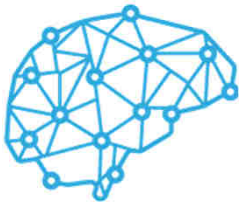


Artificial Neural Network Mapping Made Simple with the STM32Cube.AI

Markus Mayr
Product Marketing Manager, MCU

life.augmented

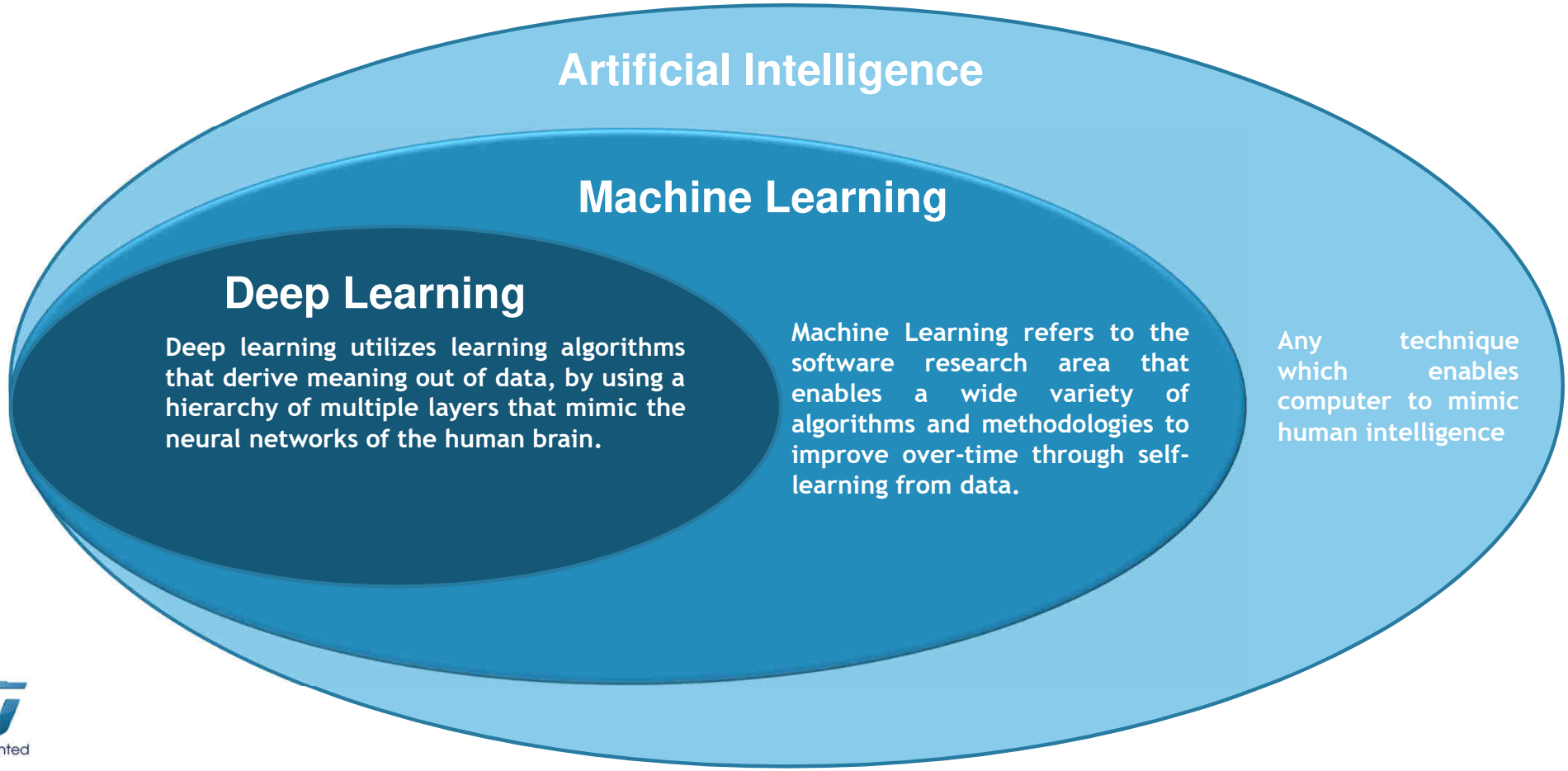
Technology Tour 2019
Schaumburg, IL | April 25

Artificial Intelligence (AI)

- AI is a superset of all the studies where machines mimic cognitive capabilities like humans. For example:
  - Interaction with the environment
  - Knowledge representation
  - Perception
  - Learning
  - Computer vision
  - Speech recognition
  - Problem solving and many more.
- Main ingredients
  - Computer science
  - Statistics
  - Mathematics

