

# Wildcount

**Inexpensive Edge sensor for recognizing and counting the presence of humans (anonymous) and animals into wild and protected areas.**

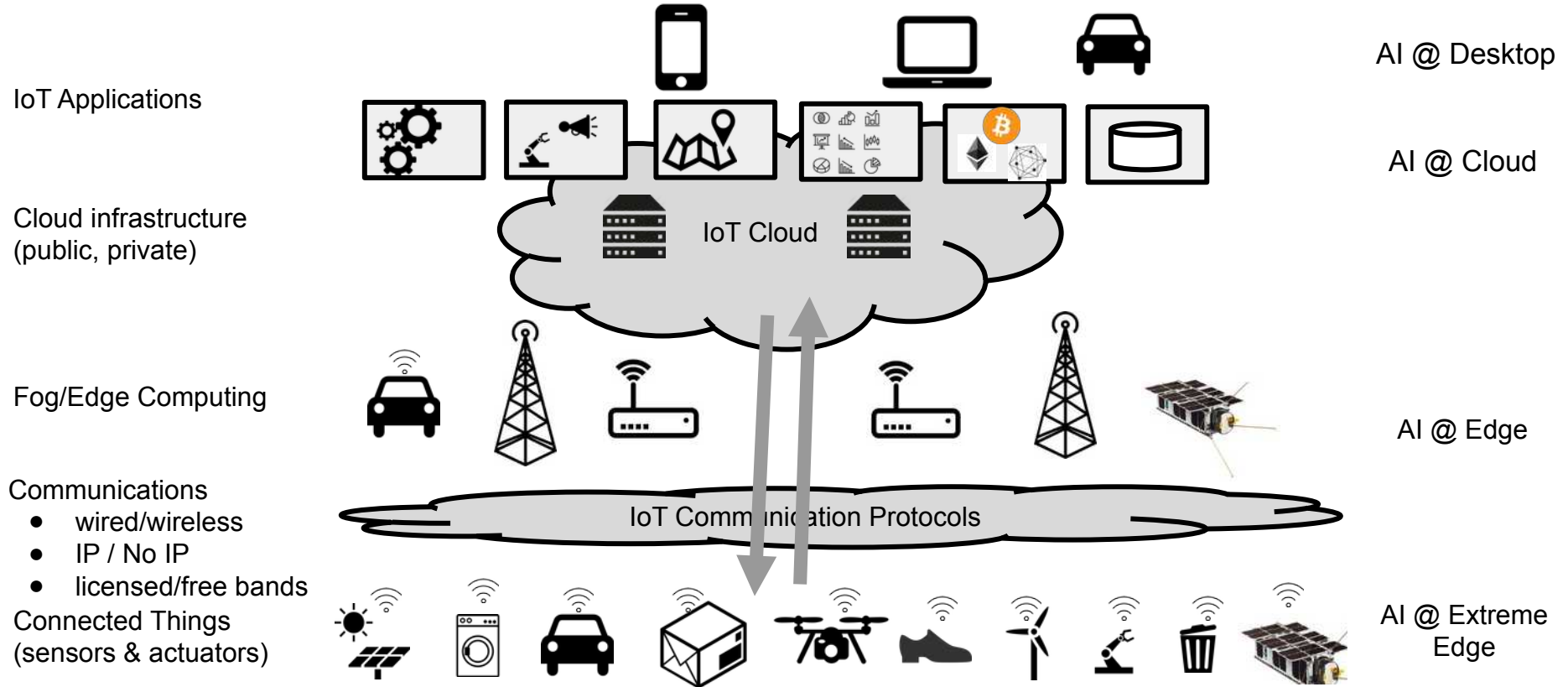
Léo Cordier, Georges QUENOT, Didier DONSEZ  
Université Grenoble Alpes  
LIG MRIM & ERODS

Grenoble



This work has been partially supported by [MIAI @ Grenoble Alpes](#). (ANR-19-P3IA-0003).

# IoT Infrastructure and AI



# Low-Power and Long Range WAN

Example: Elsys ELT2 (LoRa endpoint with temperature sensor)

- Battery lifetime from **6** to **18** years (**1 temp. msg/hour**)



Sample time: 3600 Seconds  
Sensor: ELT2 HP  
Battery capacity: 2700 Capacity(mAh)  
Battery performance: 80 Performance(%)

Spreading factor:  
 SF7  SF8  SF9  SF10  SF11  SF12

Result:

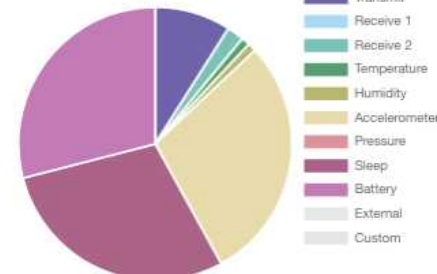
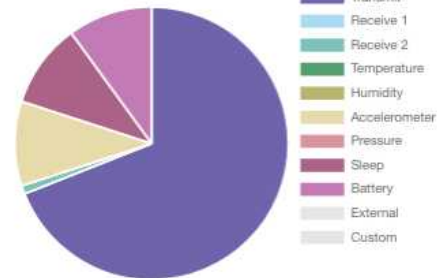
The battery will last for **6.2** years\*. The sensor will draw **40uA** and **351mAh** in one year.

