



**INDUSTRIAL
SUMMIT 2023**
SHENZHEN, CHINA | 28 SEPTEMBER



STM32Cube.AI STM32Cube.AI Develop Cloud Boost Your Edge AI Application

Wendy LI
Asia AI Competence Center
STMicroelectronics



Agenda

- 1 Introduction to Edge AI
- 2 STM32Cube.AI- AI optimization tool for STM32
- 3 STM32Cube.AI for Desktop
- 4 STM32Cube.AI Developer Cloud

Introduction to Edge AI

Signals turning into data

Embedded applications will collect more data in the future



Growing demand for data-driven insights



Increasing use of sensors



Proliferation of IoT devices

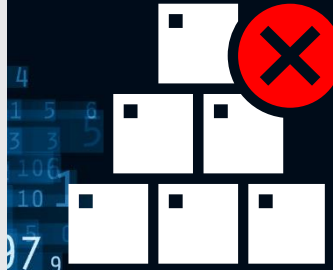


life augmented

AI is offering the best approach to process this growing amount of data



Algorithms and **predefined models** to analyze data and make predictions or decisions

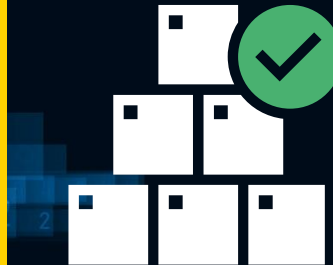


Traditional approaches show their limitations:

- when dealing with **large datasets**
- when the **phenomena are too complex**



Machine learning algorithms to automatically **learn** patterns and relationships **from the data**



AI-based data processing offers a more flexible and powerful approach to analyzing and making decisions from large data collection



The Augmented

The raise of Edge AI

Edge AI will benefit to many application domains:



Ultra-low latency
Real-time applications

01 Reduced data transmission
10 Generate meaningful information



Enhanced privacy and security
No data sharing in the cloud



Sustainable on energy
Low-data / Low-power



Improved accuracy
analyze data from a wide range of sensors and sources

Industrial maintenance
Condition monitoring
Predictive maintenance



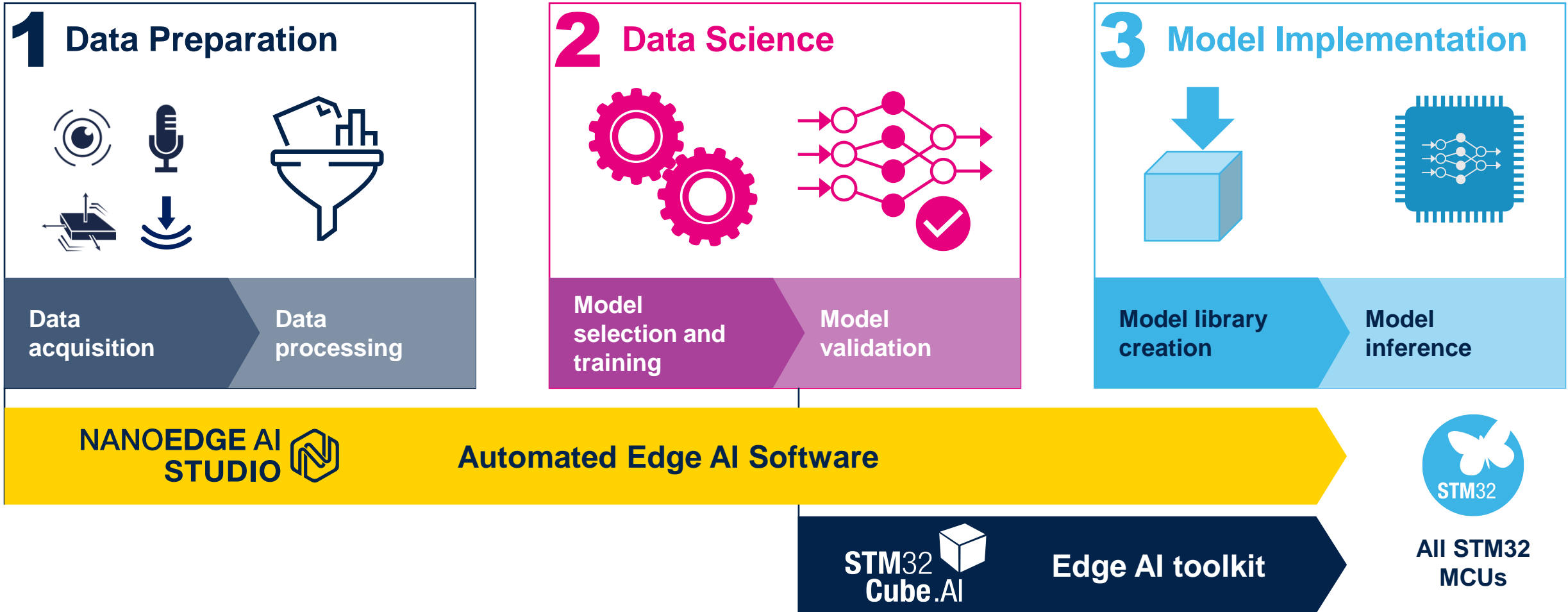
Control systems
From home heating systems
to industrial machines



Internet of Things (IoT)
smart cities, smart buildings,
connected homes, and
industrial automation



AI development workflow – ST software offering



ST ecosystem ease your AI to reach production level



	Edge AI toolkit for model optimization on STM32	Automated ML Software for end-to-end Edge AI solution design on STM32
Key benefits	<ul style="list-style-type: none">✓ Get optimized C-code from your trained model✓ Desktop and online versions✓ Benchmark service on remote hardware (online version)✓ On-device performance validation	<ul style="list-style-type: none">✓ The easiest way to integrate AI into your system✓ Save resource and development cost✓ Reach the highest performance with the automated model finder embedded in the tool
Application domain	All	Time series (except voice and speech)
Business model	Free of charge	Free for prototyping on STM32 dev boards Production requires right of use

A proven technology already adopted by multiple clients



INDUSTRIAL | DEMO

Fan anomaly detection based on vibrations

Learn to detect abnormal behavior at the edge on a vibrating machine.



INDUSTRIAL | CUSTOMER

AI solution for industrial predictive maintenance with NKE Watteco

Predictive maintenance solution for industrial equipment.



TRANSPORTATION | CUSTOMER

AI solution for monitoring automatic doors with Crouzet

Predictive maintenance on motors for automatic door motors.



INDUSTRIAL | DEMO

Anomaly detection in an electric motor

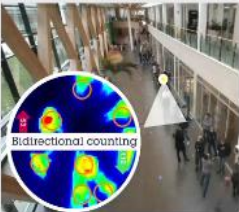
Current sensing to detect abnormal behaviors in motors.



INDUSTRIAL | CUSTOMER

AI solution for industrial predictive maintenance with Oxytronic

Predictive maintenance solution for industrial equipment.



SMART OFFICE | CUSTOMER

People flow counting Sensor with Schneider Electric

An innovative approach to measure people flows using an in-house thermal sensor.



SMART CITY | DEMO

Acoustic scene classification

Identify different environments (indoor, outdoor, in-car) using a simple microphone.



WEARABLES | DEMO

Human Activity Recognition

Easily identify 5 different activities with a 3D accelerometer.



INDUSTRIAL | DEMO

People presence detection (visual wake word)

Human detection on high-performance MCU.