life.augmented

AI vs Machine Learning vs Deep Learning



life.augmented

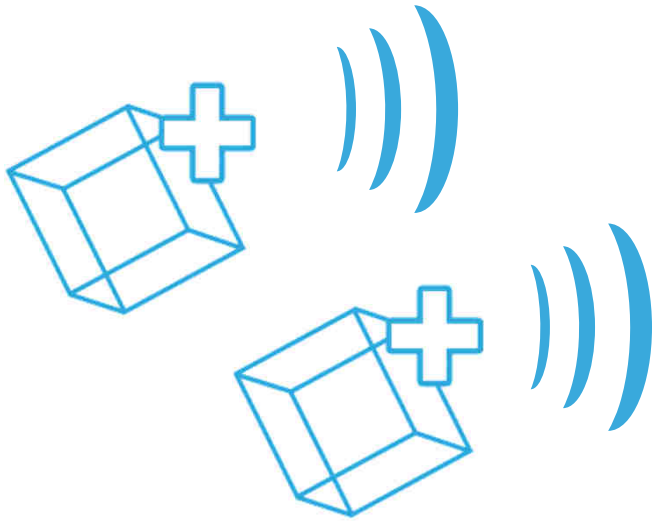
Why Deep Learning is So Important

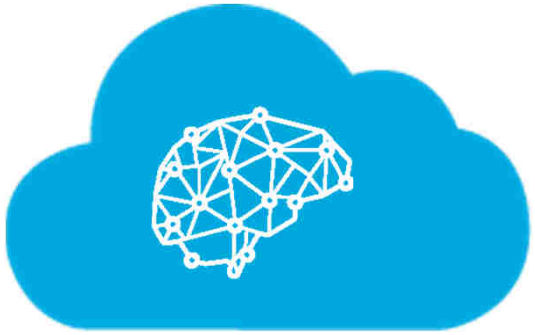
- Convolutional Deep Neural Networks outperform previous methods on a number of tasks:

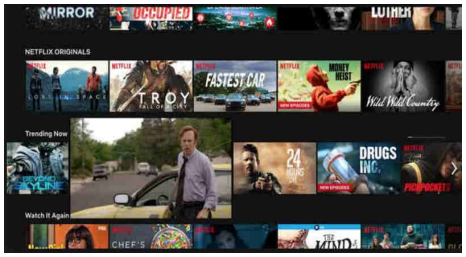
Problem	Dataset	Best Accuracy w/o CNN	Best Accuracy with CNN	Diff
Object classification	ILSVRC	**73.8%**	**95.1%**	**+21.3%**
Scene classification	SUN	**37.5%**	**56%**	**+18.5%**
Object detection	VOC 2007	**34.3%**	**60.9%**	**+26.6%**
Fine-grained class	200Birds	**61.8%**	**75.7%**	**+13.9%**
Attribute detection	H3D	**69.1%**	**74.6%**	+5.5%
Face recognition	LFW	**96.3%**	**99.77%**	**+3.47%**
Instance retrieval	UKB	**89.3%** (CDVS: 85.7%)	**96.3%**	+7.0%

May 2015

Traditional AI Processing in the Cloud



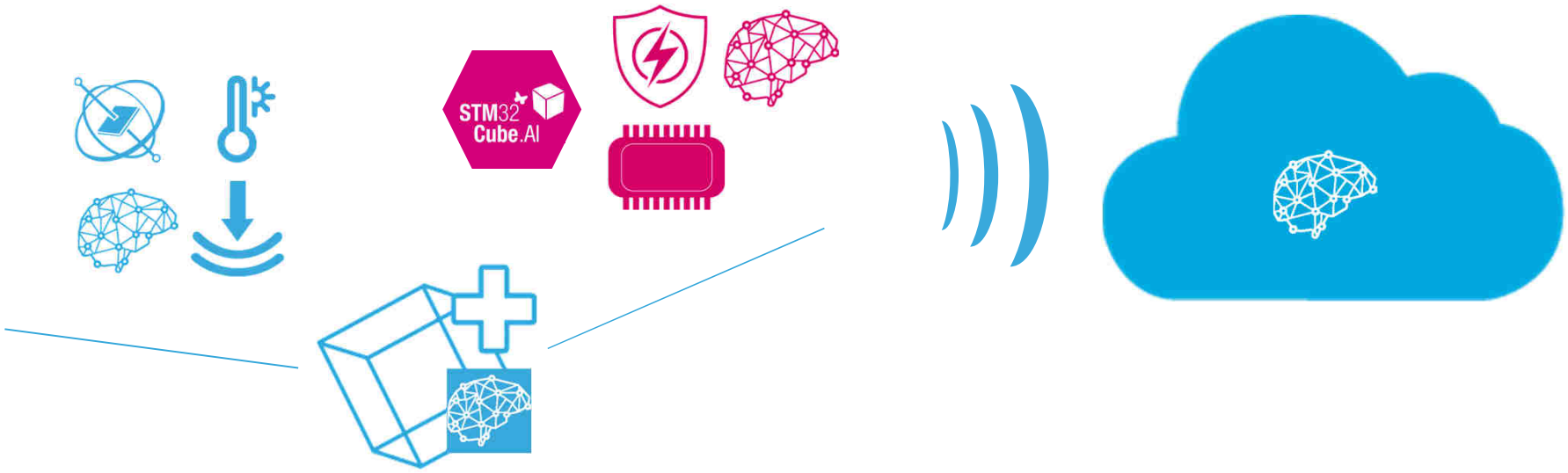




Analysis of vast amount of information

life.augmented

AI on the Edge



life.augmented

Artificial Intelligence and STM32

Application trend

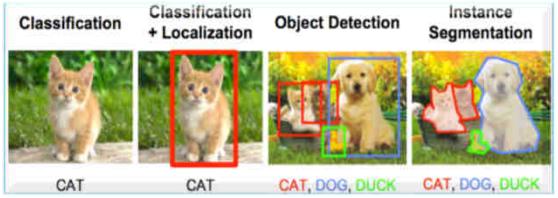
Sensors

- Activity Recognition with Inertials (DCNN, ESN, LSTM)
- Stress Analysis or Attention Analysis (DCNN, SON), etc