Sample time: 3600 Seconds  
Sensor: ELT2 HP  
Battery capacity: 2700 Capacity(mAh)  
Battery performance: 80 Performance(%)

Spreading factor:  
 SF7  SF8  SF9  SF10  SF11  SF12

Result:

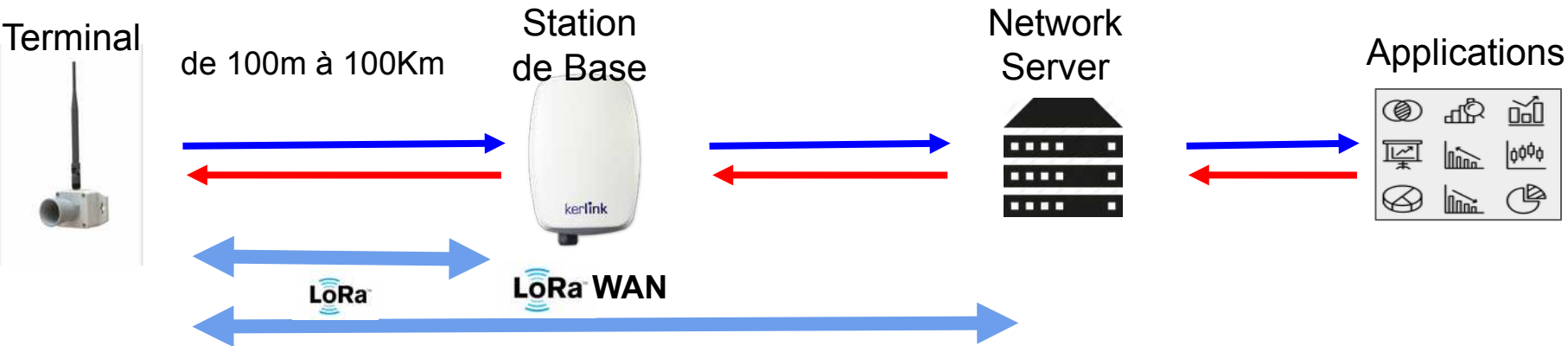
The battery will last for **18** years\*. The sensor will draw **14uA** and **120mAh** in one year.



# LoRa vs LoRAWAN



- LoRa : Technologie Radio LPWAN
  - Définit la couche physique d'une pile réseau : LoRa PHY
- LoRaWAN : un protocole réseau "sur" LoRa
  - Définit la couche MAC (Medium Access Control)



# Terra Forma (Equipex 2021-2029)

<https://terra-forma.cnrs.fr/>

Réseaux IoT pour observer les transitions  
dans les espaces naturels sensibles.

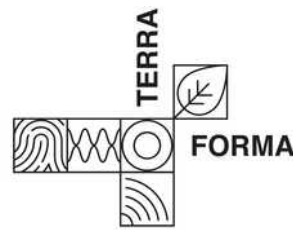
Technologies : LPWAN, LPGAN (SatIoT), AI @ Edge, AI @ Extreme Edge

Zones atelier

dont Col du Lautaret (SAJF) et Parc National des Ecrins (PNE)



# Context : Eco-biology



2 Renards !



latency ~1-2  
months

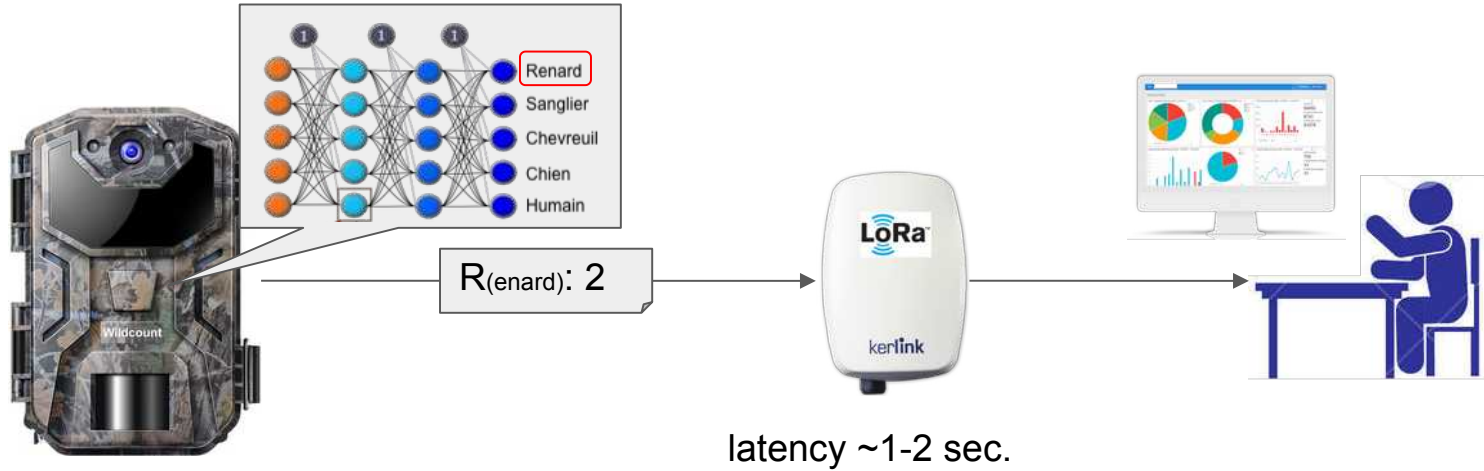
Remark: Outdoor Hunting Trail Camera can be stolen or destroyed  
→ photo campaign is lost

