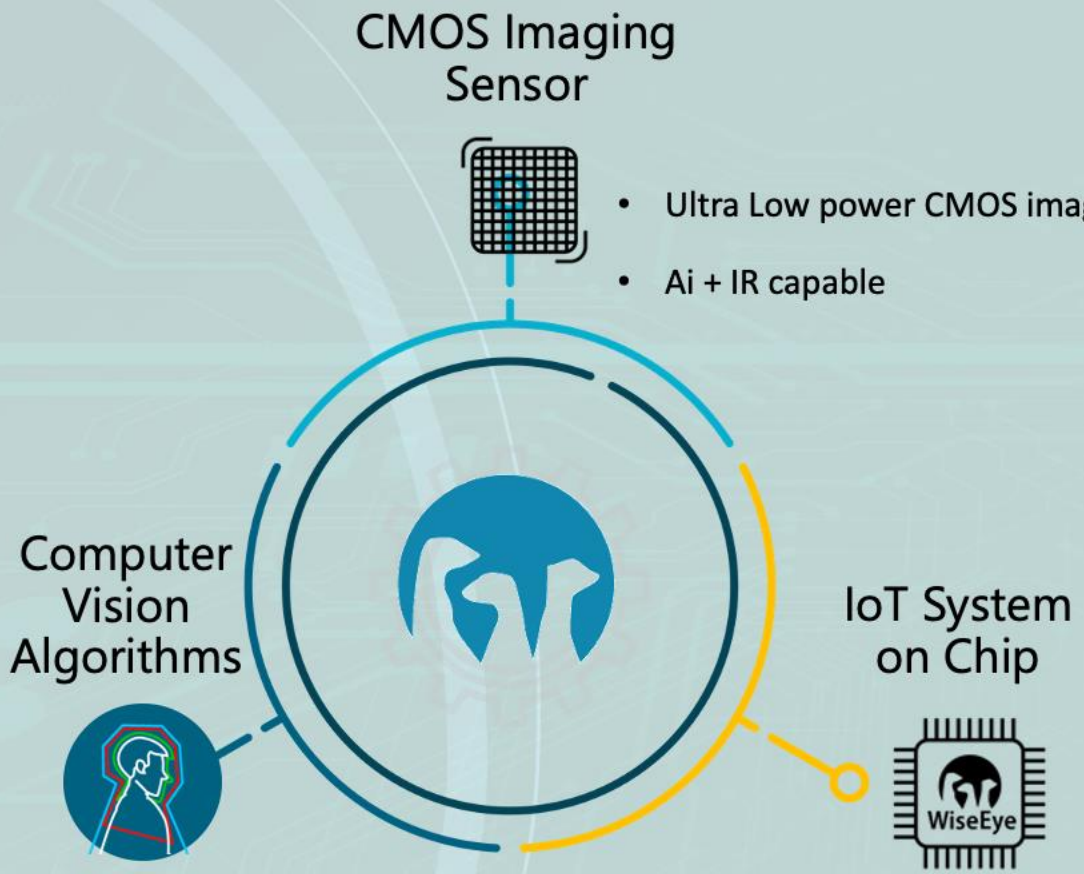


www.edgeimpulse.com

emza
visual sense

The Eye in IoT

Edge AI Visual Sensors



- Ultra Low power CMOS imager
- Ai + IR capable

- Machine Learning algorithm
- <1MB memory footprint
- Microcontrollers computing power
- Trained algorithm
- Processing of low-res images
- Human detection and other classifiers

- Machine Learning edge computing silicon
- <1mW always-on power consumption
- Computer Vision hardware accelerators

info@emza-vs.com

Enabling the next generation of Sensor and Hearable products to process rich data with energy efficiency

Visible Image

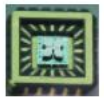
Sound

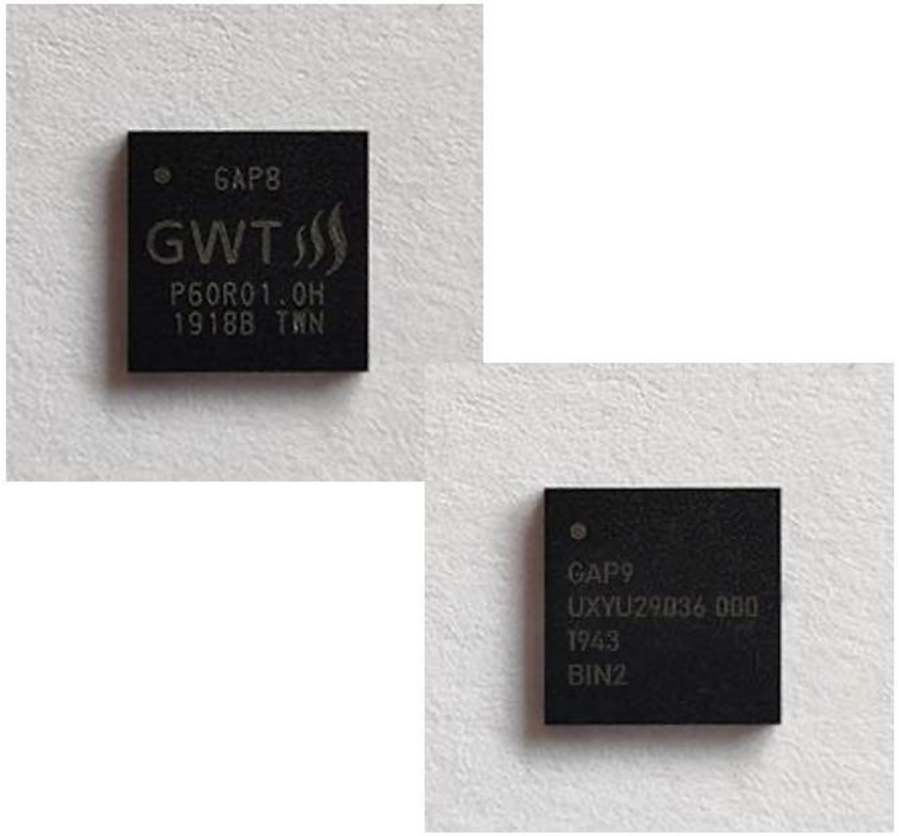
IR Image

Radar

Bio-sensor

Gyro/Accel

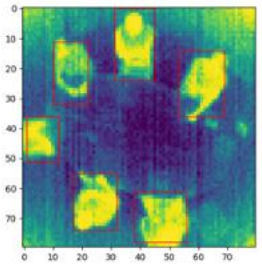




Wearables / Hearables

Battery-powered consumer electronics

IoT Sensors



GREENWAVES TECHNOLOGIES

Grovety Inc.

SOFTWARE DEVELOPMENT SERVICES FOR TINYML SOLUTIONS

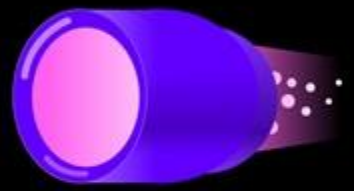
1 **Development tools**
SDK, IDE, compilers, leveraging on TVM, uTVM & LLVM

2 **Firmware**
Drivers, BSP, protocols, etc.

arm AI PARTNER

Distributed infrastructure for TinyML apps

HOTG
Decoupling intelligence



Develop at warp speed

Automate deployments

Device orchestration

HOTG is building the **distributed infrastructure** to pave the way for **AI** enabled **edge applications**