Audio Processing

- Speech Recognition (DeepSpeech,Wave2Letter)
- Speech Syntesys (WaveNet, Tacotron)

Video Analysis



- Classification (Alexnet, Inception, VGG)
- Detection (Yolo, SSD)

STM32 → **Dedicated AI hardware needed**

STM32 CubeMX.AI

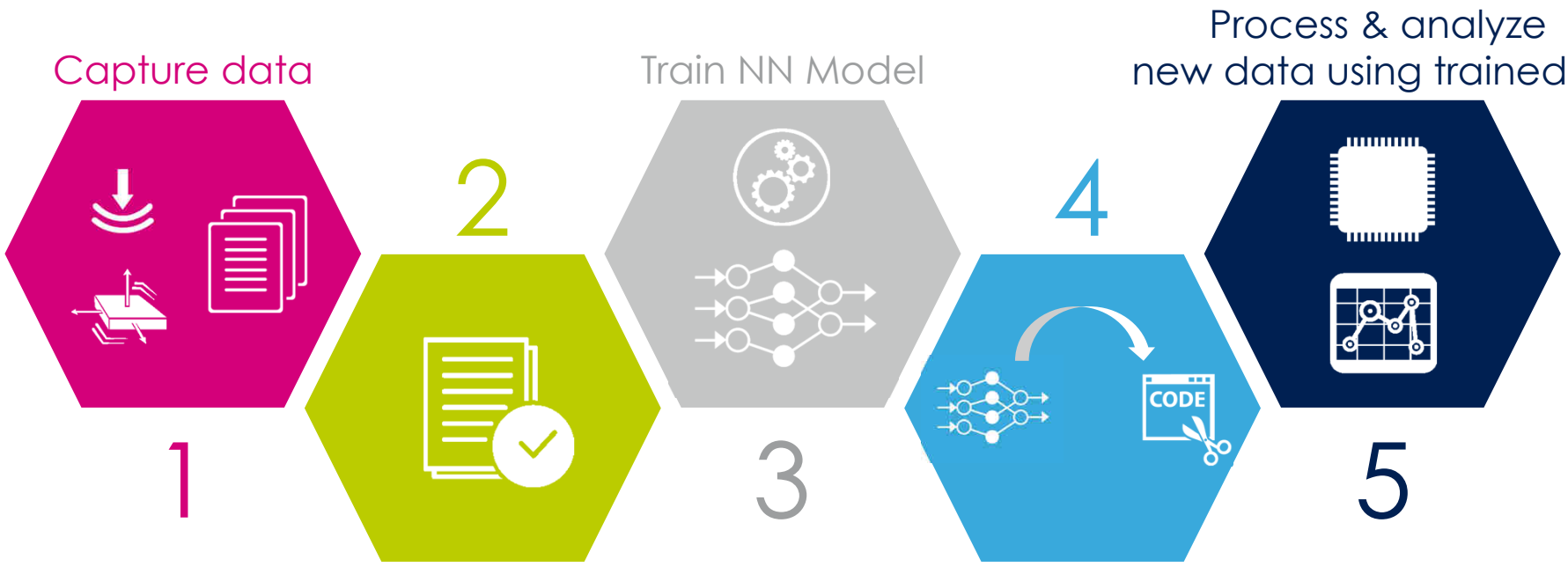
- Audio use cases with individual commands
- Classic motion sensor use cases

- Video analysis cannot be done in timely manner with MCU
- Advanced Audio use cases with Natural language understanding not yet accessible for MCUs