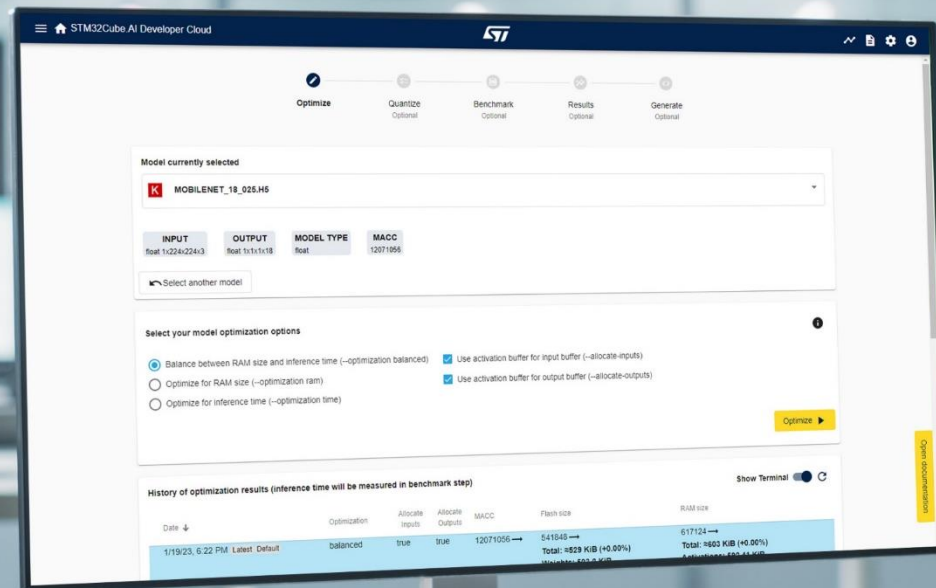
INDUSTRIAL | DEMO

Aftermarket wireless digit reader

Equip meters with aftermarket wireless & low-power readers.

STM32Cube.AI

AI optimization tool for STM32



STM32Cube.AI overview



STM32Cube.AI

The original desktop front end AI optimizer for STM32



X-CUBE-AI
for STM32Cube.MX

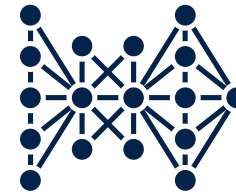


X-CUBE-AI
Command Line Interface



STM32Cube.AI Developer Cloud

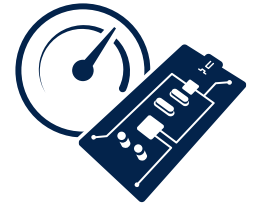
The brand-new online AI services front end for STM32



ST model zoo



Web GUI
+ REST API

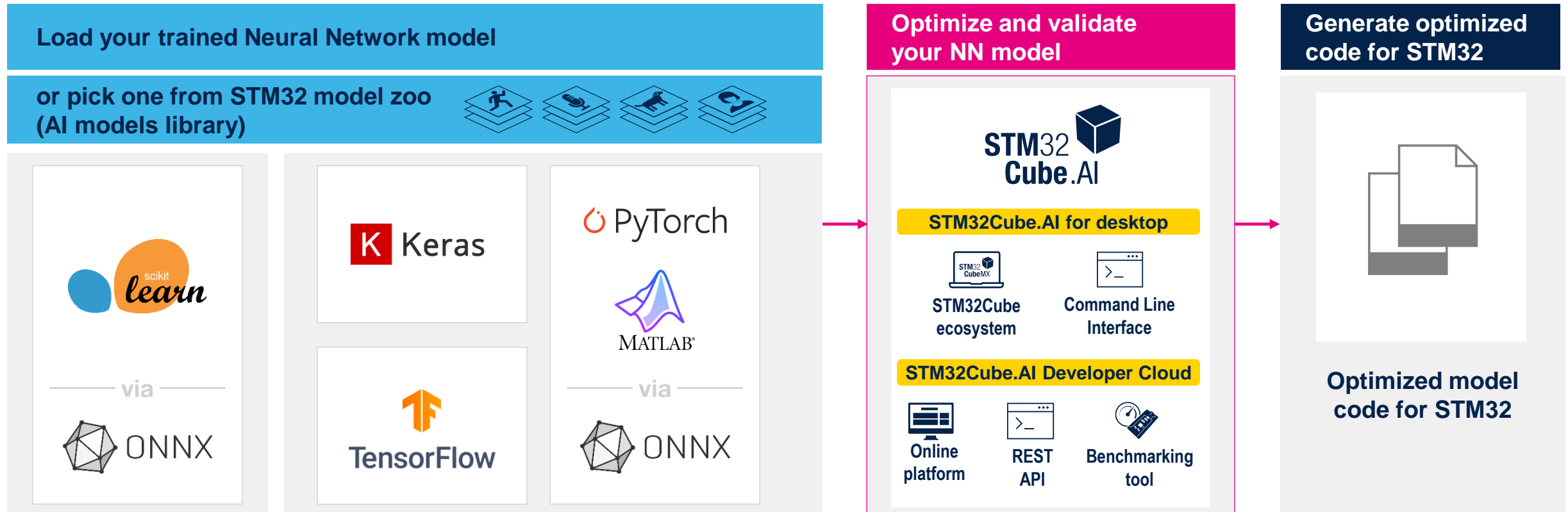


Board farm



Core engine technology

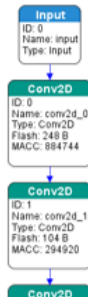
One tool – two versions to deploy AI on STM32



The 3 pillars of STM32Cube.AI

Graph optimizer

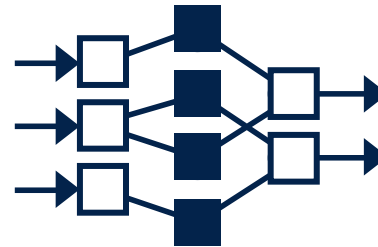
Automatically improve performance through graph simplifications & optimizations that benefit STM32 target HW architectures



- Auto graph rewrite
- Node/operator fusion
- Layout optimization
- Constant-folding...
- Operator-level info to fine-tune memory footprint and computation

Quantized model support

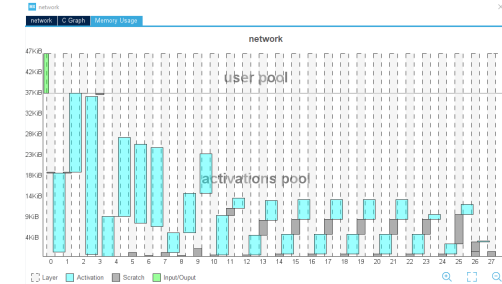
Import your quantized ANN to be compatible with STM32 embedded architectures while keeping their performance



- From FP32 to Int8 or mixed-precision
- Minimum loss of accuracy
- Code validation on target
 - Latency
 - Accuracy
 - Memory footprint

Memory optimizer

Optimize memory allocation to get the best performance while respecting the constraints of your embedded design



- Memory allocation
- Internal/external memory repartition
- Model-only update option

STM32Cube.AI is **free of charge**, available both in graphical interface and in command line.

Start with edge AI optimized models

STM32 model zoo

A collection of application-oriented models optimized for STM32

Human activity



Motion Sensing

Image classification



Computer vision

Audio event detection



Audio classification

Object detection



Computer vision



Hosted on Github



Model training scripts

- Scripts to generate and validate



Application code example

- Designed to host optimized NN models
- Automatically generated from the trained models
- Easy to deploy for end-to-end evaluation

What's new in STM32Cube.AI v8.0.0?

v8.0.0

Bringing higher degree of versatility with STM32Cube.AI

#

ONNX quantized models support

Introducing the support of **ONNX Tensor-oriented file format (QDQ)**:

Initial model converted in ONNX QDQ can be:

- ONNX models
- Quantization-Aware training (QAT) models from Tensorflow or exported from PyTorch
- Quantized models converted from TFLite and other frameworks

#

Up-to-date and improved code generation

- **Support for TensorFlow 2.11 models**
- **Support Keras.io 2.11**
- **Support ONNX Runtime 1.12.1**
- **New kernel performance improvements.**



