Neural Networks on STM32
Artificial Intelligence Solutions
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

Neural Network (NN) Model Creation

Operating Mode
ST Toolbox for Neural Networks

Capture data

Clean, label data
Build NN topology

Process & analyze new data using trained NN

Convert NN into optimized code for MCU
Form Factor Hardware to Capture and Process Data

SensorTile

Capture data

Process & analyze new data using trained NN

Capture data using STM32L4

STM32L476: Cortex-M4

MP34DT04: Microphone

Balun Filter

BlueNRG-MS: Bluetooth low-energy

LPS22HB: Barometer

Motion MEMS

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

SensorTile.Box

Capture data

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
Form Factor Hardware
AI IoT Node for More Connectivity

IoTNode

Capture data

+ Sub-1GHz
Wi-Fi
Dynamic NFC Tag

Process & analyze new data using trained NN

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

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 mobile phone application
Collect and label data from the SensorTile.

Selected partners
Neural Networks engineering services support. Data scientists and Neural network architects.
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

Process & analyze new data using trained NN

Convert NN into optimized code for MCU

* TensorFlow used as a Keras backend. Not all operators accessible to MCUs
### STM32Cube.AI Roadmap

<table>
<thead>
<tr>
<th>2019</th>
<th>Apr</th>
<th>Jun</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Floating Point Support</td>
<td>IoTNode</td>
<td>SensorTile.Box</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2020</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Quantization</td>
</tr>
<tr>
<td>• TensorFlow for MCU</td>
</tr>
<tr>
<td>• Command line interface</td>
</tr>
<tr>
<td>• UI Improvements</td>
</tr>
<tr>
<td>• Additional layers</td>
</tr>
</tbody>
</table>

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

---

**Tools:**
- ConvNetJS
- Caffe
- TensorFlow Lite
- TensorFlow
- ONNX
- CNTK
- Chainer
- PyTorch

**Projects:**
- Keras
- Lasagne
- IoTNode
- SensorTile.Box

**Platforms:**
- STM32
- Microcontrollers

**Images:**
- IoT devices and applications
- STM32 development board
- Industrial automation and robotics

---

**STM32Life:**
- Life augmented by AI
- Deep learning in your browser

---
ST Toolbox for Neural Networks
More Than Just a Conversion Tool

- Function packs for quick prototyping
- Audio and Motion examples
- STM32 Community with dedicated Neural Networks topic
- For support and ideas exchange

Process & analyze new data using trained NN

Convert NN into optimized code for MCU
Audio Scene Classification (ASC)
Audio Example in FP-AI-SENSING1 Package

- **Embedded audio**
- **Labelling** controlled by smartphone application
- **Data stored on the device**
  - **SD card for future learning**
- **3 classes**
  - Indoor, Outdoor, In vehicle

- **Embedded audio pre-processing**
- **Inferences** running on the microcontroller
- **Inference result** displayed on mobile app
Human Activity Recognition (HAR)
Motion Example in FP-AI-SENSING1 Package

- **Embedded motion**
- **Labelling** controlled by smartphone application
- Data stored on the device SD card for future learning
- **5 classes**
  - Stationary, walking, running, biking, driving

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

- **Capture Data**
- **Train NN**
- **Run on STM32**

- **STM32 Data AI**

- **NN & example dataset provided**

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

- **Labelling** controlled by smartphone application

- **Inferences** running on the microcontroller

- **Inference result** displayed on mobile app
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

Function packs for Quick prototyping
Audio and motion examples

Reference hardware
To run inferences or data collection

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

… And more coming!
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
447 ULPBench™ #1 ULP
Leadership Ultra-low-power Cortex-M4

#1 Performance 2400 CoreMark

Mainstream Cortex-M0+ MCUs
Efficiency at its best!

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

Introduction of Cortex-M33
Excellence in ULP
with more security


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

www.st.com/STM32CubeAI