Remark: Indexing @ Home: human or automatic ([DeepFaune])

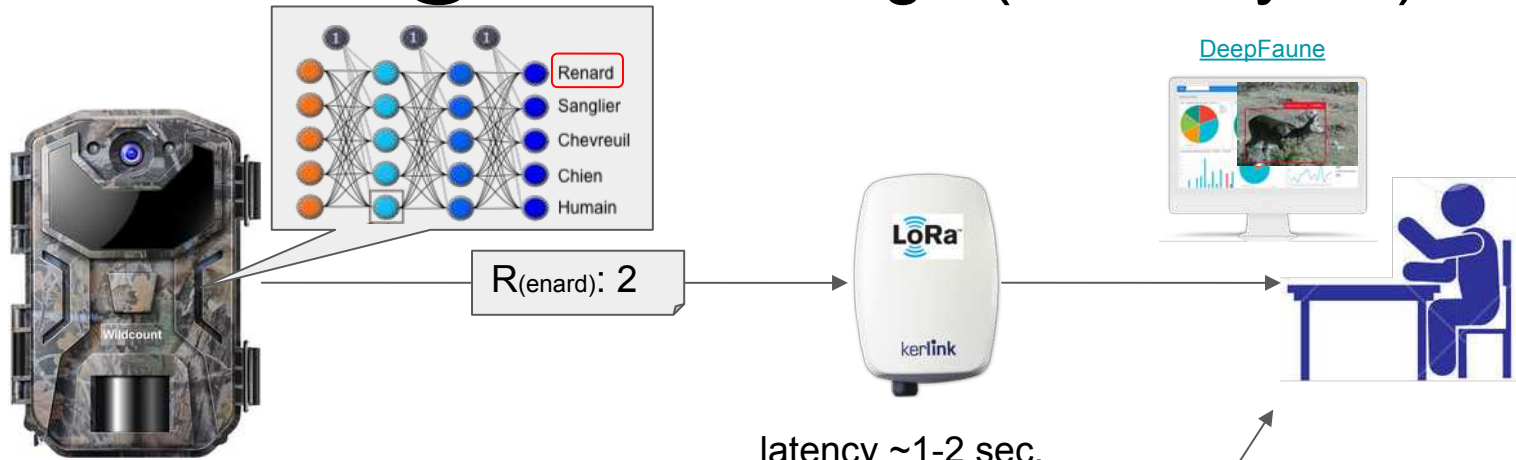
[DeepFaune] <https://www.deepfaune.cnrs.fr/>



# Wildcount



# Wildcount : AI @ Extreme Edge (aka TinyML)



Fix embedded inferences



latency ~1-2



# Quick reminder about ML on DL

## Data

small/huge set of labeled/unlabelled time-series (temps), texts, photos, videos ...

## Supervised

Training on labelled dataset

Linear models, Tree-based models, ((Very) Deep) Neural Networks

## Unsupervised

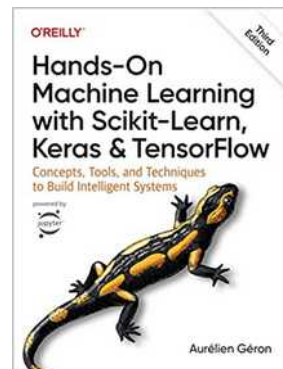
Clustering, Association (Apriori), Matrix Profile (for time-series) ...

## Frameworks


Keras, Tensorflow, PyTorch, [MindSpore](#) ...

## Studios

Jupyter, [Google Colab](#), [Edge Impulse](#) ...



