Neural Networks on STM32
Artificial Intelligence Solutions

life.augmented
Neural Networks on STM32
Simple, fast, optimized

STM32 Cube.AI
The Key Steps Behind Neural Networks

1. Capture data
   - Clean, label data
   - Build NN topology

2. Train NN Model

3. Convert NN into optimized code for MCU

4. Process & analyze new data using trained NN
ST Toolbox for Neural Networks

- Capture data
- Clean, label data
- Build NN topology
- Convert NN into optimized code for MCU
- Process & analyze new data using trained NN
Input your framework-dependent, pre-trained Neural Network into the **STM32Cube.AI** conversion tool

Automatic and fast generation of an STM32-optimized library

**STM32Cube.AI** offers interoperability with state-of-the-art Deep Learning design frameworks

Train NN Model

Convert NN into optimized code for MCU

Process & analyze new data using trained NN

**STM32Cube.AI** Extension

**AI Conversion Tool**
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 mobile phone application
Collect and label data from the SensorTile.

Selected partners
Neural Networks engineering services support.
Data scientists and Neural network architects.
ST Toolbox for Neural Networks
More Than Just a Conversion Tool

- Function packs for **quick prototyping**
- **Audio, Motion** and **Vision** examples

- **STM32 Community** with dedicated Neural Networks **topic**
- For **support** and **idea** exchange

Convert NN into optimized code for MCU

Process & analyze new data using trained NN
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

- **Software examples** for Quick prototyping
  - Audio, Motion and Vision Function packs
  - On ST development Hardware

- **STM32 Community** with dedicated Neural Networks topic

- Trainings, hands on, MOOCs and partners videos

- STM32 AI Partner Program with dedicated Partners providing
  - Machine or Deep Learning engineering services
Example Form Factor Hardware to Capture and Process Data

SensorTile

Capture data

Process & analyze new data using trained NN

www.st.com/SensorTile
www.st.com/SensorTile-edu
Fast Go to Market Module to Capture Data with More Accuracy

SensorTile.Box

- More advanced, high accuracy and low power sensors
  - First Inertial module with Machine Learning capabilities.
  - Motion (accelerometer and gyroscope, magnetometer) and slow motion (inclinometer)
  - Altitude (pressure), environment (pressure, temperature, humidity, compass) and sound (sound and ultrasound analog microphone)
  - Microsoft IoT services ready to make available on a web dashboard the result of the embedded processing

www.st.com/SensorTileBox
Distributed AI: Sensor + STM32
Optimize Performance and Power Consumption

Smart Sensor
with Machine Learning Core

- Best ultra-low-power sensing at high performance:
  - 550µA (gyroscope and accelerometer)
    ➔ 200µA less than closest competitor
  - 20~40µA (Accelerometer only for HAR)
- Efficient Finite State Machines: 2µA
- Configurable Machine Learning Core: 4~8µA

Smart STM32
second level of AI processing

- More advanced and complex NNs
- Decisions on multiple sensors
- NN input can be sensor data and/or sensor Machine Learning decisions
- Multiple Neural Networks support
- Actuation & communication

Inertial Sensor
New LSM6DSOX

FSM up to 16
MLC up to 8
Raw Data
Event Decision
FSM and MLC Re-configuration

Deep Learning
Neural Networks
Machine Learning
Form Factor Hardware
AI IoT Node for More Connectivity

IoTNode

Capture data

Sub-1GHz Sub-1GHz
NFC Dynamic NFC Tag

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

Process & analyze new data using trained NN

https://www.st.com/IoTnode
STM32Cube.AI
Tool Roadmap
### STM32Cube.AI Roadmap

#### 2019

- **Floating Point Support**

#### Jul

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

#### Oct

- **Integer Quantization**
- **External memory support (weights & activations)**
- **Lambda layers support (Alpha customers only)**

#### 2020

- **ONNX support introduction**
- **Additional layers**
- **Debug improvements**

---

**NN Training Tools Supported**
- Lasagne
- Caffe
- ConvNetJS
- TensorFlow
- TensorFlow Lite

**Software Examples**
- SensorTile
- IoTNode
- SensorTile.Box
- STM32H747 Discovery Kit

**FP-AI-Sensing1 Support**
- SensorTile
- IoTNode
- SensorTile.Box

**FP-AI-Vision1 Support**
- SensorTile
- IoTNode
- SensorTile.Box

**FP-AI-Sensing1 Support**

**FP-AI-Vision1 Support**

**OpenMV H7 CAM**

NN libraries generated by STM32Cube.AI compatible OpenMV MicroPython ecosystem
Function Packs
(Software examples)
Audio Scene Classification (ASC)
Audio Example in FP-AI-SENSING1 Package

Audio Data capture

Labelling controlled by smartphone application

Data stored on the device SD card for future learning

Indoor, Outdoor, In vehicle labelling

Embedded audio pre-processing

Inferences running on the microcontroller

NN & example dataset provided

Inference result displayed on mobile app
Human Activity Recognition (HAR)
Motion Example in FP-AI-SENSING1 Package

Motion Data Capture
Labelling controlled by smartphone application
Data stored on the device SD card for future learning
Stationary, walking, running, biking, driving labelling

Embedded motion pre-processing
Inferences running on the microcontroller
Inference result displayed on mobile app

NN & example dataset provided
Enjoy the food classification demo
- Default demo based on 18 classes (224x224 RGB pictures)
- Several camera image output size possible

Full end-to-end optimized software example
- from camera acquisition to image pre-processing before feeding the NN
- Multiple memory mapping possibilities to optimise and test impact on performances
- Retrain this NN with your own dataset
- Quantize your trained network to optimized inference time and memory usage

Embedded image pre-processing (SW) on the STM32H747

Inferences running on the microcontroller

NN & example dataset provided

Inference result displayed on STM32H747
Discovery board LCD display

Pizza
99%
150ms
Making AI Accessible Now

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

Compatible with Deep Learning
STM32Cube.AI ecosystem

Compatible with Machine Learning
Partner ecosystems

World 1st Cortex-M MCU
World 1st Cortex-M Ultra-low-power
1st High Perf. 120 MHz, 90nm
1st High Perf. Cortex-M4 168 MHz
1st Mixed Signal DSP + Analog Cortex-M4
Entry Cost Cortex-M0
Entry Cost Ultra-low-power
World 1st Cortex-M7
Leadership Ultra-low-power Cortex-M4
447 ULPBench™ #1 ULP
#1 Performance 2400 CoreMark
Efficiency at its best!

First STM32 MPU
Dual Cortex-A7 + Cortex-M4
STM32 meets Linux

STM32Cube.AI ecosystem
Compatible with Deep Learning
Compatible with Machine Learning

Introduction of Cortex-M33
Excellence in ULP with more security

Mainstream Cortex-M0+ MCUs
Efficiency at its best!

STM32L4
STM32L1
STM32F1
STM32F2
STM32F4
STM32F7
STM32L6
STM32L8
STM32WB
STM32L5
STM32MP1


More than 40,000 customers
Over 4 Billion STM32 shipped since 2007

Life augmented
www.st.com/STM32CubeAI