life.augmented

The Key Steps Behind Neural Networks

Neural Network (NN) Model Creation | Operating Mode



life.augmented

ST Toolbox for Neural Networks

life.augmented

Capture data

Process & analyze new data using trained NN

STM32

STM32 Cube.AI

Clean, label Data
Build NN topology

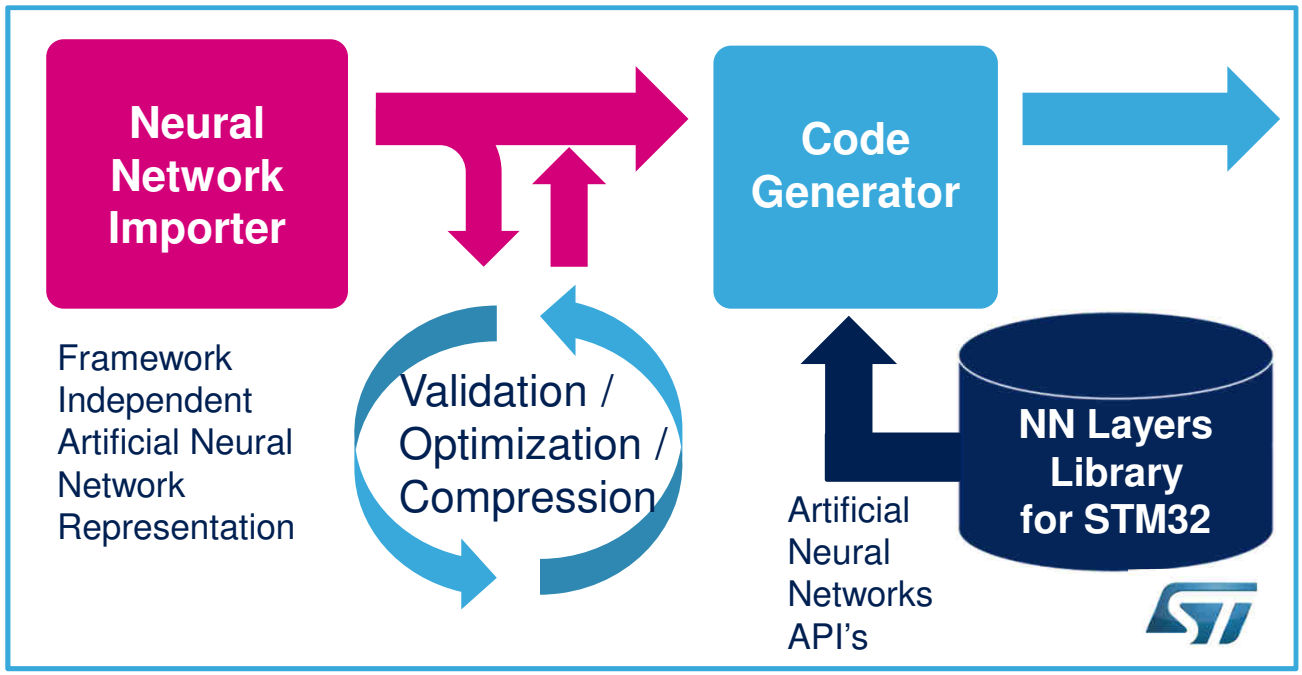
Convert NN into optimized code for MCU

life.augmented

X-Cube-AI : Architecture

Off-the-shelf :
Pre-trained Artificial Neural Network Model

Deep Learning Framework dependent

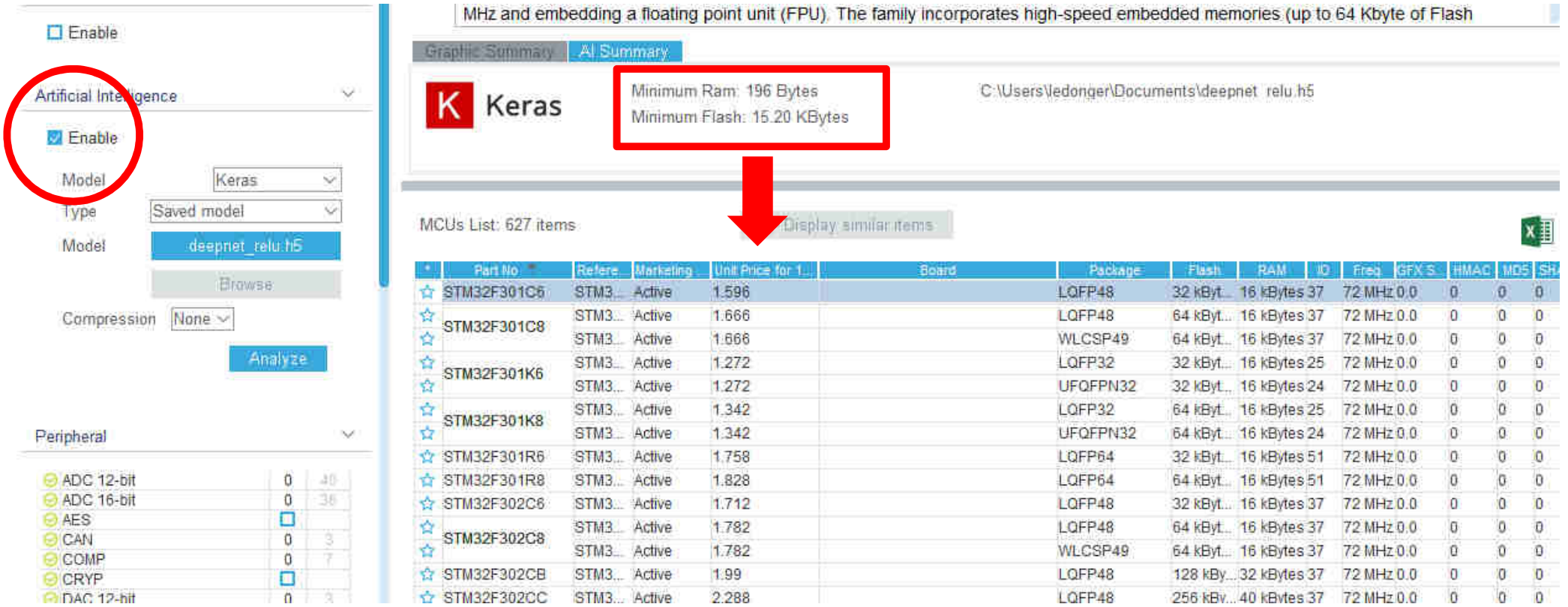


Embedded Solution
Optimized Artificial Neural Network Code generated for STM32

This optimized STM32 Artificial neural network model can be included into the user project (using KEIL, IAR, OpenSTM32) and can be compiled and ported onto the final device for field trials