Making Edge AI accessible to all STM32 portfolio

Take advantage of STM32Cube.AI on all STM32 series

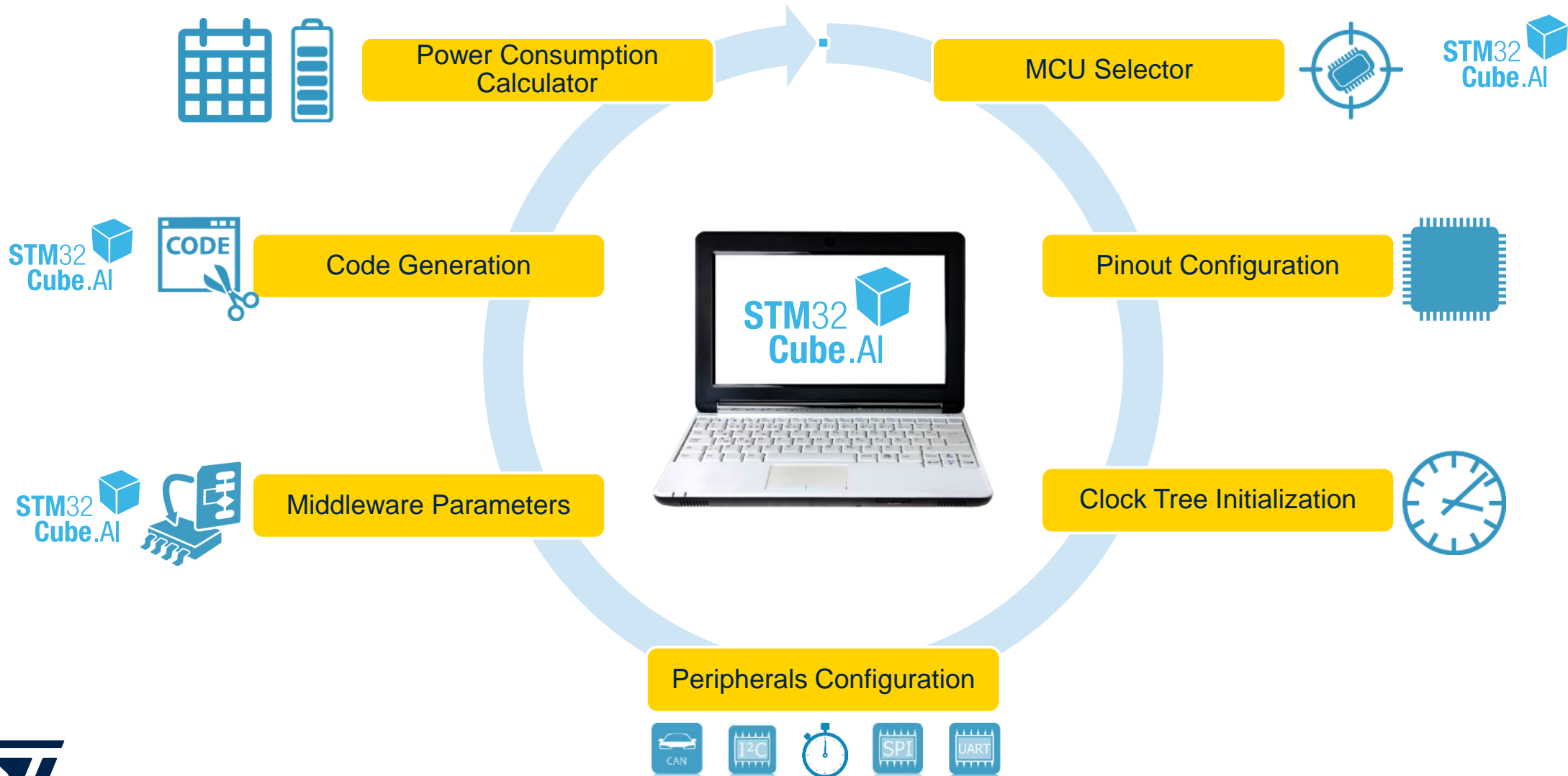
High Perf MCUs	STM32F2 Up to 398 CoreMark 120 MHz Cortex-M3		STM32F4 Up to 608 CoreMark 180 MHz Cortex-M4	STM32F7 1082 CoreMark 216 MHz Cortex-M7	STM32H7 Up to 3224 CoreMark Up to 550 MHz Cortex -M7 240 MHz Cortex -M4
	STM32F3 245 CoreMark 72 MHz Cortex-M4		STM32G4 569 CoreMark 170 MHz Cortex-M4	Mixed-signal MCUs	
Mainstream MCUs	STM32C0 114 CoreMark 48MHz Cortex M0+	STM32F0 106 CoreMark 48 MHz Cortex-M0	STM32G0 142 CoreMark 64 MHz Cortex-M0+	STM32F1 177 CoreMark 72 MHz Cortex-M3	
Ultra-low Power MCUs	STM32L0 75 CoreMark 32 MHz Cortex-M0+	STM32L1 93 CoreMark 32 MHz Cortex-M3	STM32L4 273 CoreMark 80 MHz Cortex-M4	STM32L4+ 409 CoreMark 120 MHz Cortex-M4	STM32L5 443 CoreMark 110 MHz Cortex-M33
Wireless MCUs			STM32WL 162 CoreMark 48 MHz Cortex-M4 48 MHz Cortex-M0+	STM32WB 216 CoreMark 64 MHz Cortex-M4 32 MHz Cortex-M0+	