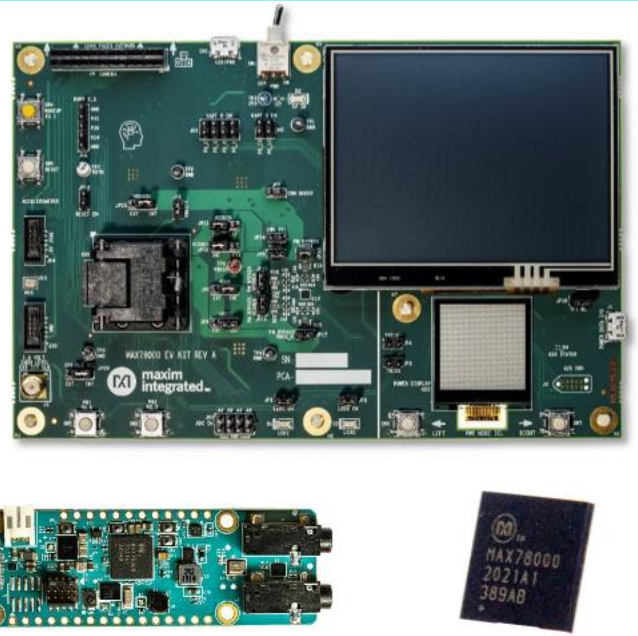
LatentAI

Adaptive AI for the Intelligent Edge

Latentai.com

Maxim Integrated: Enabling Edge Intelligence

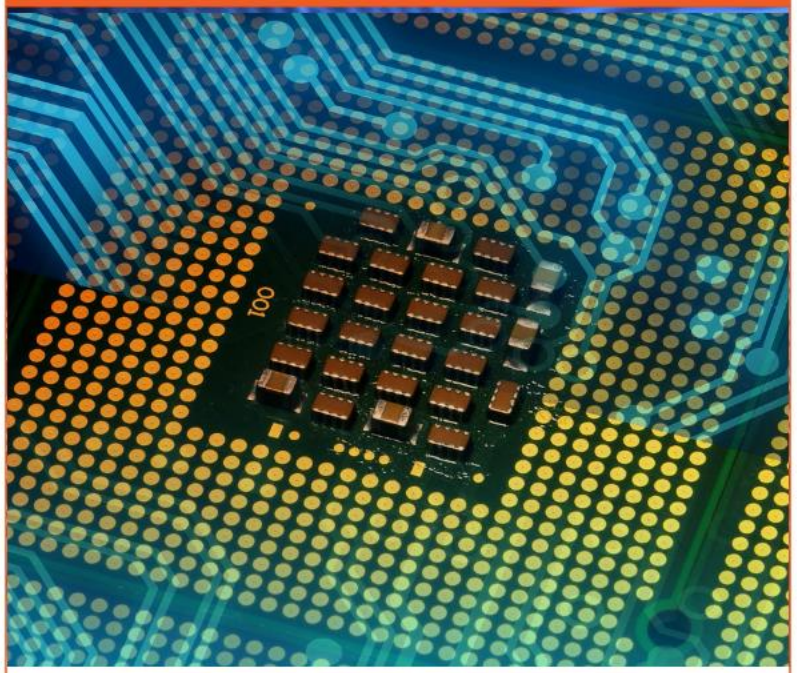
Advanced AI Acceleration IC



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

Low Power Cortex M4 Micros



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

Sensors and Signal Conditioning

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

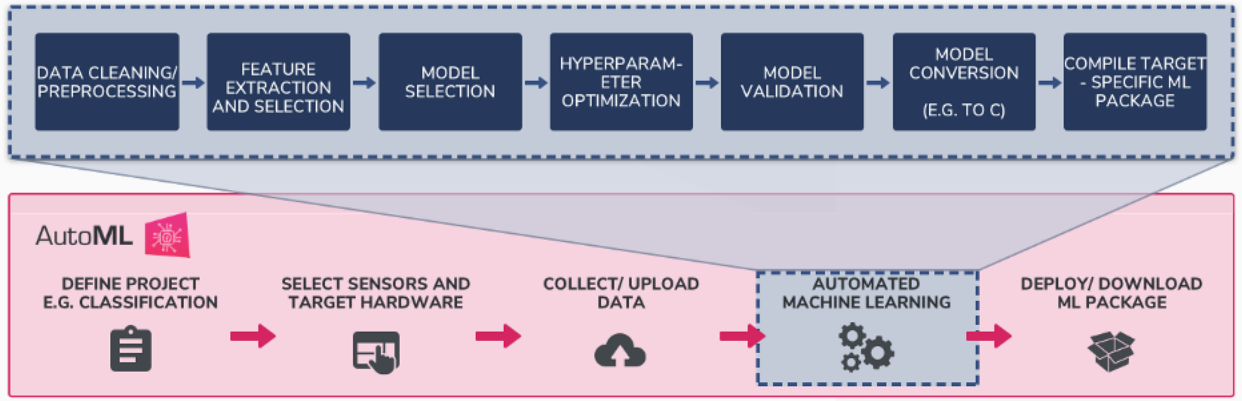
Qeexo AutoML

Automated Machine Learning Platform that builds tinyML solutions for the Edge using sensor data

Key Features

- Supports 17 ML methods:
  - Multi-class algorithms: GBM, XGBoost, Random Forest, Logistic Regression, Gaussian Naive Bayes, Decision Tree, Polynomial SVM, RBF SVM, SVM, CNN, RNN, CRNN, ANN
  - Single-class algorithms: Local Outlier Factor, One Class SVM, One Class Random Forest, Isolation Forest
- Labels, records, validates, and visualizes time-series sensor data
- On-device inference optimized for low latency, low power consumption, and small memory footprint applications
- Supports Arm® Cortex™- M0 to M4 class MCUs

End-to-End Machine Learning Platform



For more information, visit: www.qeexo.com

Target Markets/Applications

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

Qualcomm
AI research

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

Reality AI®

Add Advanced Sensing to your Product with Edge AI / TinyML

https://reality.ai info@reality.ai @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

SensiML™

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

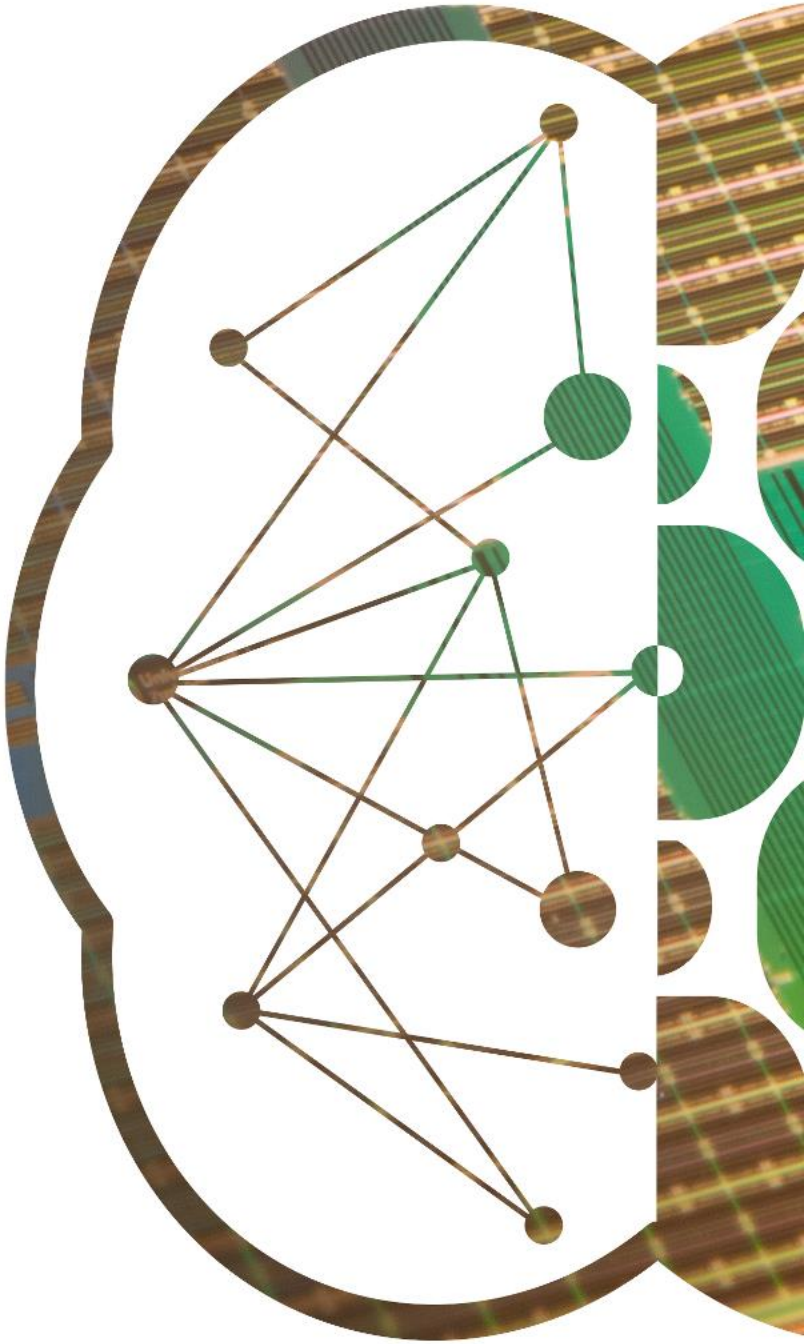
We enable the creation of production-grade smart sensor devices.