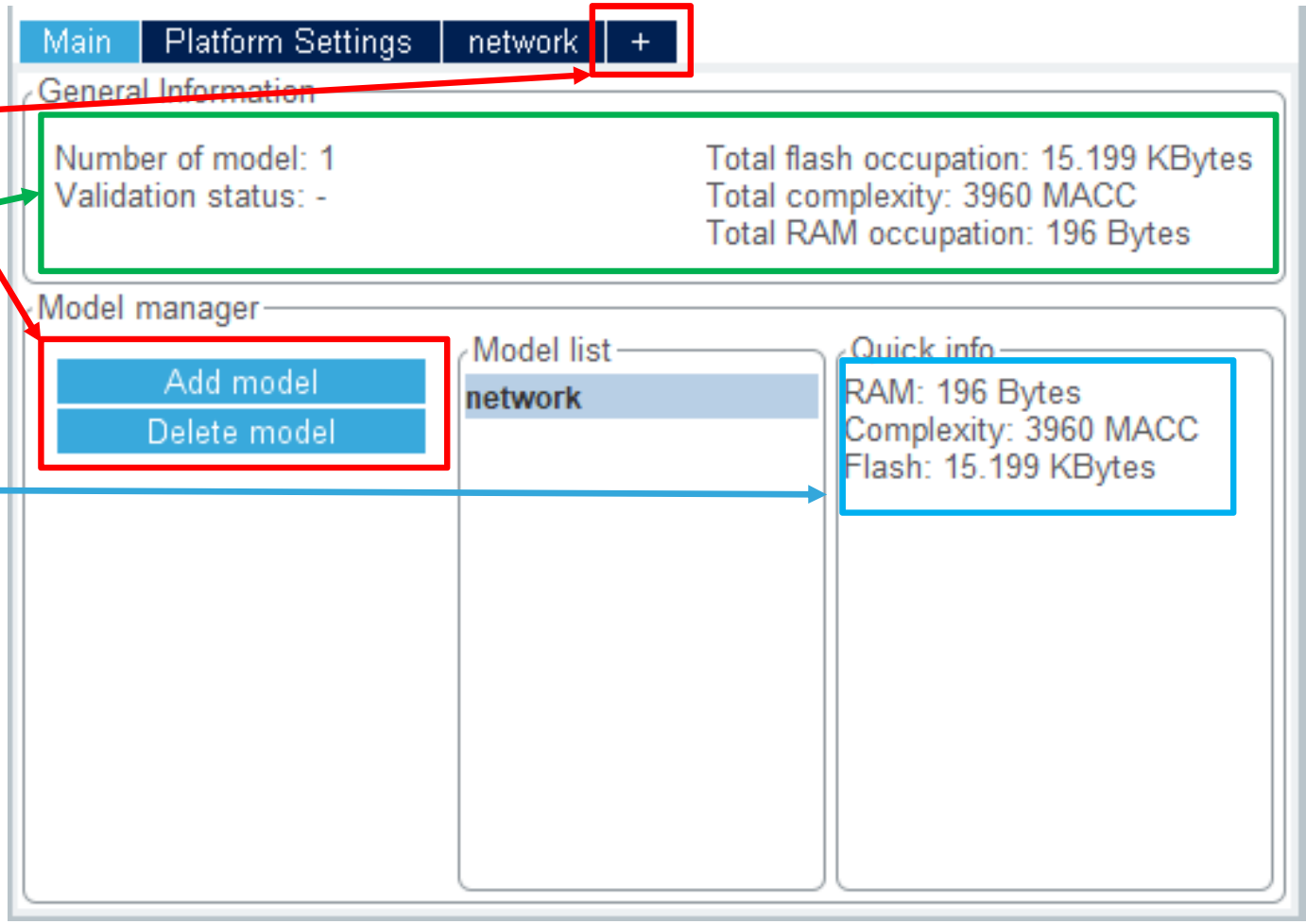
life.augmented

MCU Finder AI Filter



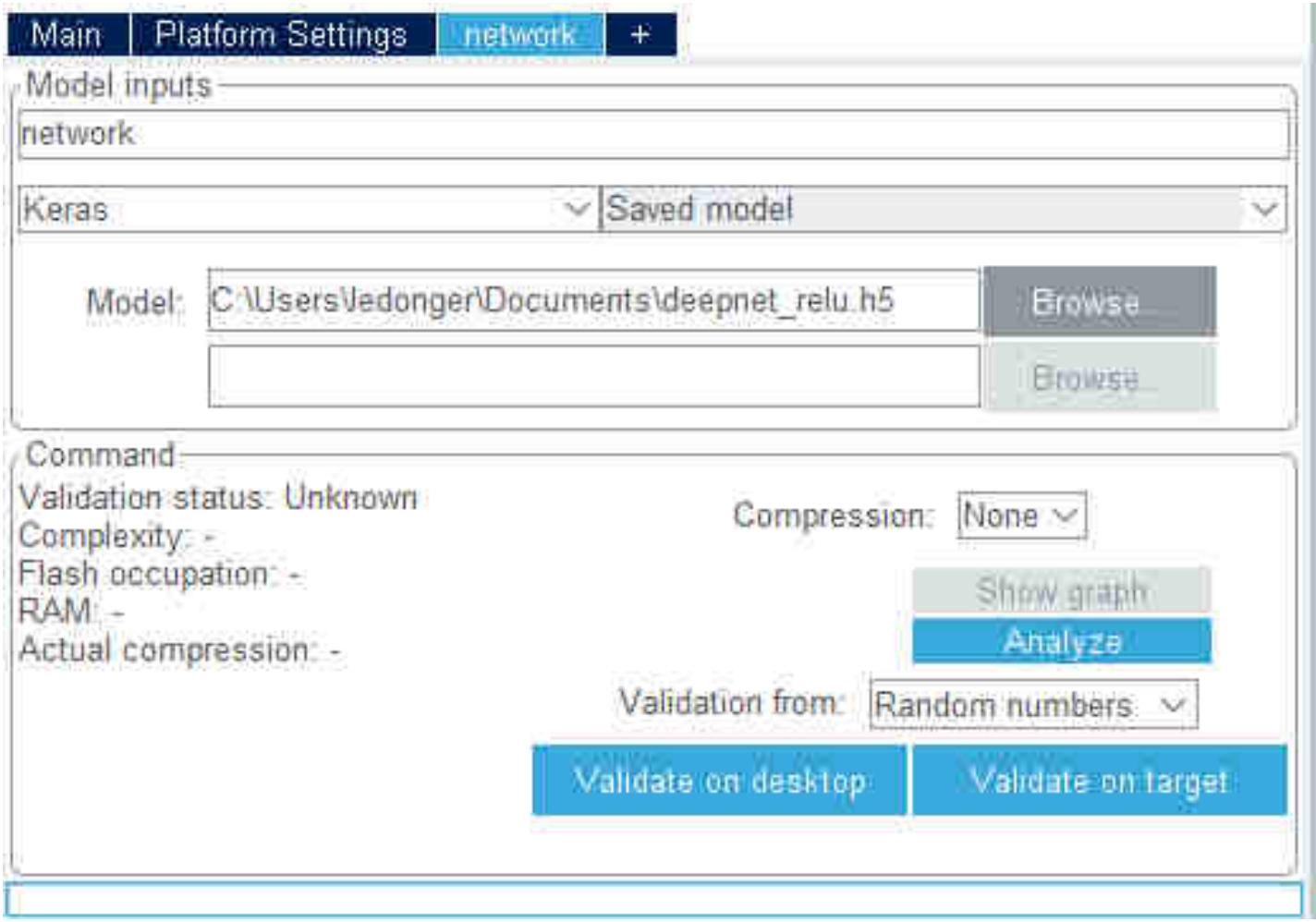
X-Cube-AI Main tab

- Add/Delete models
- Get general information
- Have a quick look on different models



X-Cube-AI Detailed View

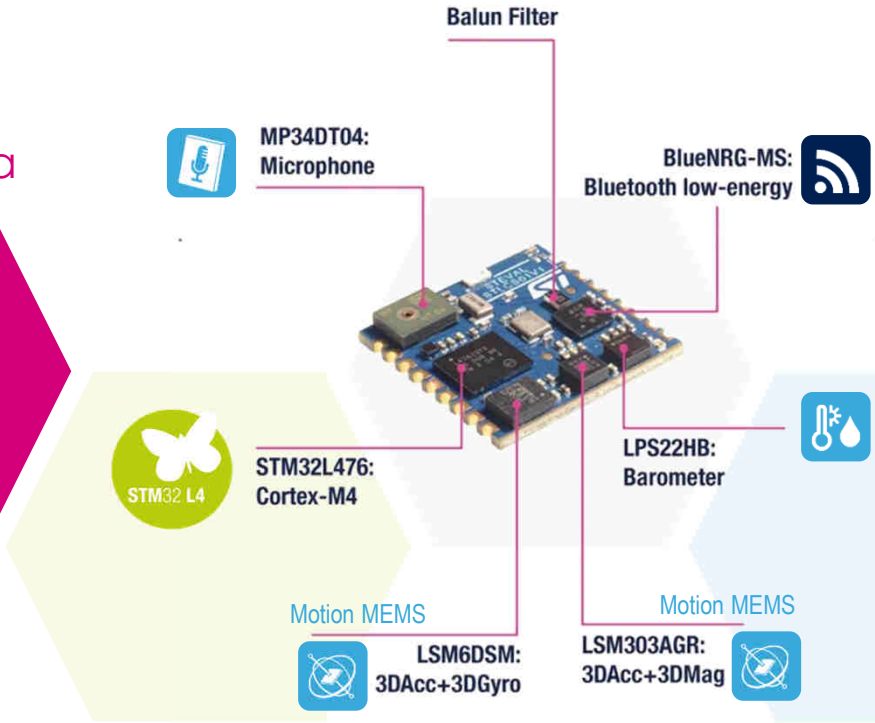
- Perform **analysis** to compute the model size, get an image of the network and the complexity
- Perform **validation on desktop**
- Perform **validation on target**
- Set a **compression** to reduce the model size (By reducing the accuracy of the model)



Form Factor Hardware
to Capture and Process Data

Capture data

Process & analyze new data using trained NN



life.augmented

www.st.com/SensorTile
www.st.com/SensorTile-edu

Form Factor Hardware

AI IoT Node for More Connectivity (Q1 2019)