# Wildcount's Aims

- Internet of Things endpoint for
  - counting of wild fauna in natural and protected areas
  - counting of (anonymous) users of natural and protected areas
- Technical requirements
  - Low power / Long battery lifetime
  - Low RAM
  - Long range transmissions of counters (in “real time”)
  - Transmissions in white zones (no 2345G coverage, private LPWAN)
- Low Power AI-embedded processors
  - Neural networks trained with Alps fauna pictures taken from hunting camera and images databases.
  - LoRaWAN communications 

# Hipref ML & DL

## GPU Processors

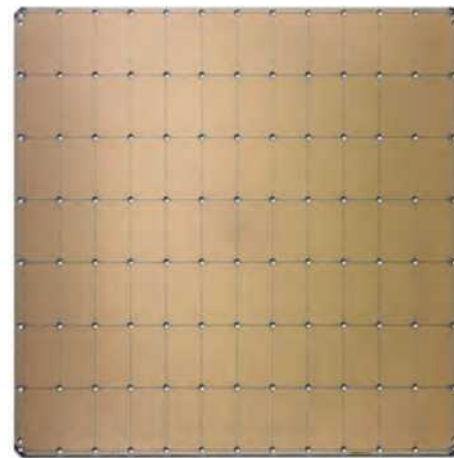
- NVidia GPU A100
- ...

## AI Processors

- Google TPU v4 (275 TFLOPS FP16 @ 170 W)
- Huawei IA Ascend 910 (320 TFLOPS FP16, 640 TOPS INT8 @ 310 W)
- ...
- Celebras Wafer Scale Engine (@ **20kW**)

## Others

- Processing-in-Memory (UpMem) → DRAM + DPU



**Cerebras WSE-2**  
2.6 Trillion Transistors  
46,225 mm<sup>2</sup> Silicon



**Largest GPU**  
54.2 Billion Transistors  
826 mm<sup>2</sup> Silicon



# Edge ML & DL

## GPU Processors

- Jetson Nano : ? TOPS @ 12.5W



## AI Processors

- Google Coral Edge TPU : 4 TOPS @ 2 W
  - MobileNet V2 model : 400 images per second
- Intel Movidius Myriad X VPU : 4 TOPS @ 1.5 W



[https://ai-benchmark.com/ranking\\_processors.html](https://ai-benchmark.com/ranking_processors.html)

[https://ai-benchmark.com/ranking\\_4\\_0\\_3.html](https://ai-benchmark.com/ranking_4_0_3.html)

[https://ai-benchmark.com/ranking\\_loT.html](https://ai-benchmark.com/ranking_loT.html)



# TinyML : AI @ Extreme Edge

ML and DL on low-power ( $\sim 1$  mW) MPUs and DSPs

Challenges for inference and for On Device Learning (ODL).

- Fragmented MCU market (heterogeneity)
  - ISAs (ARM Cortex M, RISC-V, ESP32, x86 ...)
  - w/o extensions (ARM CMSIS-NN ...)
    - required specific optimizations
- SRAM (64KB to 1.5MB), FlashRAM (128KB to 8MB), w/o FPU, w/o File System
- Cost by unit (< 10 USD)
- standard tools and frameworks
- benchmarks for comparison
- ...

# Tensorflow Lite Micro (aka TF Micro)



Tensorflow DNN for tiny MCU and DSP

*low power* CPU, w/o FPU, few RAM ...

Design

130 operations instead of 1400 for TF

default implementations

platform-optimized operator implementations (ie CMSIS-NN)

operator implementations can exploit multiple cores

list of TFL operations (no DAG)

optimization for memory and latency

quantisation (float32 → int8)

memory management (bin-packed arena)

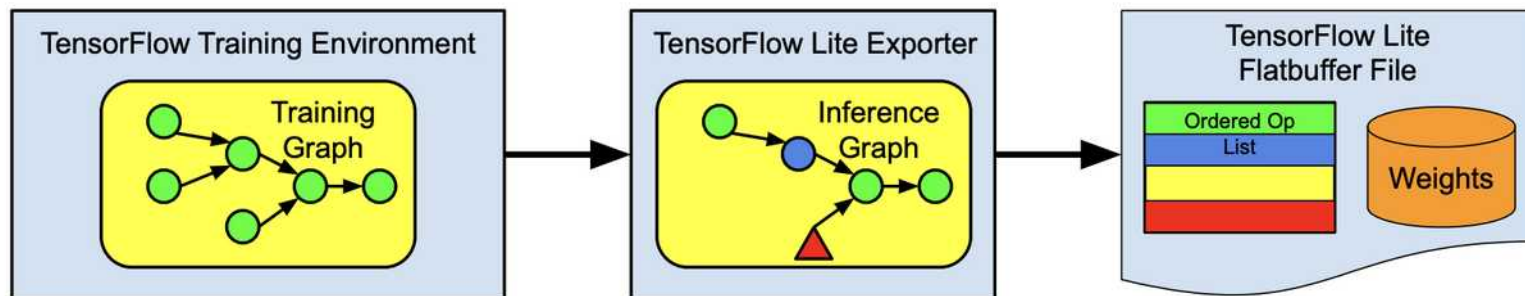
multi-tenancy (*multiple models*)

thread-safe

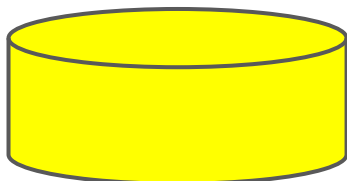
# TF Micro workflow



lib C (weight + code)



ie jupyter notebook on GPUs, collab ...



optimization (f32->i8 quantization, pruning, dag -> list ...)