sensiml.com

SynSense

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

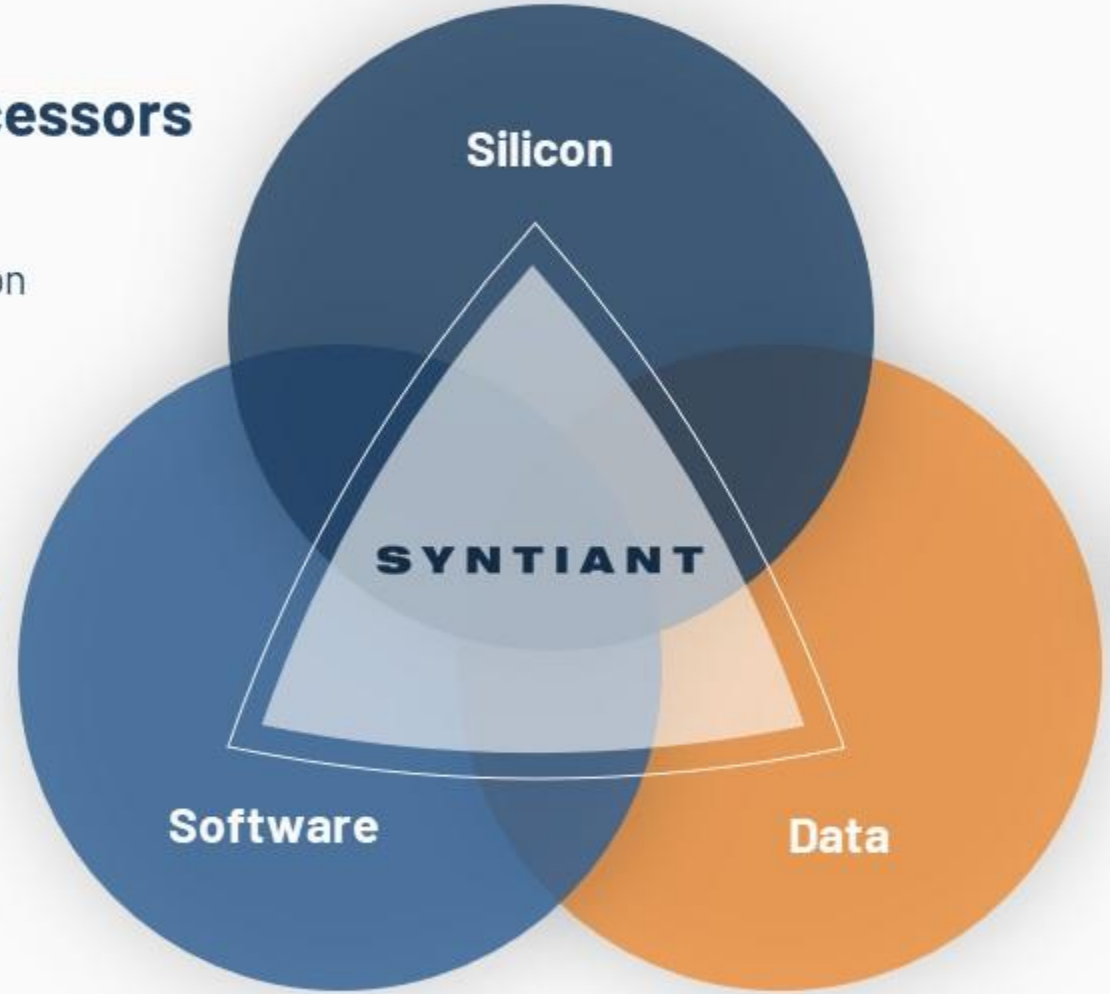


SYNTIANT

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

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

partners@syntiant.com

www.syntiant.com

tinyML Trailblazers Series

Success Stories with Pete Warden

LIVE ONLINE **December 3rd, 2021 at 8 am PST**

Trailblazers

Register now!

tinyML Summit 2022

Miniature dreams can come true…

March 28-30, 2022

Hyatt Regency San Francisco Airport

https://www.tinyml.org/event/summit-2022/

Registration will be open on **December 15**.

Deadline for poster submission is **December 17**, 2021.

*The Best Product of the Year and the Best Innovation of the Year awards are open for nominations between **November 15 and February 28.***

tinyML Research Symposium 2022

March 28, 2022

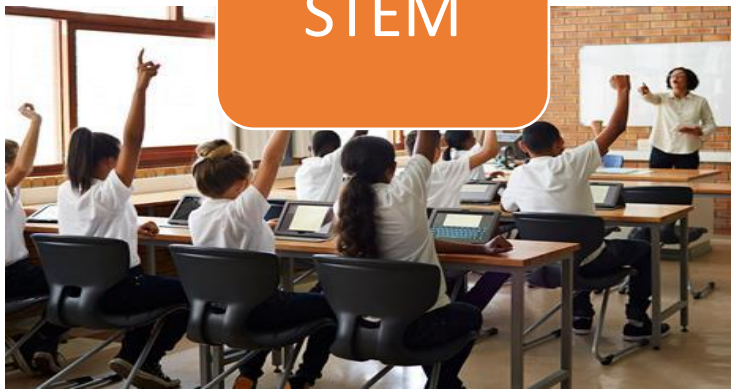
https://www.tinyml.org/event/research-symposium-2022

Call for papers – Submission deadline is **December 17**, 2021.

More sponsorships are available: sponsorships@tinyML.org

tinyML for Good – Workshop, November 17th(7 am PDT)