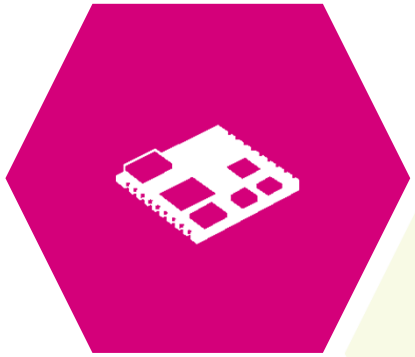
Sub-1GHz

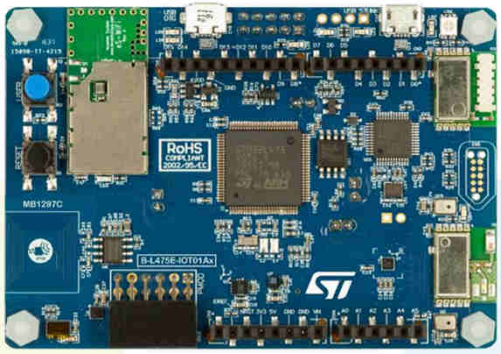
Dynamic NFC Tag

\+

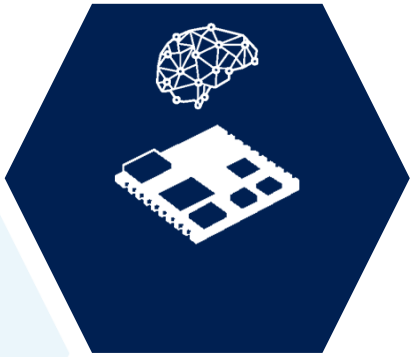
Wi-Fi

Capture data





Process & analyze new data using trained NN



More debug capabilities

- Integrated ST-Link/V2.1
- PMOD extension connector
- Arduino Uno extension connectors

life.augmented

https://www.st.com/iotnode

Collecting Data
& Architecting a NN Topology

Services provided by Partners

Capture data

Clean, label Data
Build NN topology

ST tools to support

ST BLE Sensor

ST BLE Sensor mobile phone application

Collect and label data from the SensorTile.

Selected partners

Neural Networks engineering services support.

Data scientists and Neural network architects.

www.st.com/STM32CubeAI#Partners?

life.augmented

STM32CubeMX Extension

AI Conversion Tool

Input your framework-dependent, pre-trained Neural Network into the **STM**32**Cube**.AI conversion tool

Automatic and fast generation of an STM32-optimized library

STM32**Cube**.AI offers interoperability with state-of-the-art Deep Learning design frameworks

Train NN Model

* TensorFlow used as a Keras backend. Not all operators accessible to MCUs

Convert NN into optimized code for MCU

Process & analyze new data using trained NN

life.augmented

STM32Cube.AI Roadmap

2019

- Floating Point Support

Keras
ConvNetJS
Deep Learning in your browser
Lasagne
Caffe

Apr

IoTNode

Jun

- Quantization
- TensorFlow for MCU
- Command line interface
- UI Improvements
- Additional layers

+ TensorFlow Lite microcontrollers

2020

- ONNX Introduction
- Additional layers
- Debug improvements

Keras + TensorFlow + ONNX + mxnet + Microsoft CNTK + Chainer + PyTorch

life.augmented

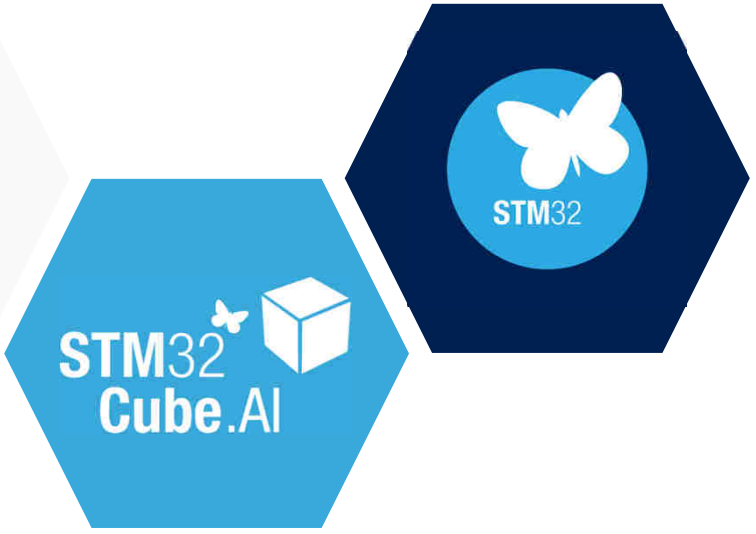
ST Toolbox for Neural Networks

More Than Just a Conversion Tool

Process & analyze
new data using trained NN

- Function packs for **quick prototyping**
- **Audio** and **motion** examples

- STM32 **Community** for **support** and **idea** exchange
- **Dedicated** topic for Neural Networks

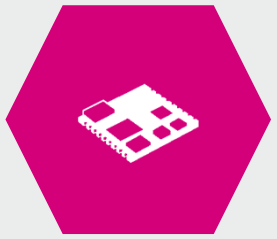


Convert NN into
optimized code for MCU

life.augmented

Human Activity Recognition (HAR)

Motion Example in FP-AI-SENSING1 Package



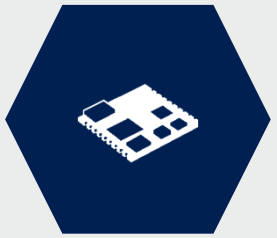
5 classes

Embedded **motion**

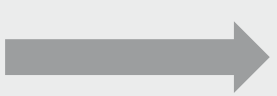
Labelling controlled by smartphone application

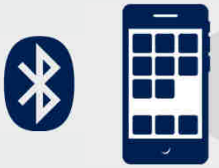
Data stored on the device SD card for future **learning**

Stationary, walking, running, biking, driving



NN & example dataset provided





Embedded **motion** pre-processing

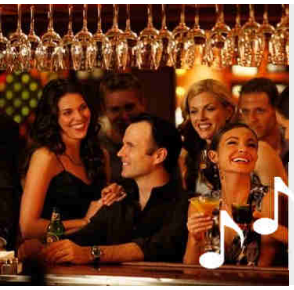
Inferences running on the microcontroller