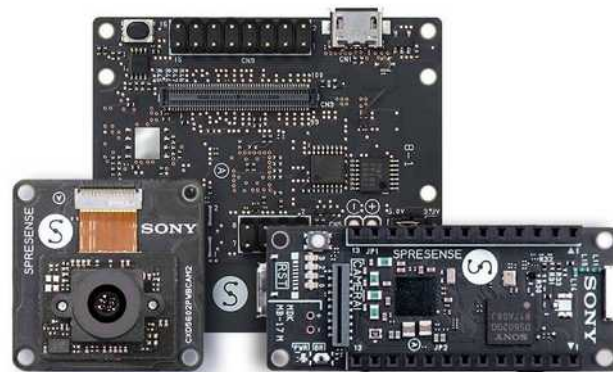
runtime

# Prototyping with Sony Spresense (Hardware)

Sony Spresense (CXD5602: Cortex M4F x 6 cores, 1.5 MB RAM, 8 MB Flash SPI)

- + ISX012 Image Sensor (CMOS 5 Mpixels, Y/C, RGB and JPEG, FOV 78°)
- + LoRaWAN modem (Microchip RN2483)
- + PIR motion sensor, SDCard, GNSS
- + 1 LiPo battery

- + Nordic PPK II  
profiling of *power consumption*  
for energy tuning





# (Fast) Prototyping with Sony Spresense and EI

Class inference by a Network

Transfert learning MobileNet V2 networks (VGG)  
trained on Edge Impulse (Keras Tensorflow)  
with ~3000 wildlife photos (from PNE's hunting cams)

Conversion to TF Lite (quantization f32→ i8)

2 models

- 10 Alps animals (61% accuracy)
- Wolf versus Other (89% accuracy)

Firmware

Arduino sketch with Tensorflow Micro\* library (and tflite models)



\* aka [TinyML](#)

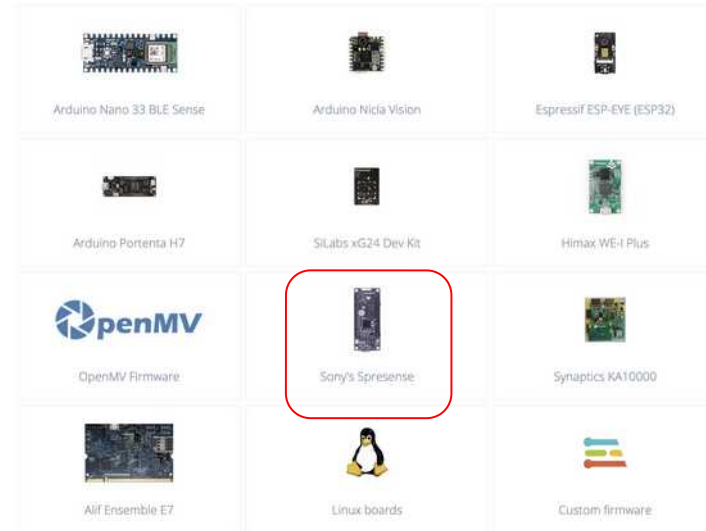
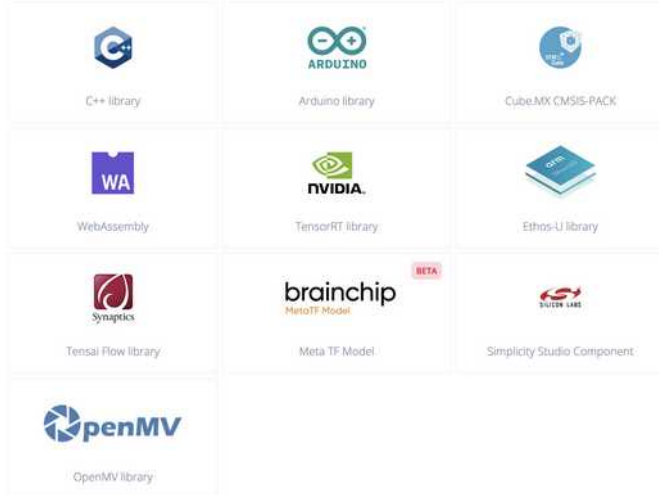


# EDGE IMPULSE

<https://docs.edgeimpulse.com/docs/>

Online platform for training and tuning models fitting TinyML constraints (Low power/low RAM MCU or DSP)

Generate (C/C++/WASM) TF Micro libs for most common MCU and eval boards (STM32, Sony, ESP32, Jetson ...)