Latest product generation

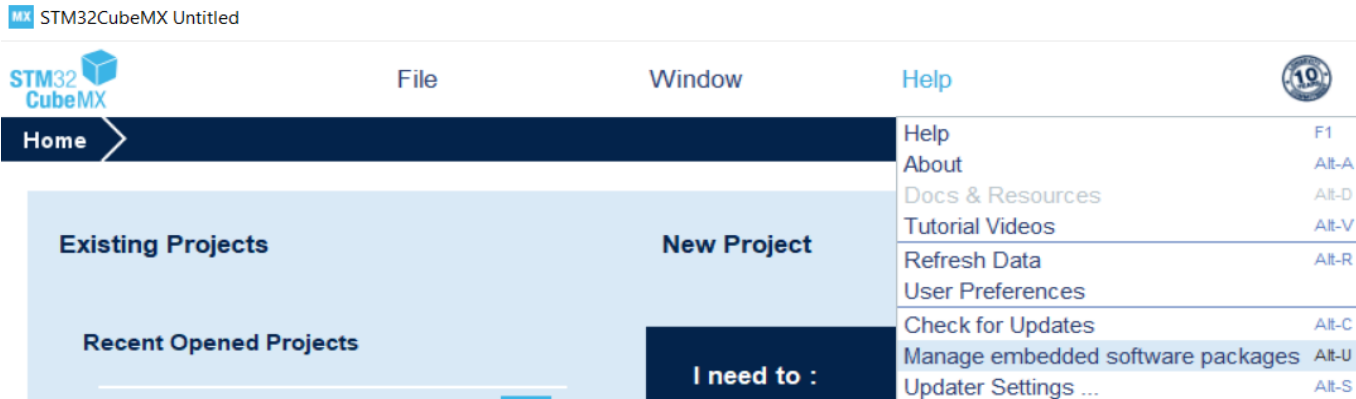
STM32Cube.AI for Desktop



STM32Cube.AI, an STM32CubeMX expansion

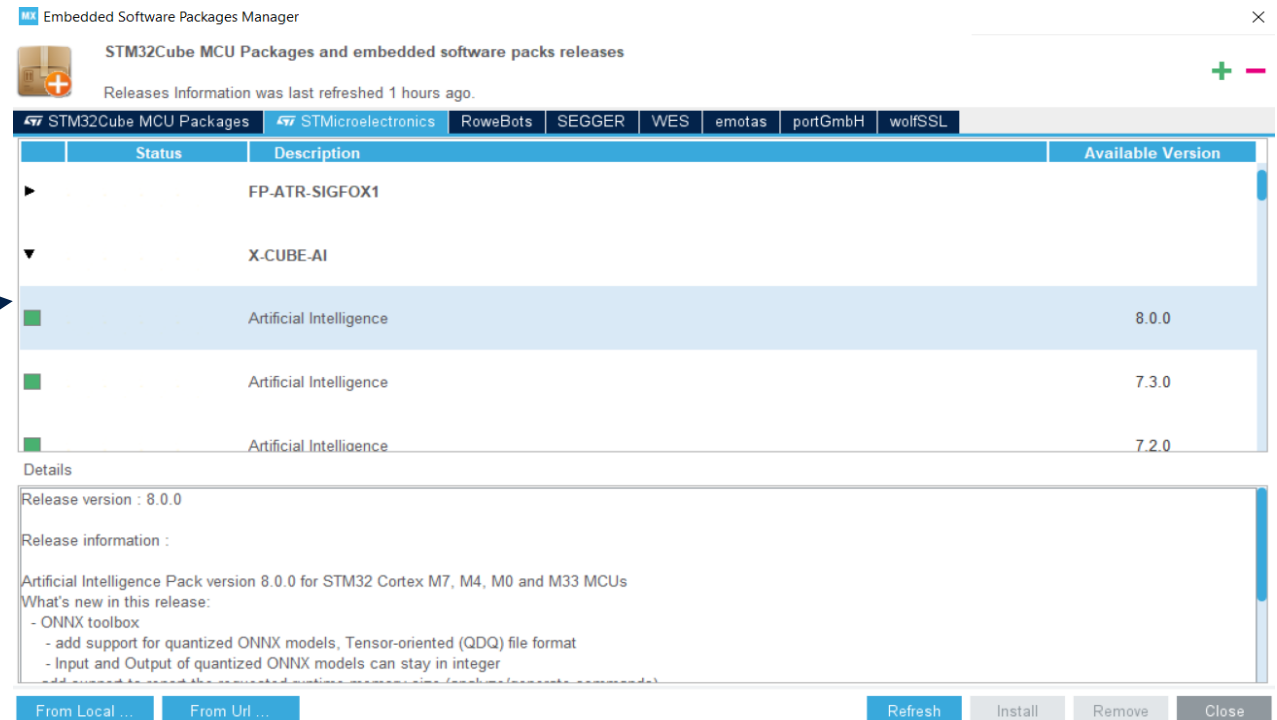


Install X-Cube-AI in STM32CubeMX



- Select “Manage embedded software packages” in Help item.

- find **X-CUBE-AI** in STMicroelectronics
- Install the latest version(8.0.0) in X-CUBE-AI.
- After installed, the square will turn on green in front of Artificial intelligence.



STM32Cube.AI user flow (1/4)

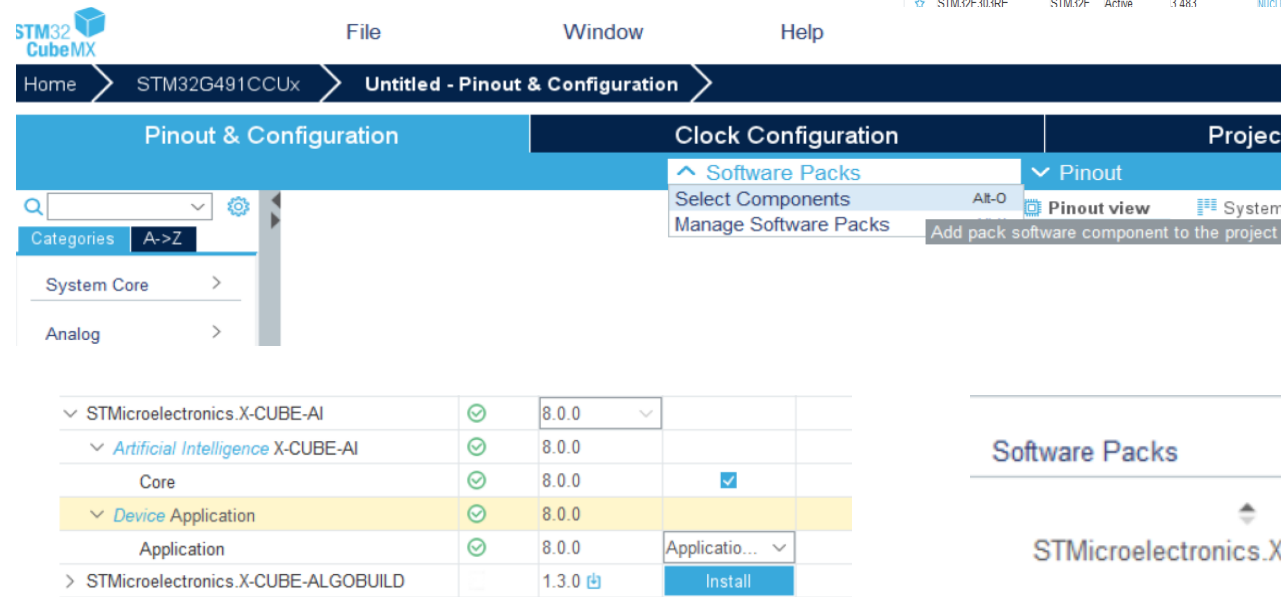
1 Select MCU, add X-CUBE-AI pack

2 Load and analyze NN Model

3 Optimize and validate

4 Generate project and integrate

- Select corresponding STM32 MCUs
- Select components X-CUBE-AI
- Find X-CUBE-AI in left software packs.



The screenshot shows the STM32CubeMX interface. The 'Pinout & Configuration' tab is active. On the left, a list of software packs is shown, including 'STMicroelectronics.X-CUBE-AI' and 'Artificial Intelligence X-CUBE-AI'. The 'Device Application' pack is highlighted. On the right, the 'Software Packs' list is expanded, showing 'STMicroelectronics.X-CUBE-AI 8.0.0' with an 'Install' button.

Part No.	Reference	Marketing	Unit Price for 10,000	Board	Package	Flash	RAM	IO	Freq.
STM32F302RD	STM32F...	Active	2.846		LQFP64	384 kBytes	64 kBytes	51	72 MHz
STM32F302RE	STM32F...	Active	3.24		LQFP64	512 kBytes	64 kBytes	51	72 MHz
STM32F302VD	STM32F...	Active	3.216		UFBGA100	384 kBytes	64 kBytes	86	72 MHz
STM32F302VE	STM32F...	Active	3.216		LQFP100	384 kBytes	64 kBytes	86	72 MHz
STM32F302ZE	STM32F...	Active	3.61		UFBGA100	512 kBytes	64 kBytes	86	72 MHz
STM32F302ZE	STM32F...	Active	3.61		LQFP100	512 kBytes	64 kBytes	86	72 MHz
STM32F302ZE	STM32F...	Active	3.795		LQFP144	384 kBytes	64 kBytes	115	72 MHz
STM32F302ZE	STM32F...	Active	4.188		LQFP144	512 kBytes	64 kBytes	115	72 MHz
STM32F303RD	STM32F...	Active	3.09		LQFP64	384 kBytes	80 kBytes	51	72 MHz
STM32F303RF	STM32F...	Active	3.483		LQFP64	512 kBytes	80 kBytes	51	72 MHz
					UFBGA100	384 kBytes	80 kBytes	86	72 MHz

AI Summary

Minimum Flash: 214.04 KiB
Minimum Ram: 46.19 KiB

C:\Data\work\Models\person_model_grayscale\model.tflite

MCUs/MPUs List: 836 items [Display similar items](#) [Export](#)

*	Part No.	Reference	Marketing	Unit Price for 10,000	Board	Package	Flash	RAM	IO	Freq.
☆	STM32F302RD	STM32F...	Active	2.846		LQFP64	384 kBytes	64 kBytes	51	72 MHz
☆	STM32F302RE	STM32F...	Active	3.24		LQFP64	512 kBytes	64 kBytes	51	72 MHz
☆	STM32F302VD	STM32F...	Active	3.216		UFBGA100	384 kBytes	64 kBytes	86	72 MHz
☆	STM32F302VE	STM32F...	Active	3.216		LQFP100	384 kBytes	64 kBytes	86	72 MHz
☆	STM32F302ZE	STM32F...	Active	3.61		UFBGA100	512 kBytes	64 kBytes	86	72 MHz
☆	STM32F302ZE	STM32F...	Active	3.61		LQFP100	512 kBytes	64 kBytes	86	72 MHz
☆	STM32F302ZE	STM32F...	Active	3.795		LQFP144	384 kBytes	64 kBytes	115	72 MHz
☆	STM32F302ZE	STM32F...	Active	4.188		LQFP144	512 kBytes	64 kBytes	115	72 MHz
☆	STM32F303RD	STM32F...	Active	3.09		LQFP64	384 kBytes	80 kBytes	51	72 MHz
☆	STM32F303RF	STM32F...	Active	3.483		LQFP64	512 kBytes	80 kBytes	51	72 MHz
						UFBGA100	384 kBytes	80 kBytes	86	72 MHz