STEM



TINY ML For Good

Healthcare

Earth
Climate
Conservation

Contact: 4good@tinyML.org

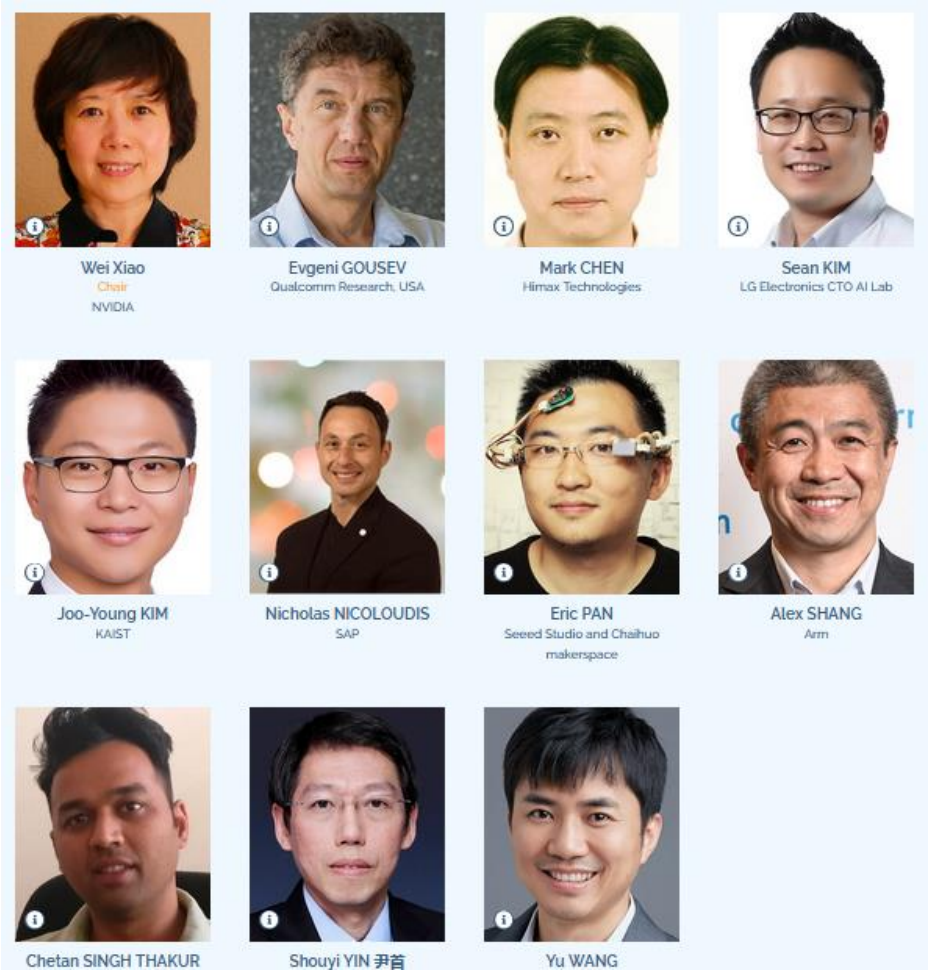
TINY ML ASIA

LIVE ONLINE **November 2-5, 2021**

(9-11:30 am China Standard time)

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

Technical Program Committee



Register today!

Free event courtesy of our sponsors and strategic partners

AONdevices arm EDGE IMPULSE emza visual sense

GREENWAVES TECHNOLOGIES Groovety Inc. HOTG imagimob

LatentAI Adaptive AI for a Smarter Edge Qualcomm Qeexo RealityAI

seeed The IoT Hardware Enabler SensiML SynSense SYNTIANT

More sponsorships are available: sponsorships@tinyML.org

Next tinyML Talks

Date	Presenter	Topic / Title
Tuesday, December 7	Chris Rogers (SensiML) and Theo Kersjes (onsemi)	The Value of Edge AI for Industrial Applications: onsemi and SensiML IIoT Solutions

Webcast start time is 8:00 am Pacific time

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

TINY ML TALKS webcast

Reminders

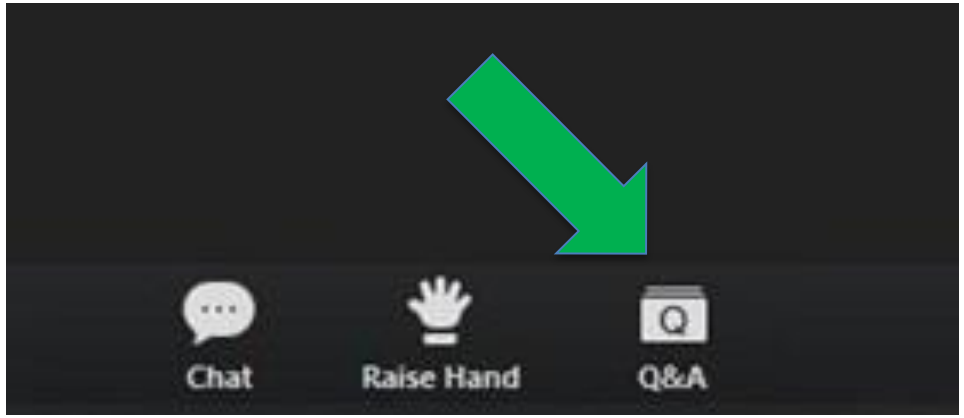
Slides & Videos will be posted tomorrow

PDF

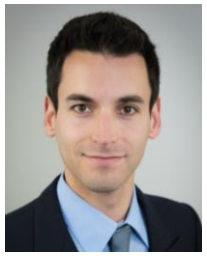
tinyml.org/forums

youtube.com/tinyml

Please use the Q&A window for your questions



Local Committee in Germany



Carlos Hernandez-Vaquero
Software Project Manager, IoT devices
Robert Bosch

Prof. Dr. Daniel Mueller-Gritschneder
Interim Head - Chair of Real-time Computer Systems
Group Leader ESL - Chair of Electronic Design Automation
Technical University of Munich

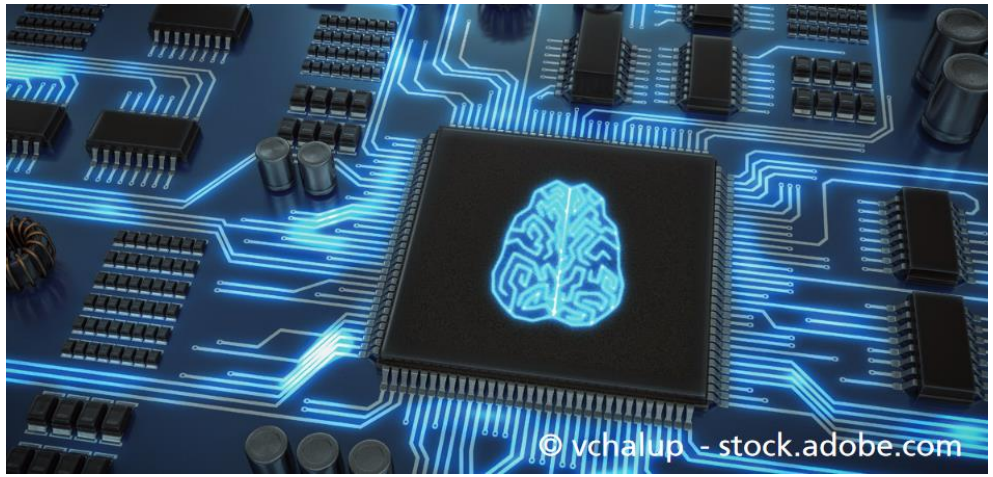
Marcus Rüb
Researcher in the field of TinyML
Hahn-Schickard

TINY ML TALKS webcast

Pierre Gembaczka

Dr. Pierre Gembaczka is Program-manager at Fraunhofer IMS. He studied Microtechnology and medical technology and holds a Master degree from the University of Applied Sciences in Gelsenkirchen. Afterwards he completed his doctorate at the Fraunhofer IMS in cooperation with the University of Duisburg Essen and obtained the academic degree of a doctor of engineering. From 2014 to 2017 he worked as a research assistant in the department Micro- and Nanosystems - Pressure Sensors at Fraunhofer IMS. From 2018 to 2020 he works as research assistant in the embedded systems group at Fraunhofer IMS and researches embedded AI solutions for various applications. He is inventor of the AI software framework AIfES (Artificial Intelligence for Embedded Systems). Since May 2020 he is Program manager “Industrial AI” and AIfES Product Manager.

What is AIfES?