## Workflow



Devices



Data sources



Data acquisition



Impulse design



Create impulse



Image



Transfer learning



EON Tuner



Retrain model



Live classification



Model testing



Performance calibration



Versioning



Deployment



## #1 - EON Tuner Primary

## Neural Network settings

## Training settings

Validation set size ②

20

%

## Neural network architecture

```

1
2 import math
3 from pathlib import Path
4 import tensorflow as tf
5 from tensorflow.keras.models import Model
6 from tensorflow.keras.layers import Sequential
7 from tensorflow.keras.layers import Dense, InputLayer, Dropout, Conv1D, Flatten, Reshape, MaxPooling1D,
8   BatchNormalization, Conv2D, GlobalMaxPooling2D, Lambda
9 from tensorflow.keras.optimizers import Adam, Adadelta
10 from tensorflow.keras.losses import categorical_crossentropy
11 sys.path.append('./resources/libraries')
12 import el_tensorflow.training
13
14 WEIGHTS_PATH = './transfer-learning-weights/edgeimpulse/MobileNetV2_0.05_96x96_color_bsize_64_lr_0.05
15   _epoch_574_val_loss_4.22.hdf5'
16
17 INPUT_SHAPE = (160, 160, 3)
18
19 base_model = tf.keras.applications.MobileNetV2(
20   input_shape = INPUT_SHAPE, alpha=0.05,
21   weights = WEIGHTS_PATH
22 )
23
24 base_model.trainable = False
25
26 model = Sequential()
27 model.add(InputLayer(input_shape=INPUT_SHAPE, name='x_input'))
28 # Don't include the base model's top layers

```

Start training

## Training output

## Model

Model version: ② Quantized (int8) ▾

## Last training performance (validation set)



ACCURACY

89.9%



LOSS

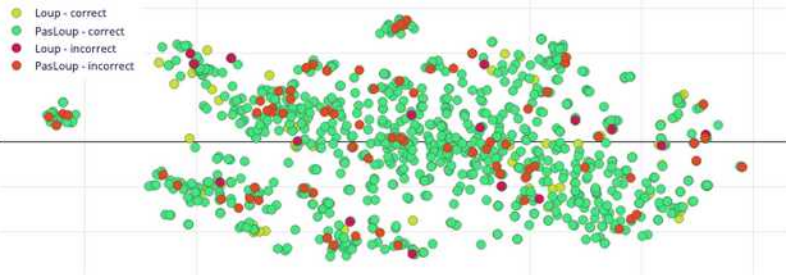
0,20

## Confusion matrix (validation set)

	LOUP	PASLOUP
LOUP	61.0%	39.0%
PASLOUP	7.1%	92.9%
F1 SCORE	0.53	0.94

## Data explorer (full training set) ②

- Loup - correct
- PasLoup - correct
- Loup - incorrect
- PasLoup - incorrect



## On-device performance ②



INFERRING TIME

270 ms.



PEAK RAM USAGE

670,4K



FLASH USAGE

173,2K

Dashboard

Devices

Data sources

Data acquisition

Impulse design

● Create impulse

● image

● Transfer learning

EON Tuner

Retrain model

Live classification

Model testing

Versioning

Deployment

## GETTING STARTED

Documentation

Forums



Enable EON™ Compiler

Same accuracy, up to 50% less memory. Open source.



# EDGE IMPULSE

## Available optimizations for Transfer learning

### Quantized (int8) ★

Currently selected

This optimization is recommended for best performance.

RAM USAGE

**670,4K**

LATENCY

**282 ms**

FLASH USAGE

**201,1K**

ACCURACY

**59.08%**

CONFUSION MATRIX ?

11.1	0	0	0	0	0	0	0	0	11.1	0	0	77.8
0	84.3	1.9	0	0	0	0.9	0	0.9	4.6	0.9	6.5	
0	3.9	64.1	0	0	1.0	1.9	0	3.9	2.9	0	22.3	
0	0	16.7	50	0	0	0	0	0	16.7	0	16.7	
0	0	20	0	40	0	0	0	0	20	20	0	
0	0	16.7	0	0	5.6	5.6	0	16.7	5.6	0	50	
0	2.2	0	0	0	0	40	0	2.2	13.3	0	42.2	
0	0	0	0	0	0	0	10.5	21.1	5.3	0	63.2	
0.9	0.9	3.6	0	0	0	2.7	0	43.6	2.7	0	45.5	
0	1.0	0	0	0	0	1.9	0	2.9	73.3	0	21.0	
0	0	5.9	0	0	0	0	0	0	0	76.5	17.6	

### Unoptimized (float32)

Click to select

RAM USAGE

**2,4M**

LATENCY

**1 044 ms**

FLASH USAGE

**414,6K**

ACCURACY

**61.28%**

CONFUSION MATRIX ?