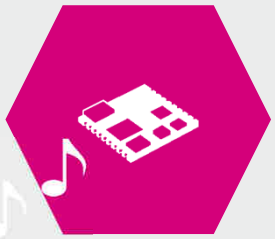
Inference result displayed on mobile app

life.augmented

Audio Scene Classification (ASC)

Audio Example in FP-AI-SENSING1 Package





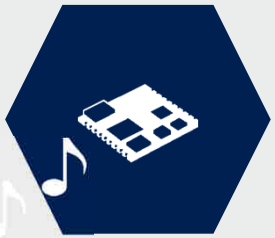
3 classes

Embedded **audio**

Labelling controlled
by smartphone application

Data stored on the device
SD card for future **learning**

Indoor, Outdoor, In vehicle



NN & example
dataset provided

Embedded **audio**
pre-processing

Inferences running
on the microcontroller

Inference result
displayed on mobile app

life.augmented

STM32 Solutions for AI

More Than Just the STM32Cube.AI

An extensive toolbox to support easy creation of your AI application

AI extension for STM32CubeMX

To map pre-trained Neural Networks onto the STM32

STM32 Cube.AI

Function packs for Quick prototyping

Audio and motion examples

SensorTile reference hardware

To run inferences or data collection

… And more coming!

STM32 Community with dedicated Neural Networks topic

Mobile phone application

To collect and label data

To display the result of inference processing on the STM32

ST Partner Program with a dedicated group of Partners providing Neural Networks engineering services

Data scientists and Neural network architects

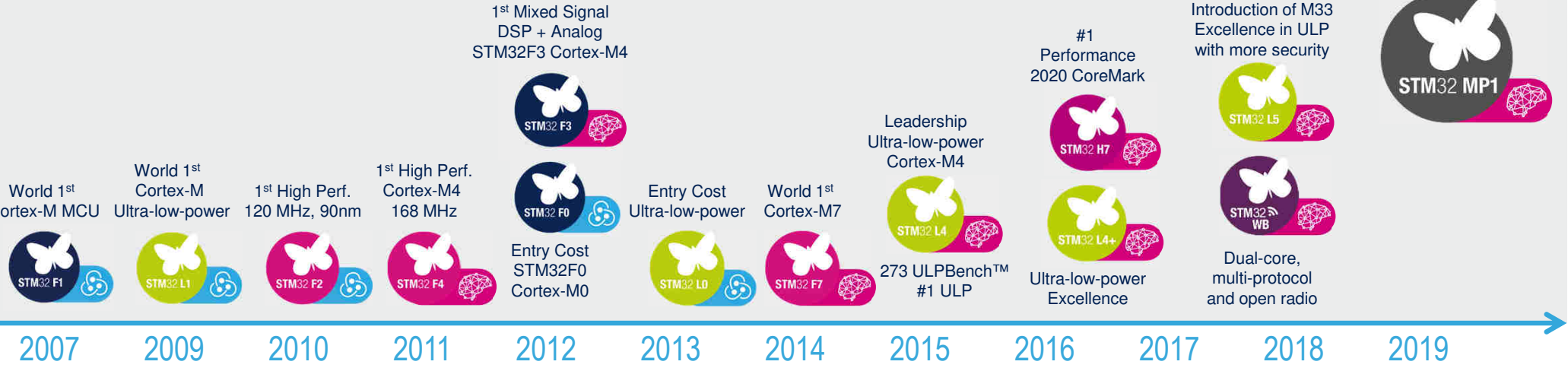
life.augmented

Making AI Accessible Now

Leader in Arm® Cortex®-M 32-bit General Purpose MCU

Compatible with STM32Cube.AI ecosystem

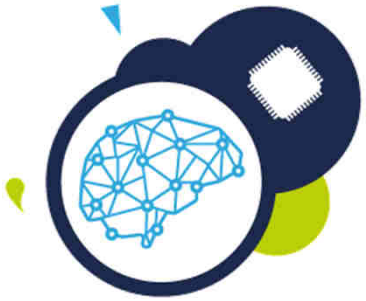
Compatible with Partner Machine Learning ecosystems



life.augmented

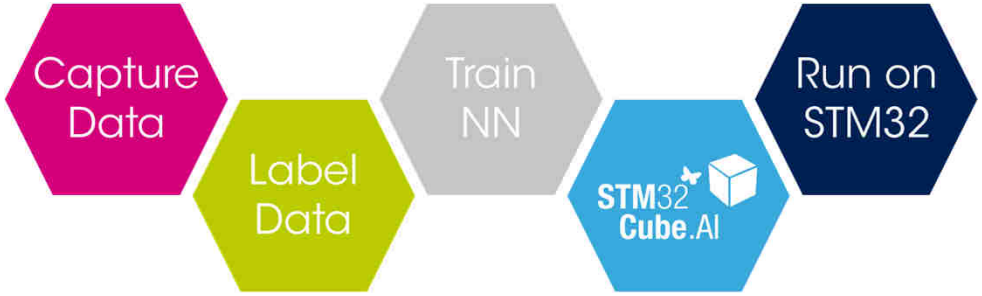
More than 40,000 customers

Over 4 Billion STM32 shipped since 2007



For More Information

www.st.com/STM32CubeAI



life.augmented