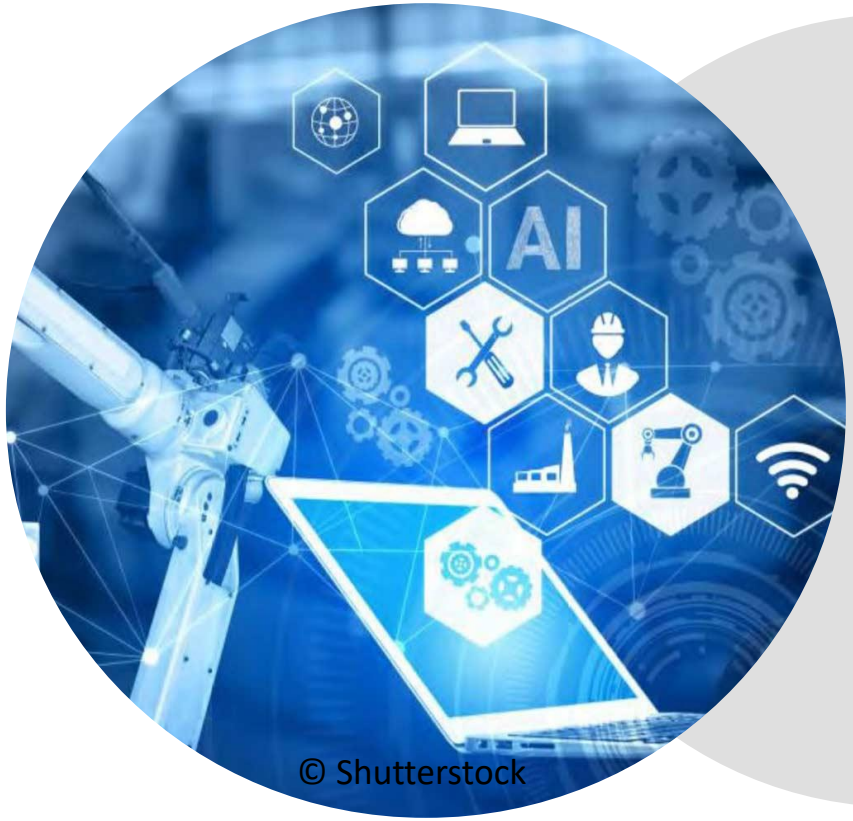


(Artificial Intelligence for Embedded Systems)

A standalone, open source, high-efficiency AI framework completely programmed in C, which allows to train and run machine learning algorithms even on the smallest microcontrollers.

Developed by Fraunhofer Institute for Microelectronic Circuits and Systems IMS

Vision & Mission



© Shutterstock

Vision

Intelligent and self-learning embedded systems.

Mission

Easy integration of machine learning (ML) right where the data is generated. In a sensor, a machine or the system independent of the hardware.

What does TinyML or Embedded-AI mean for us?

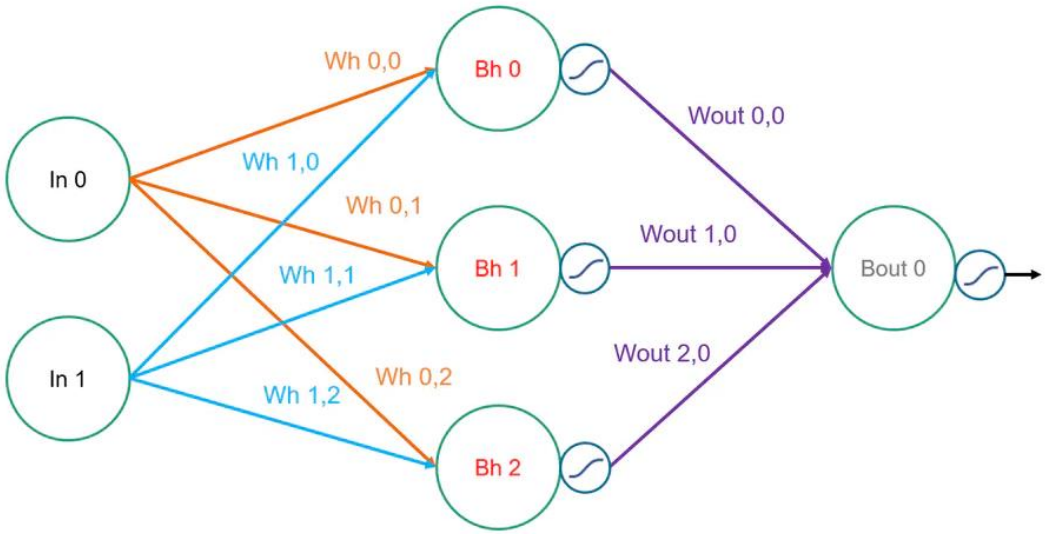
Embedded-AI as a solution for resource-limited systems

Decentralized, highly integrated AI at the point of data generation (Sensor, component, product, device) has the following advantages:

- **Fast processing**, no transmission delays
- **Increased security**, only preprocessed, protected data is transmitted
- **Increased reliability** through decentralized architecture
- **Saving resources**, reduced data volumes, reduced overall processor performance
- **Saving energy**, small and resource-saving systems like microcontrollers
- **Personalizable AI**, that autonomously optimizes itself to the application or user

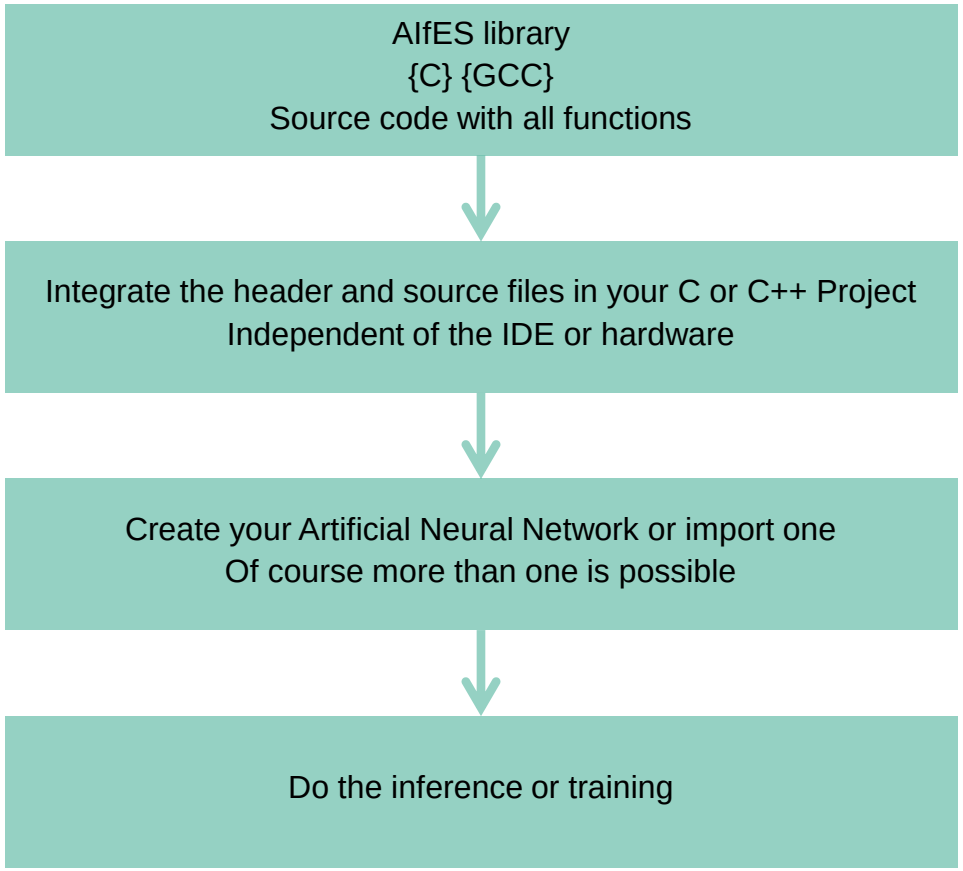
TINY ML TALKS webcast

What makes AIfES so special?



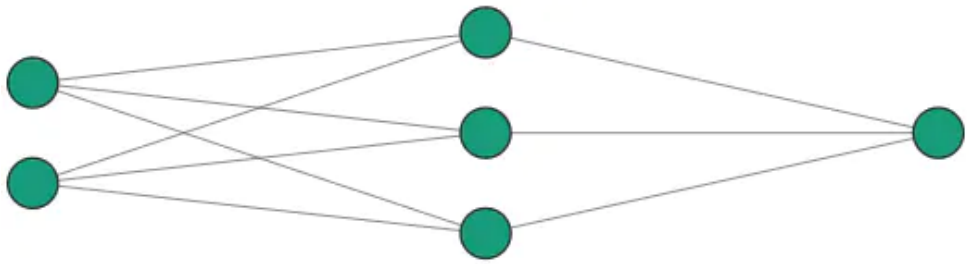
- Open source library
- Standalone AI framework, no conversions required
- Can be integrated into any C or C++ IDE
- Develop your ANN directly on the target hardware
- Inference and training of artificial neural networks (ANN)
- Programmed in ISO C and GCC compatible
- Runs on any hardware that supports GCC,
  from an 8-Bit microcontroller or Embedded Linux to a PC
- Multiple ANNs possible on one system
- Reconfigure the ANN at runtime
- Import of ANNs from other frameworks possible

How to integrate AIfES?