1

Select MCU, add X-CUBE-AI pack

2

Load and analyze NN Model

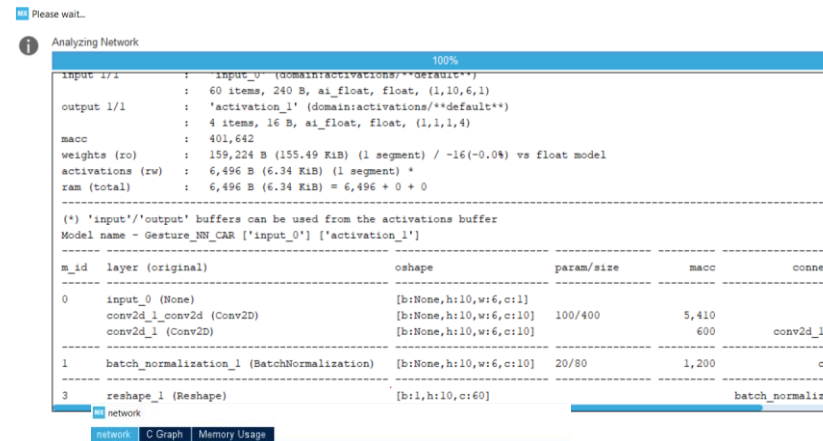
3

Optimize and validate

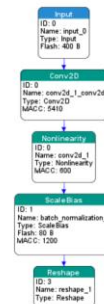
4

Generate project and integrate

- Load NN model. (h5, tflite, onnx)
- Model complexity and footprint analysis
- Check Flash, Ram usage.



network



STMicroelectronics X-CUBE-AI 8.0.0 Mode and Configuration

Mode

☒ Artificial Intelligence X-CUBE-AI

☒ Device Application

Configuration

Reset Configuration Add network Delete network

Main network +

Model inputs

network

Keras Saved model

Model: 095549_epochs50_modelCAR/Gesture_NN_CAR.h5 Browse... Browse...

Compression: None Optimization: Balanced

Validation inputs: Random numbers

Validation outputs: None

Complexity: 401642 MACC
Used Flash: 179.96 KiB (179.96 KiB over 256.00 KiB Internal)
Used Ram: 9.67 KiB (9.67 KiB over 112.00 KiB Internal)
Achieved compression: -

Show graph

Analyze

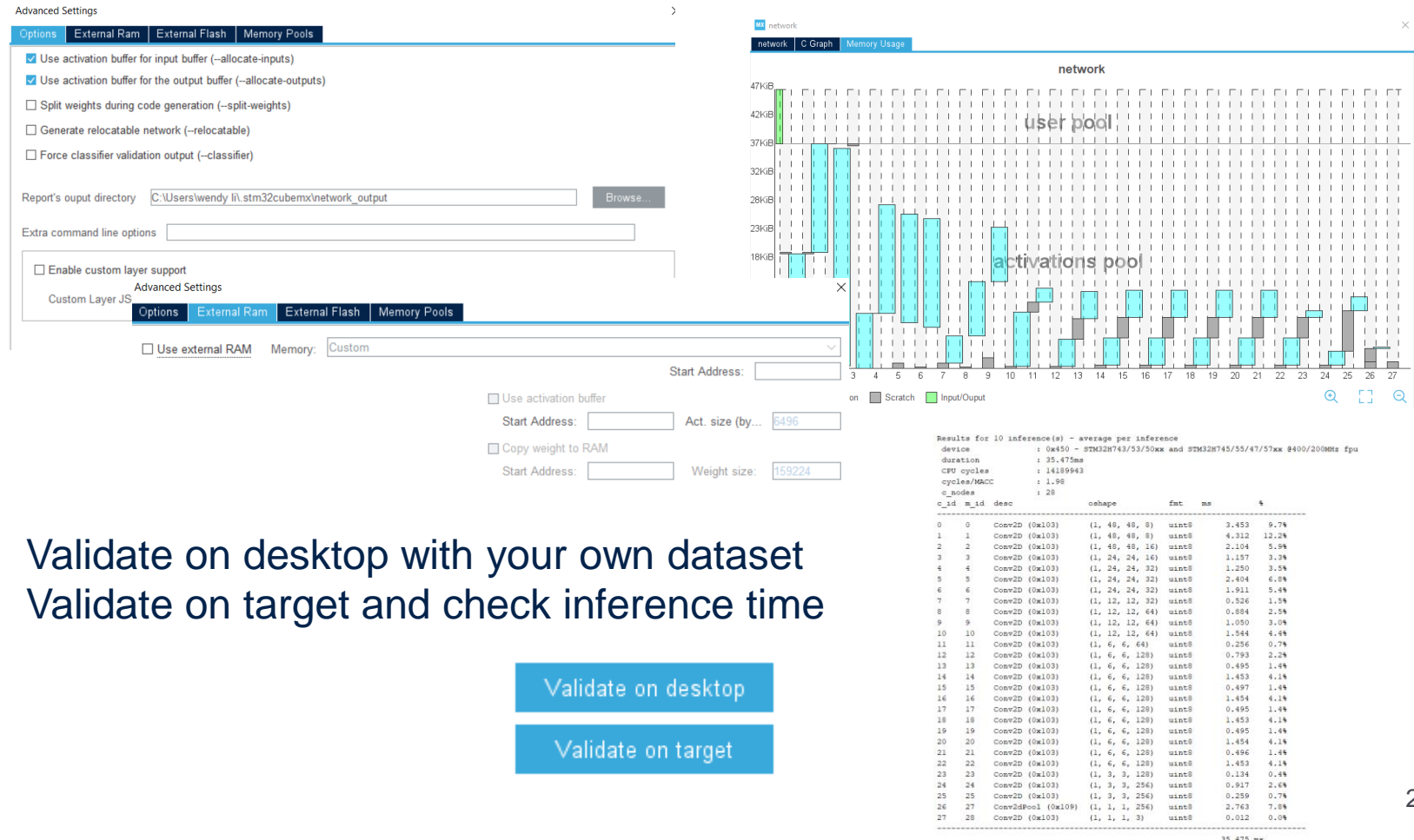
Validate on desktop

Validate on target

STM32Cube.AI user flow (3/4)

- 1 Select MCU, add X-CUBE-AI pack
- 2 Load and analyze NN Model
- 3 Optimize and validate
- 4 Generate project and integrate