22.2	0	0	0	0	0	0	0	0	11.1	0	0	66.7
0	85.2	1.9	0	0	0	0.9	0	1.9	3.7	0.9	5.6	
0	1.9	65.0	0	0	1.0	1.0	0	2.9	3.9	0	24.3	
0	0	16.7	50	0	0	0	0	0	16.7	0	16.7	
0	0	20	0	40	0	0	0	0	20	20	0	
0	0	11.1	0	0	5.6	5.6	0	16.7	5.6	0	55.6	
0	2.2	0	0	0	0	48.9	0	0	8.9	2.2	37.8	
0	0	0	0	0	0	0	10.5	26.3	5.3	0	57.9	
1.8	0	4.5	0	0	0	2.7	0	47.3	0.9	0	42.7	
0	1.0	1.0	0	0	0	1.9	0	3.8	74.3	0	18.1	
0	0	0	0	0	0	0	0	0	0	76.5	23.5	

# Accuracy and Uncertainty

 ACCURACY  
61.28%

	BLAIRE	CHAMC	CHEVR	CHIEN	ECUREI	LIEVRE	LOUP	MUSTE	RENAR	SANGL	VIDE	UNCER
BLAIREAU	22.2%	0%	0%	0%	0%	0%	0%	0%	11.1%	0%	0%	66.7%
CHAMOIS	0%	85.2%	1.9%	0%	0%	0%	0.9%	0%	1.9%	3.7%	0.9%	5.6%
CHEVREUIL	0%	1.9%	65.0%	0%	0%	1.0%	1.0%	2.9%	3.9%	0%	0%	24.3%
CHIEN	0%	0%	16.7%	50%	0%	0%	0%	0%	0%	16.7%	0%	16.7%
ECUREUIL	0%	0%	20%	0%	40%	0%	0%	0%	0%	20%	20%	0%
LIEVRE	0%	0%	11.1%	0%	0%	5.6%	5.6%	0%	16.7%	5.6%	0%	55.6%
LOUP	0%	2.2%	0%	0%	0%	0%	48.9%	0%	0%	8.9%	2.2%	37.8%
MUSTELIDE	0%	0%	0%	0%	0%	0%	0%	10.5%	26.3%	5.3%	0%	57.9%
RENARD	1.8%	0%	4.5%	0%	0%	0%	2.7%	0%	47.3%	0.9%	0%	42.7%
SANGLIER	0%	1.0%	1.0%	0%	0%	0%	1.9%	0%	3.8%	74.3%	0%	18.1%
VIDE	0%	0%	0%	0%	0%	0%	0%	0%	0%	0%	76.5%	23.5%
F1 SCORE	0.31	0.90	0.74	0.67	0.57	0.10	0.59	0.19	0.58	0.78	0.79	

 ACCURACY  
89.54%

	LOUP	PASLOUP	UNCERTAIN
LOUP	64.4%	28.9%	6.7%
PASLOUP	4.8%	91.8%	3.4%
F1 SCORE	0.59	0.94	

# TF Lite Model sizes



Enable EON™ Compiler

Same accuracy, up to 50% less memory. Open source.



## Available optimizations for Transfer learning

### Quantized (int8)



Currently selected

This optimization is recommended for best performance.

### Unoptimized (float32)

Click to select

RAM USAGE

777,6K

LATENCY

270 ms

CONFUSION MATRIX

53.3	33.3	13.3
3.8	92	4.2

FLASH USAG

237,5K

ACCURACY

88.81%

RAM USAGE

2,9M

LATENCY

1 001 ms

CONFUSION MATRIX

64.4	28.9	6.7
4.8	91.8	3.4

FLASH USAG

343,4K

ACCURACY

89.54%

Estimate for Sony Spresense (Cortex-M4F 156MHz)



Enable EON™ Compiler

Same accuracy, up to 50% less memory. Open source.



## Available optimizations for Transfer learning

### Quantized (int8)



Currently selected

This optimization is recommended for best performance.

### Unoptimized (float32)

Click to select

RAM USAGE

670,4K

LATENCY

270 ms

CONFUSION MATRIX

53.3	33.3	13.3
3.8	92	4.2

FLASH USAG

173,2K

ACCURACY

88.81%

RAM USAGE

2,4M

LATENCY

1 001 ms

CONFUSION MATRIX

64.4	28.9	6.7
4.8	91.8	3.4

FLASH USAG

303,3K

ACCURACY

89.54%

Estimate for Sony Spresense (Cortex-M4F 156MHz)

Default : tflite convertor

# Hunting Cam Prototype





# Démonstration Architecture



# Field test case(s)

- Parc National des Ecrins (Entraigues, secteur Valbonnais)
  - Fauna, Users (October 2022)
    - [Kerlink iStation Gateway @ Maison du Parc](#)
  - PGHM Le Versoud
    - “outdoor” security (to be confirmed)
- Next
  - PIA Equipex Terra-Forma (2021-2029)
    - ~50 labs, Zones Atelier
  - Deep Faune Database
    - ~1.5 millions images
  - Internet of Sounds ...