What does AIfES look like?

- AIfES is quite similar to the Python AI frameworks
- An ANN is described layer by layer
- Tensors are also used
- The example on the right describes a pre-trained 2-3-1 feedforward neural network (FNN)
- You can now perform the inference but also continue training



```
// Input layer
uint16_t input_layer_shape[] = {1, 2};
ailayer_input_t input_layer;
input_layer.input_dim = 2;
input_layer.input_shape = input_layer_shape;
// Dense layer (hidden layer)
float weights_data_dense_1[] = {-10.1164f, -8.4212f, 5.4396f,
7.297f, -7.6482f, -9.0155f};
float bias_data_dense_1[] = {-2.9653f,  2.3677f, -1.5968f};
ailayer_dense_t dense_layer_1;
dense_layer_1.neurons = 3;
dense_layer_1.weights.data = weights_data_dense_1;
dense_layer_1.bias.data = bias_data_dense_1;
// Sigmoid activation function
ailayer_sigmoid_t sigmoid_layer_1;
// Output dense layer
float weights_data_dense_2[] = {12.0305f, -6.5858f, 11.9371f};
float bias_data_dense_2[] = {-5.4247f};
ailayer_dense_t dense_layer_2;
dense_layer_2.neurons = 1;
dense_layer_2.weights.data = weights_data_dense_2;
dense_layer_2.bias.data = bias_data_dense_2;
// Sigmoid activation function
ailayer_sigmoid_t sigmoid_layer_2;
// ------- Define the structure of the model -------------------------
aimodel_t model;
ailayer_t *x;
// Passing the layers to the AIfES model
model.input_layer = ailayer_input_f32_default(&input_layer);
x = ailayer_dense_f32_default(&dense_layer_1, model.input_layer);
x = ailayer_sigmoid_f32_default(&sigmoid_layer_1, x);
x = ailayer_dense_f32_default(&dense_layer_2, x);
model.output_layer = ailayer_sigmoid_f32_default(&sigmoid_layer_2, x);
aialgo_compile_model(&model);
```

Are hardware accelerators supported?

Arm®

The Arm® CMSIS DSP library can be included in AIfES

- Hardware acceleration is possible for all Arm® controllers that support CMSIS DSP
- AIfES is a partner in Arm's AI Ecosystem

AIRISC by Fraunhofer IMS (link)

AIRISC: RV32IMEFC implementation – about 2,7 Coremark/MHz

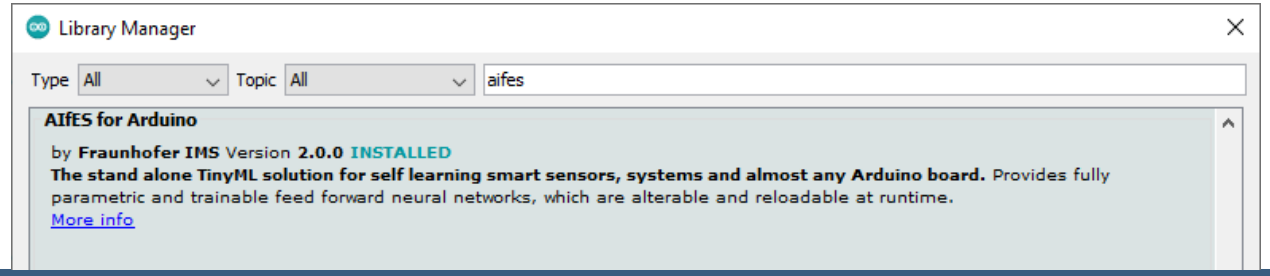
Extensions for AI (specialized AIfES support)

Functional safety (Lockstep, ECC etc.) incl. ISO 26262 ASIL-D ready certification

Crypto functions for information security

AIfES for Arduino

- The Fraunhofer IMS with AIfES and Arduino prepare to enter a partnership
- For this purpose, a version compatible with the Arduino IDE was realized
- It runs on almost any Arduino or Arduino compatible board
- You can easily install it via the Arduino Library Manager
- Published under the GNU GPL V3 license
- Free AIfES Tutorials and Projects (link)
- Also usable for the PC or other hardware
- AIfES for Arduino GitHub
- AIfES for Arduino library



AIfES licensing and partners

AIfES is offered as Dual License Model and is Open Source

- GNU GPL V3: Private or Free Open Source Software
- Paid license agreement: Commercial use (contact us)

Other partners

- Arduino
- Arm AI Ecosystem
- Open Roberta Lab

Paid license / Modules and extensions

Commercial use of the open source version possible

Changes to the code possible

Python-Modell-Wrapper für Keras und TensorFlow

Modules and extensions (closed source)

We also work together with other Fraunhofer Institutes on modules and extensions

- Federated learning
- Handwriting recognition
- Embedded human detection
- Complex gesture recognition
- Automated optimization of the network architecture

Function overview

Feedforward neural network inference

Float

Freely configurable (inputs, hidden layer, outputs)

Many activation functions

- Sigmoid, softsign, linear, RELU, Leaky RELU, softmax, tanh, ELU

Feedforward neural network Training

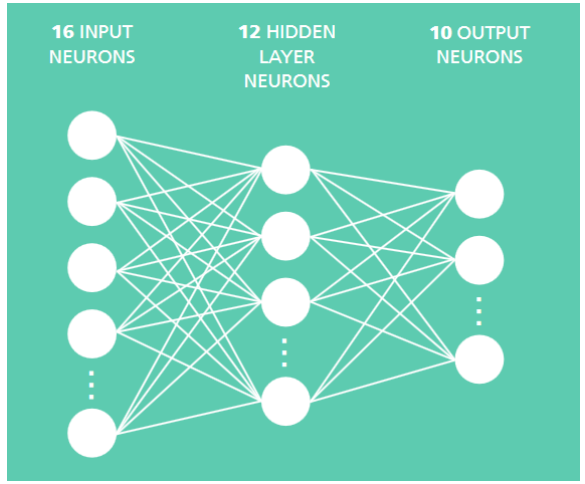
Full SGD and ADAM algorithm

Training types

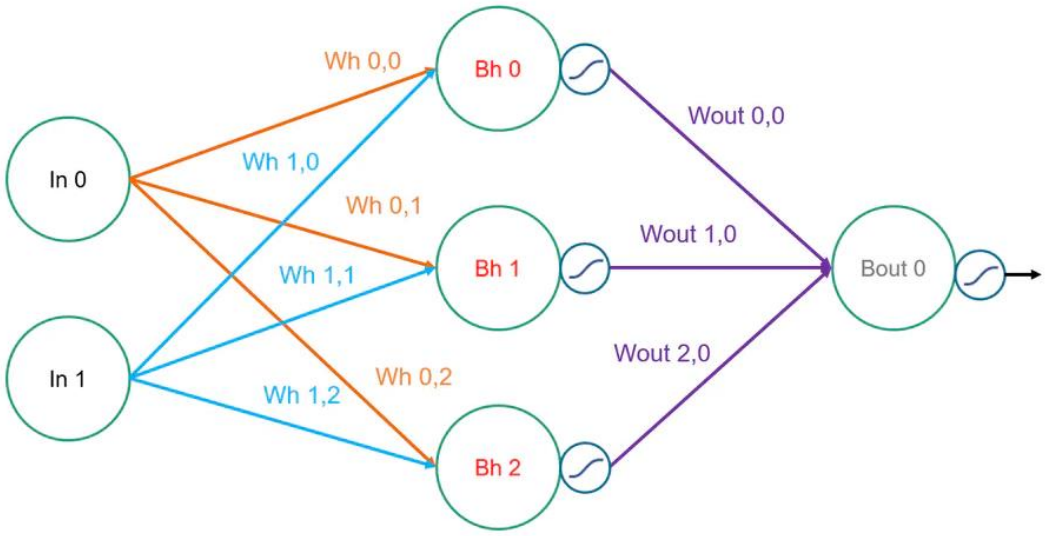
- Online, Batch, Minibatch

Various loss functions

- mean squared error (MSE), cross-entropy



How can I import a trained FNN?