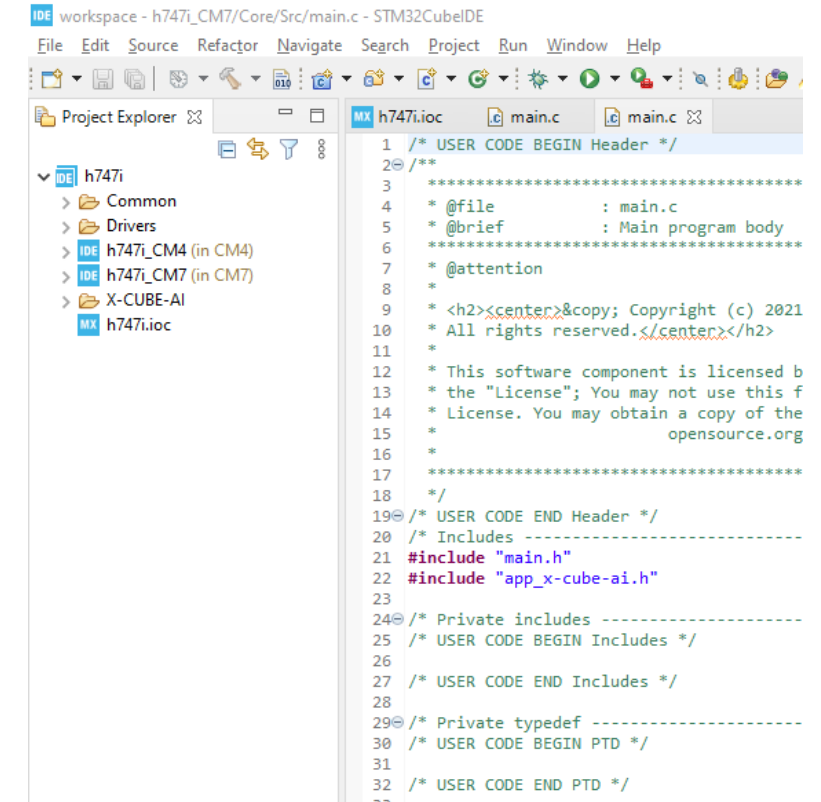
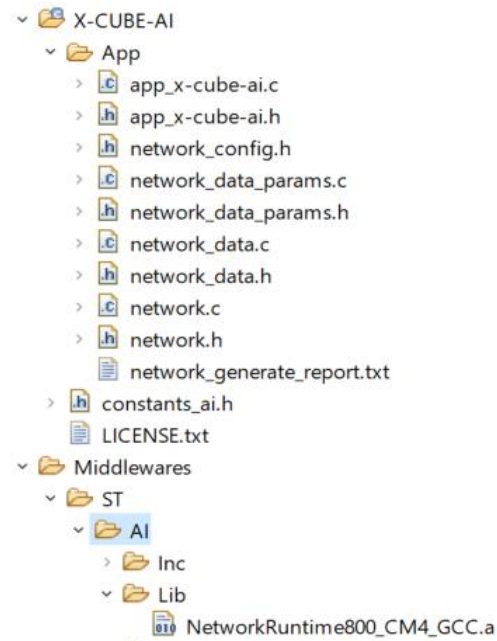
- Fine tune memory allocation with optimizations and GUI
- Optimize system parameters and clock tree
- Extend model with your own customer layers



- 1 Select MCU, add X-CUBE-AI pack
- 2 Load and analyze NN Model
- 3 Optimize and validate
- 4 Generate project and integrate

GENERATE CODE

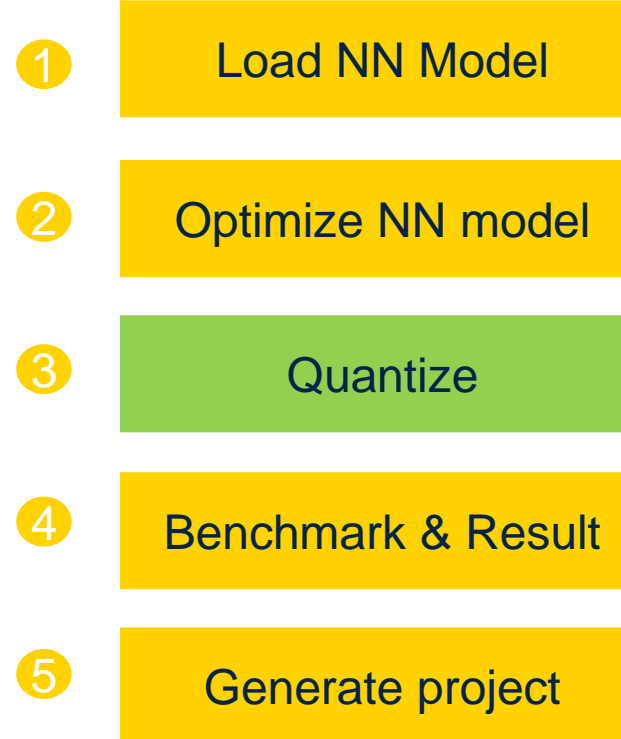
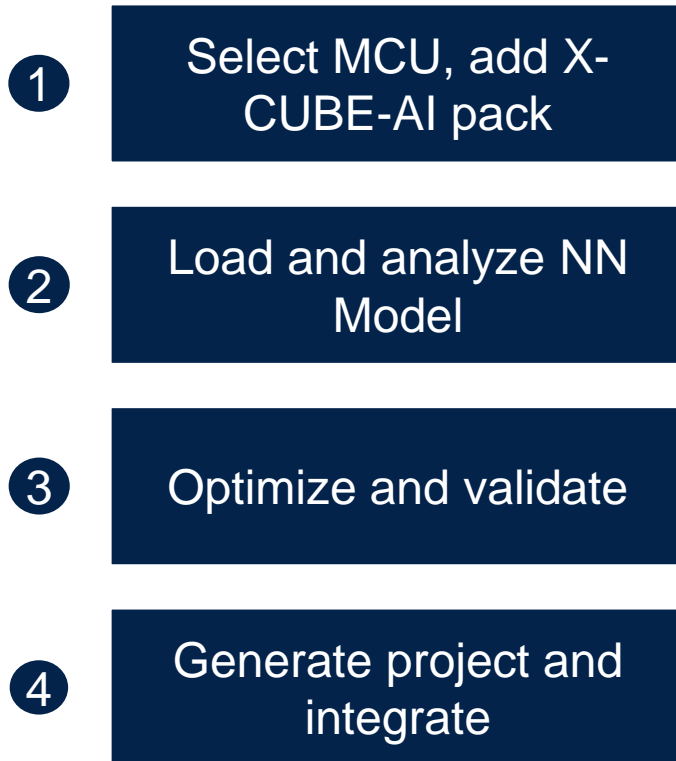
- Generate Application Template
- Integrate with your application-specific code in your favorite IDE
- Perform system tests



STM32Cube.AI Developer Cloud

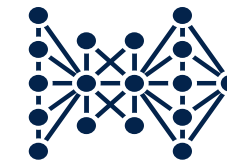


Cube AI vs Developer Cloud

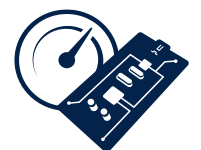


STM32Cube. AI Developer Cloud has these new benefits:

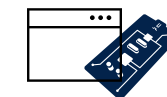
- Model Zoo
- Quantized tool
- Board farm
- Online, No need for tool installation



ST model zoo



Board farm



Benchmarking tool

How to Find STM32Cube.AI Developer Cloud?

- Website: <https://stm32ai.st.com/stm32-cube-ai-dc/>

- Open online platform through ST website.
- Login in with your email, password

Already registered?

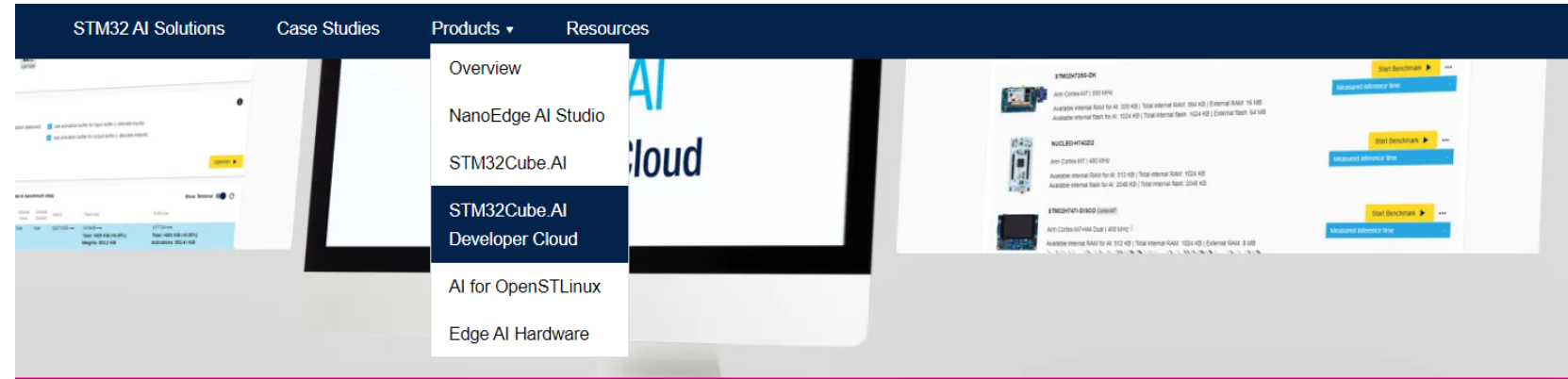
Enter your e-mail address and password to login your myST user.