# Next prototypes

Brand new MCUs for Embedded (Very Low Power) AI

M5 Stack / ESP32 Cam

Maix Speed (YOLO)

Arduino Nicla Vision (STM32 H7)

RPI Pico

Image sensors

Sony Spresense HDR

OmniVision sensors (Night / Day)

IR Illuminator

**Thermal 24x32 Image Sensor (MLX90640)**

<https://github.com/CampusIoT/tutorial/tree/master/wiotalterminal/examples/MLX90640>



# Seeestudio SenseCap A1101

TinyML @ STM32WLEx (@48 MHz with 256 Kbytes of Flash memory, 64 Kbytes of SRAM)

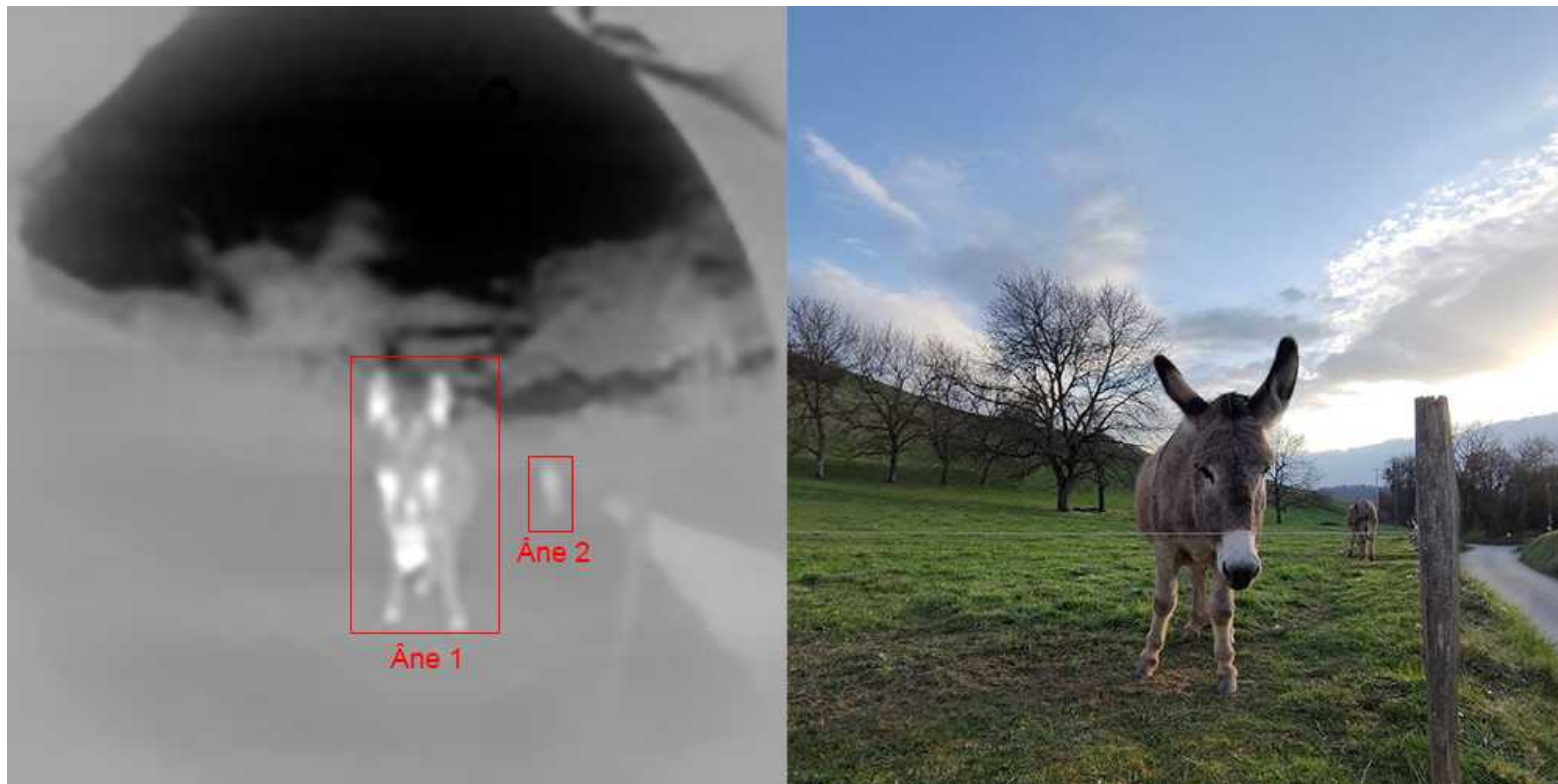
Communication: BLE, LoRa, IR-FHSS, Mioty, CLS (constellation SatIoT Kinéis)



# MLX90640 (24x32) thermal sensor



Lynred 80x80 IR sensor in GAPPoc B



# More information

<https://gitlab.com/wildcount>

<https://gitlab.com/wildcount/firmware-spresense-arduino>

<https://campusiot.github.io/>

<http://edge-intelligence.imag.fr>

<https://terra-forma.cnrs.fr/>

**Prendre le pouls de la Terre**

<https://www.ecrins-parcnational.fr/>