- You can import a pre-trained FNN from other frameworks
- You need the trained weights and biases of the model
- The network structure can then be replicated in AIfES
- After the import the inference can be executed
- Even a further training is possible

Weights in AIfES

LayeredWeights

Hidden layer weights:

Wh 0,0 | Wh 0,1 | Wh 0,2 | Wh 1,0 | Wh 1,1 | Wh 1,2

Hidden layer bias weights:

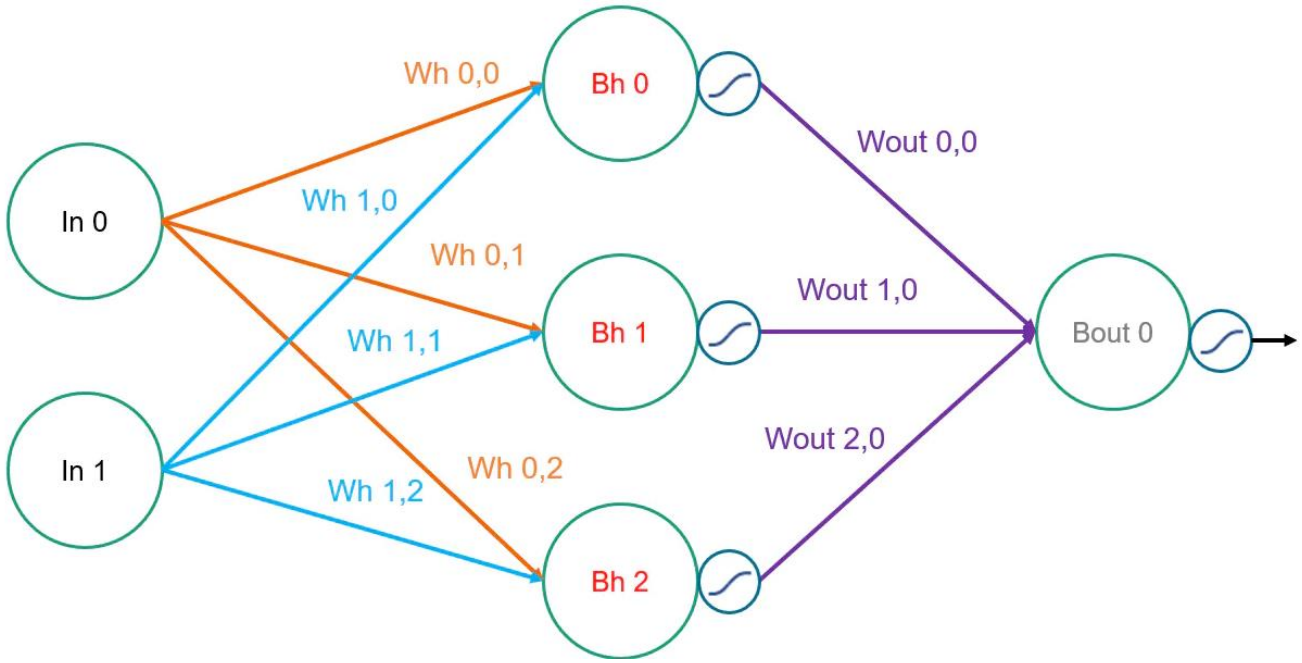
Bh 0 | Bh 1 | Bh 2

Output layer weights:

Wout 0,0 | Wout 1,0 | Wout 2,0

Output layer bias weight:

Bout 0



FlatWeights

Wh 0,0 | Wh 0,1 | Wh 0,2 | Wh 1,0 | Wh 1,1 | Wh 1,2 | Bh 0 | Bh 1 | Bh 2 | Wout 0,0 | Wout 1,0 | Wout 2,0 | Bout 0

What's next?

AIfES update in the next weeks

AIfES-Express API

- Simplified API that is directly integrated
- Inference and training with one function call

Fixpoint calculation with quantization of weights

- Automated Q7 quantization

Storage of weights in flash memory

Of course there are also new examples

Currently in development

CNN / ConvNets

Reinforcement learning

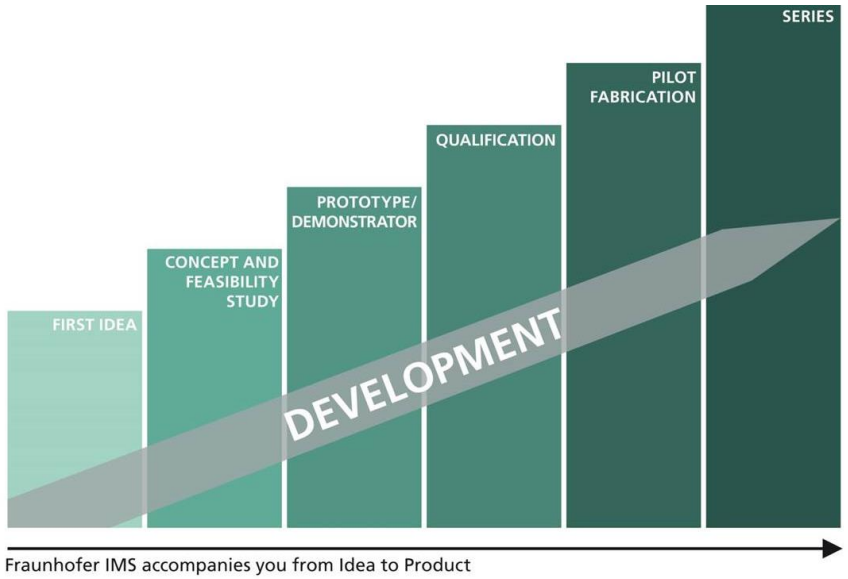
From the idea to the product - AI services at the IMS

Collaborative development of a product

- Concepts of how the use of AI can improve a product
- Development of AI models
- Integration of AI models into the customer's toolchain
- Upgrade of existing products with AI
- AI hardware and software codesign
- Verification and validation of AI

Consulting and training

- Customer consulting in the AI environment
- AIfES Workshops



TINY ML TALKS webcast

AIfES

Demonstrators and Projects

AIfES – Demonstrator: Handwriting recognition

Handwriting recognition - digits from 0-9

- Runs on an 8-bit microcontroller (Arduino UNO)
- Uses a standard capacitive PS/2 touchpad
- A special feature extraction was developed
- Very small ANN with only 12 neurons in a hidden-layer
- Recognition needs about 25ms (16 MHz clock frequency)
- Pre-trained on PC
- 10 persons were trained

[Video on YouTube]