E-mail address

Password

☐ Remember me on this computer. ⓘ

Login



Online tool for fast AI optimization and benchmark

Open online platform

STM32Cube.AI Developer Cloud is a free-of-charge online platform and services allowing the creation, optimization, benchmarking, and generation of AI for the STM32 microcontrollers. It is based on the STM32Cube.AI core technology.

Benefits include access to online service to benchmark AI on STM32 boards. It comes with the ST Model zoo including reference model, training script and application examples.

Main features

Online GUI secured by ST user credentials

AI optimizer and visualizer

AI benchmark service on various STM32

STM32 model zoo with training scripts

Application code examples

Model quantization tool

Load NN Model

- Load a model from **ST model zoo** or **own your trained model** from your favorite AI framework: Tensorflow lite, Keras, Pytorch, ONNX, scikit Learn, ...
- Supported format are FLOAT32 or quantized INT8

The model you selected or uploaded before

STM32Cube.AI Developer Cloud

8.0.1

OPTIMIZE YOUR TRAINED NEURAL NETWORK

Optimize and measure performance of your Artificial Intelligence library for STM32 MCUs
This free online tool allows you to generate and test optimized STM32 AI libraries from your trained Neural Networks

Drop your model or click here to open a filebrowser
Supported models are Keras, ONNX and TFLite (.h5, .hdf5, .onnx, .tflite)
Max file size: 32 MiB

Upload

CHOOSE A MODEL FROM YOUR WORKSPACE

Show previously quantized models

Models

Sort by: date

	FoodReco_MobileNet_Derivative_Float.h5 Content Length: 1.7 MiB Last modified: Apr 6, 2023	
	fdmobilenet_0.25_128_int8_image_classification_flow...	

CHOOSE A MODEL FROM STM32 MODEL ZOO

	Description: MiniResNet 2 stacks 64x50 trained on E... Use case: audio_event_detection Dataset used: esc10	
	miniresnetv2_1stacks_64x50_int8_audio_event_dete... Content Length: 141.75 KiB Description: MiniResNetv2 1 stacks 64x50 trained on ... Use case: audio_event_detection Dataset used: esc10	
	miniresnetv2_2stacks_64x50_int8_audio_event_dete... Content Length: 469.37 KiB Description: MiniResNetv2 2 stacks 64x50 trained on ... Use case: audio_event_detection Dataset used: esc10	

Optimize NN model(1/2)

- Select your preferred model optimization options between the Ram or latency most critical parameter.
- Click “**Optimize**” to check the optimization result

STM32Cube.AI Developer Cloud

8.0.1

Optimize Quantize Optional Benchmark Optional Results Optional Generate Optional

Model currently selected

K FOODRECO_MOBILENET_DERIVATIVE_FLOAT.H5

INPUT	OUTPUT	MODEL TYPE	MACC
float 1x224x224x3	float 1x1x1x18	float	12071056

Select another model

Select your model optimization options

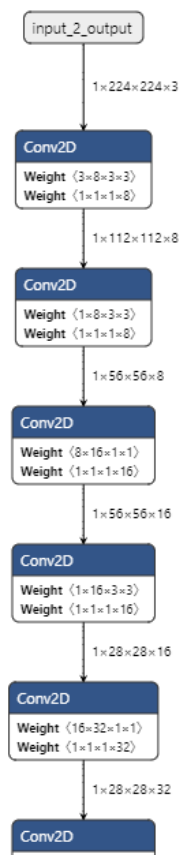
☒ Balance between RAM size and inference time (--optimization balanced) ☒ Use activation buffer for input buffer (--allocate-inputs)

☐ Optimize for RAM size (--optimization ram) ☒ Use activation buffer for output buffer (--allocate-outputs)

☐ Optimize for inference time (--optimization time)

Optimize ▶

Optimize NN model(2/2)



History of optimization results (inference time will be measured in benchmark step)

Show terminal ☒

Date	Version	Optimization	Allocate Inputs	Allocate Outputs	MACC	Flash size ↑	RAM size
4/6/23, 10:53 AM	8.0.1	time	true	false	12071056 →	541740 ↘	1014084 ↗
4/6/23, 10:54 AM Latest	8.0.1	balanced	true	true	12071056 →	541784 → Total: ≈529 KiB (+0.00%) Weights: 503.2 KiB Library (estimation) ≈26 KiB	617120 → Total: ≈603 KiB (+0.00%) Activations: 592.41 KiB Library (estimation) ≈10 KiB Input/Output: 0 B / 0 B
<div> <div>Show graph </div> <div>Delete </div> </div> <div>Compare with default </div>							<div>Go to quantize </div> <div>Go to benchmark </div>
4/6/23, 10:53 AM	8.0.1	time	true	true	12071056 →	541784 →	1014016 ↗
4/6/23, 10:53 AM	8.0.1	ram	true	true	12071056 →	541784 →	617120 →
4/6/23, 10:51 AM Default	8.0.1	balanced	true	true	12071056 →	541784 →	617120 →

Items per page: 5 1 - 5 of 5 < < > >

Terminal

```
>>> stm32ai validate --model FoodReco_MobileNet_Derivative_Float.h5 --workspace workspace --output output --allocate-inputs --allocate-outputs --relo
Neural Network Tools for STM32AI v1.7.0 (STM.ai v8.0.1-19409)
Applied 6 of 7 optimizations
Copying the AI runtime files to the user workspace: workspace

Exec/report summary (validate)
-----
model file      : FoodReco_MobileNet_Derivative_Float.h5
```

Quantize(optional)

- Perform post-training quantization to further reduce your NN model parameter size

STM32Cube.AI Developer Cloud 8.0.1

Optimize **Quantize Optional** Benchmark Optional Results Optional Generate Optional

Model currently selected

FOODRECO_MOBILENET_DERIVATIVE_FLOAT.H5

INPUT: float 1x224x224x3 OUTPUT: float 1x1x1x1 MODEL TYPE: float MACC: 12071056

Select another model

Current parameters

ALLOCATE INPUTS: true ALLOCATE OUTPUTS: true OPTIMIZATION: balanced

Change parameters Skip quantize >

Apply post-training quantization (using TensorFlow Lite Converter)

Input Type: ☒ int8 ☐ uint8 ☐ float32

Output Type: ☒ int8 ☐ uint8 ☐ float32

Load a dataset to check the accuracy obtained after quantization

Load file (.npz)

If no quantization file is provided, quantization will occur with random values

Launch quantization

Quantized models

FoodReco_MobileNet_Derivative_Float_quant_int8_int8_random_1.tflite
Content Length: 196.45 KiB
Last Modified: 4/6/23, 11:03 AM

Optimize selected quantized model

History of optimization results (inference time will be measured in benchmark step)

Show terminal

Date	Version	Optimization	Allocate Inputs	Allocate Outputs	MACC	Flash size	RAM size
4/6/23, 11:03 AM	Quantized	8.0.1	balanced	true	true	11801432	196256
							Total: ~192 KiB (-63.8%) Weights: 131.62 KiB Library (estimation) ~60 KiB
							Total: ~168 KiB (-72.1%) Activations: 152.05 KiB Library (estimation) ~16 KiB Input/Output: 0 B / 0 B

Show graph Delete

Compare with default

Go to benchmark

Items per page: 5 1 - 1 of 1

Terminal

```
>>> stm32ai validate --model FoodReco_MobileNet_Derivative_Float_quant_int8_int8_random_1.tflite --workspace workspace --output output --allocate-inp
Neural Network Tools for STM32AI v1.7.0 (STM.ai v8.0.1-19400)
Copying the AI runtime files to the user workspace: workspace

Exec/report summary (validate)
```

Benchmark (optional)

STM32Cube.AI Developer Cloud 8.0.1

Optimize Quantize Optional Benchmark Optional Results Optional Generate Optional

Model currently selected

FOODRECO_MOBILENET_DERIVATIVE_FLOAT_QUANT_INT8_INT8_RAND0

INPUT float 1x224x224x3 OUTPUT float 1x1x1x18 MODEL TYPE float MACC 12071056



Select another model

Current parameters

ALLOCATE INPUTS true ALLOCATE OUTPUTS true OPTIMIZATION balanced

Change parameters

Schedule a benchmark

 <p>STM32H735G-DK Arm Cortex-M7 550 MHz Available internal RAM for AI: 560 KB Total internal RAM: 564 KB External RAM: 16 MB Internal flash: 1024 KB External flash: 64 MB</p>	<p>Measured inference time 45.33 ms</p>
 <p>NUCLEO-H743ZI2 Arm Cortex-M7 480 MHz Available internal RAM for AI: 1056 KB Total internal RAM: 1024 KB Internal flash: 2048 KB</p>	<p>Start Benchmark</p> <p>Measured inference time -</p>
<p>STM32H747I-DISCO CortexM7</p>	<p>Start Benchmark</p>

- Access the online benchmark service to remotely evaluate your AI performance on various STM32 boards.
- 9 boards available for checking inference time.

[STM32H735G-DK](#)
[NUCLEO-H743ZI2](#)
[STM32H747I-DISCO](#)
[STM32H7B3I-DK](#)
[STM32F769I-DISCO](#)
[B-U585I-IOT02A](#)
[NUCLEO-G474RE](#)
[STM32F469I-DISCO](#)
[STM32L4R9I-DISCO](#)

Result (optional)

- Visualize and optionally export the performance benchmark report

STM32Cube.AI Developer Cloud

Optimize Quantize Optional Benchmark Optional Results Optional Generate Optional

Model currently selected

FOODRECO_MOBILENET_DERIVATIVE_FLOAT_QUANT_INT8_INT8_RANDO

INPUT OUTPUT MODEL TYPE MACC

float 1x224x224x3 float 1x1x1x18 float 12071056

Select another model

Current parameters

ALLOCATE INPUTS true ALLOCATE OUTPUTS true OPTIMIZATION balanced

Change parameters Go next

History of benchmark results

Show performance summary Export as CSV Refresh Delete

Date	Model Name	Model Type	MACC	Duration (ms)	Weights Size	Activation Size	Flash size	RAM size	Device	Clock	Use External Flash
4/6/2023, 11:15:18 AM	FoodReco_M...	tfliite	11801432	45.33	131.62 KiB	387.48 KiB	195.65 KiB	403.62 KiB	STM32H735G-DK	550MHz	false
4/4/2023, 6:30:47 PM	fdmobilenet_...	tfliite	3952906	16.21	128.32 KiB	203.95 KiB	193.59 KiB	220.09 KiB	STM32H735G-DK	550MHz	false
2/8/2023, 5:26:26 PM	arc_detect_m...	keras	37216	1.982	144.76 KiB	2.25 KiB	160.12 KiB	5.3 KiB	NUCLEO-G474RE	170MHz	false
2/8/2023, 3:44:19 PM	arc_detect_m...	tfliite	265448	4.753	259.57 KiB	12.06 KiB	279.61 KiB	15.07 KiB	NUCLEO-G474RE	170MHz	false

Show graph Show details per layer Memory pools

Generate with this



Generate (optional)

- Generate your optimized AI code in C-code or in a format compatible with the STM32CubeMX software suites.

Select CPU
and boards
you are using

The screenshot displays the STM32Cube.AI Developer Cloud interface. At the top, a progress bar shows five steps: Optimize (checked), Quantize (checked, Optional), Benchmark (checked, Optional), Results (checked, Optional), and Generate (checked, Optional). Below this, the 'Model currently selected' section shows 'FOODRECO_MOBILENET_DERIVATIVE_FLOAT_QUANT_INT8_INT8_RAND0'. The 'Current parameters' section shows 'ALLOCATE INPUTS' (true), 'ALLOCATE OUTPUTS' (true), and 'OPTIMIZATION' (balanced). The 'CPU/Series filters' section shows 'Arm Cortex-M7' and 'Ex: STM32F4'. The 'Boards' section shows 'STM32F746G-DISCO'. Below the progress bar, there are four download options: 'Download C Code' (with a code icon), 'Download STM32CubeMX IOC file' (with an MX icon), 'Download STM32CubeIDE Project' (with an IDE icon), and 'Download Firmware' (with an ELF icon). Each option includes a description of what is generated and a 'Download' button with a download icon. On the right side, the 'STM32F746G-DISCO' board is detailed with its specifications: Arm Cortex-M7 | STM32F7 Series | 216 MHz, RAM: 320 KB (Max: 320 KB), Flash: 1024 KB (Max: 1024 KB), External RAM: 16 MB, and External Flash: 16 MB. It also includes an image of the board and a description of the kit as a complete demonstration and development platform for STM32F746NG (STM32F746NGH6 order code) microcontroller.

Model currently selected

FOODRECO_MOBILENET_DERIVATIVE_FLOAT_QUANT_INT8_INT8_RAND0

INPUT: float 1x224x224x3, OUTPUT: float 1x1x1x18, MODEL TYPE: float, MACC: 12071056

Select another model

Current parameters

ALLOCATE INPUTS: true, ALLOCATE OUTPUTS: true, OPTIMIZATION: balanced

Change parameters

CPU/Series filters

Arm Cortex-M7 Ex: STM32F4

Boards

STM32F746G-DISCO

Download C Code

Generates STM32-optimized C Code associated with your pre-trained Neural Network

Download

Download STM32CubeMX IOC file

Version: 6.8.0 or greater

Generates a ZIP package containing an IOC file and the selected model, ready to start STM32CubeMX locally on your machine

Download

Download STM32CubeIDE Project

Version: 1.12.0 or greater

Generates STM32CubeIDE project, including IOC file, Project File, file tree and STM32-optimized C Code

Download

Download Firmware

Generates ELF file(s) associated to a selected board, to flash locally your target using STM32CubeProgrammer

Download

STM32F746G-DISCO

Arm Cortex-M7 | STM32F7 Series | 216 MHz

RAM: 320 KB (Max: 320 KB)

Flash: 1024 KB (Max: 1024 KB)

External RAM: 16 MB

External Flash: 16 MB

The 32F746GDISCOVERY Discovery kit is a complete demonstration and development platform for STMicroelectronics Arm® Cortex®-M7-core-based STM32F746NG (STM32F746NGH6 order code) microcontroller.

The 32F746GDISCOVERY Discovery kit is used as reference design for user application

Don't go alone

We have created a network of companies to support you

Partner
Program



Trust our **authorized partners** to ensure the success of your project. Learn more at st.com/stm32ai



Wish to discuss a co-development partnership for ML/AI projects? Contact us at edge.ai@st.com



life.augmented

Please Scan the QR Codes
and Stay Tuned with Us.



PDSA Wechat Subscription



Power & SPIN Microsite



Our technology starts with You



Find out more at www.st.com

© STMicroelectronics - All rights reserved.

ST logo is a trademark or a registered trademark of STMicroelectronics International NV or its affiliates in the EU and/or other countries.

For additional information about ST trademarks, please refer to www.st.com/trademarks.

All other product or service names are the property of their respective owners.



life.augmented