AIfES – Demonstrator: Color recognition

Recognizes the colors of objects

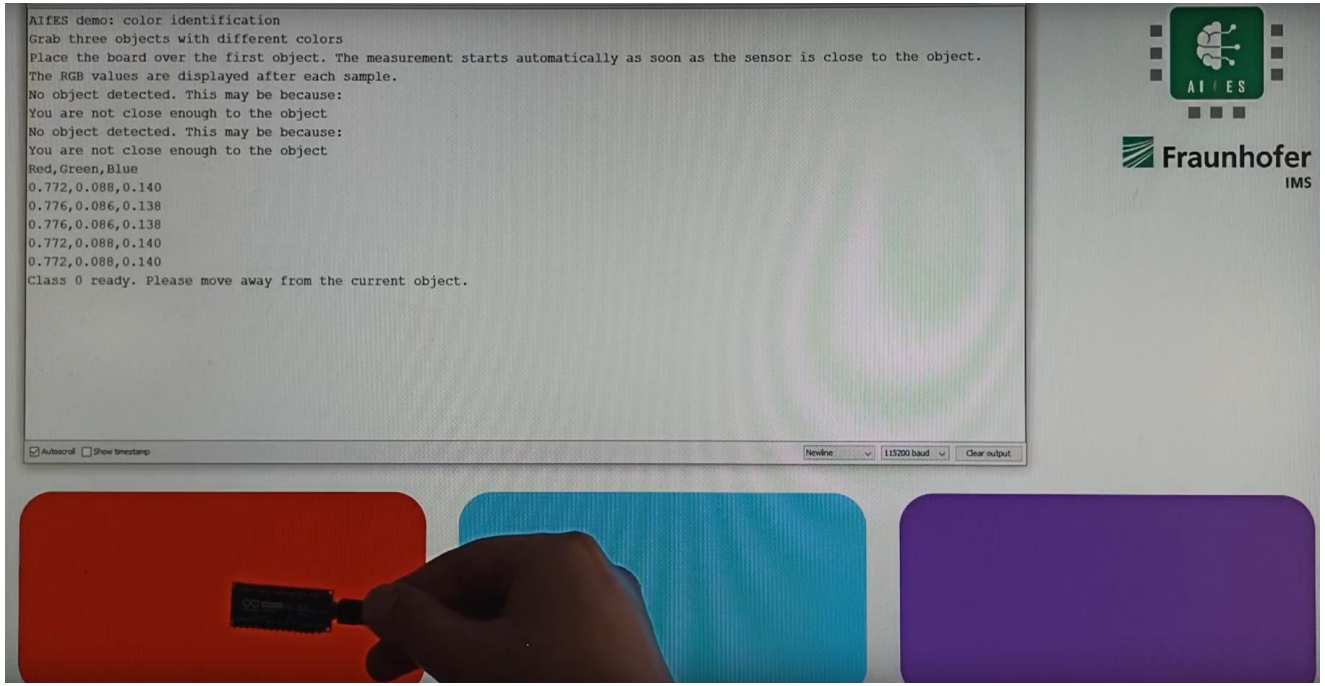
- Arduino Nano BLE Sense
- RGB Sensor
- 3 colors trainable in demo

Included in AIfES for Arduino

Open source available

Tutorial

Video on YouTube



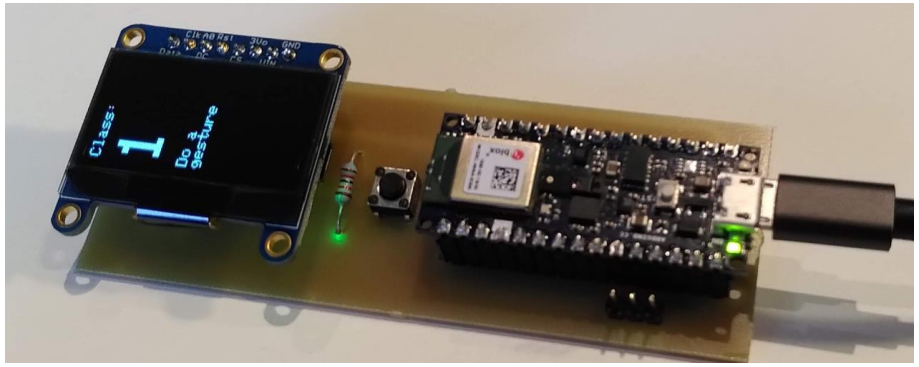
AIfES – Demonstrator: Gesture recognition with training

Recognizes complex gestures

- Recognizes figures written in the air → special feature extraction
- Can learn individual gestures directly in the system ADAM algorithm
- Can train up to 10 individual gestures (limited only by memory)
- Uses an accelerometer
- AIfES creates a KNN with the appropriate structure and trains it
- Three repetitions per gesture are sufficient
- Recognition takes about 20 - 100 ms (Cortex M4)
- Training three gestures takes less than 2 seconds

[YouTube video 1]

[YouTube video 2]



Project noKat: Embedded human recognition

Current ZIM - Project (Central Innovation Program for medium-sized businesses)

- IMS and company van Rickelen GmbH & Co. KG

Low-power and low-cost camera system (RGB)

- Camera remains stationary

Recognition of people in moving images

- Other classes (cars, bicycles, animals, etc.)

Specialized feature extraction and a very small ANN (artificial neural network)

- No ConvNet (convolutional neural network)

Reduction of the required parameters by more than 99%

- EfficientDet-D7 (77 million parameters) / noKat (1125 parameters)

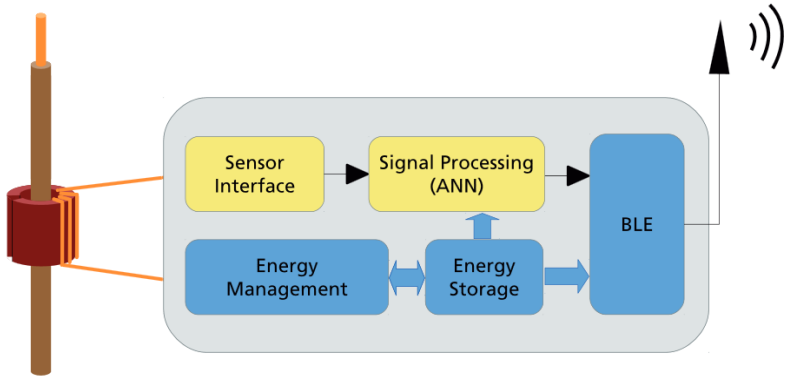
Processing time on a microcontroller (160 MHz) approx. 120ms

AIfES – Demonstrator: Current sensor

Wireless current sensor for condition monitoring

- Wireless and energy self-sufficient operation for easy retrofitting
- Learns the states of a device based on its power consumption
- Learning algorithm on the microcontroller (ATMega32U4)
- Configuration via BLE
- Sends only the device status via BLE
- No measured values have to be transmitted

Read more



AIfES – Demonstrator: AI based LIDAR

Real-time sensor signal preprocessing in a LIDAR camera

The use of AI should improve performance in high ambient light

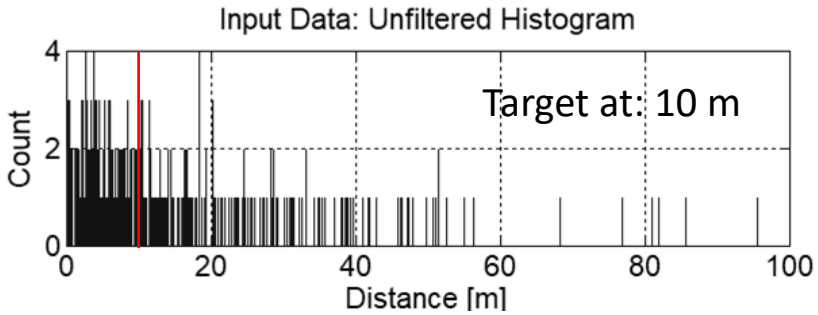
The ANN evaluates the histogram of the time correlated single photon count

- The histogram consists of **95%** noise and only **5%** target information
- Dynamic background noise due to ambient light disturbs the measurement

The ANN calculates the distance of the object and replaces all filters

Improvement over the classical method in increased ambient light

- Improvement of the accuracy by about 20%.



Contact

Dr. Pierre Gembaczka

Program Manager: Industrial AI & Product Manager: AIfES

Fraunhofer Institute for Microelectronic Circuits and Systems IMS

Finkenstraße 61, 47057 Duisburg
Phone +49 203 3783-220
Email pierre.gembaczka@ims.fraunhofer.de

AIfES - Artificial Intelligence for Embedded Systems
www.aifes.ai
aifes@ims.fraunhofer.de

AIfES for Arduino on GitHub

Follow us on LinkedIn

Copyright Notice

www.tinyml.org

Copyright Notice

www.tinyML.org