Wildcount

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

Contact: Georges QUENOT, Didier DONSEZ
Université Grenoble Alpes
LIG MRIM & ERODS

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

Context : Eco-biology

2 Renards !

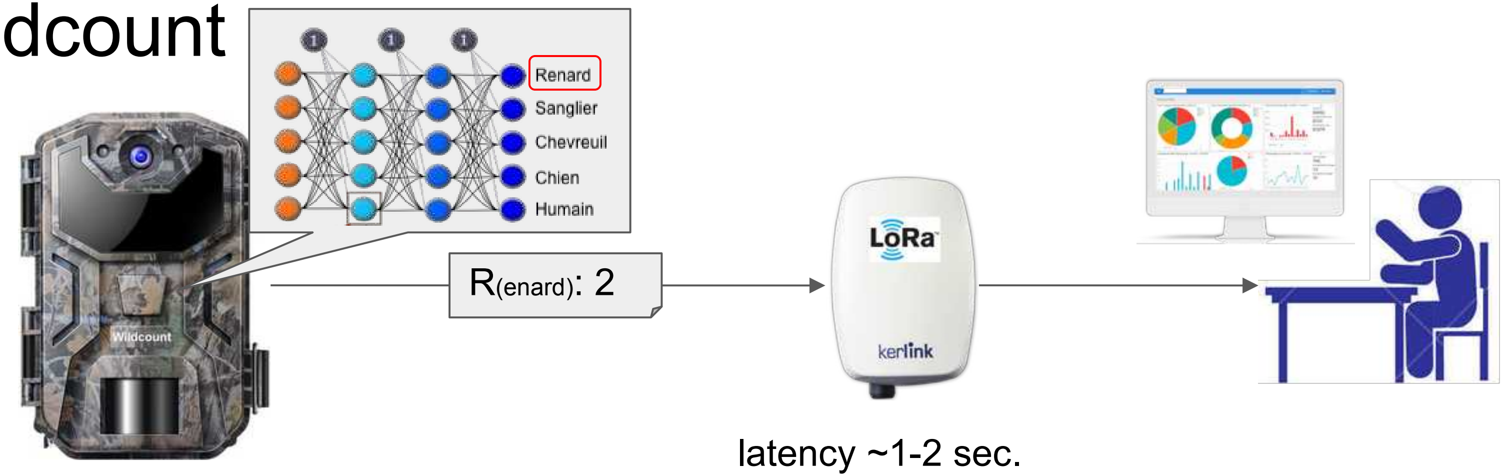
latency ~1-2 months

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

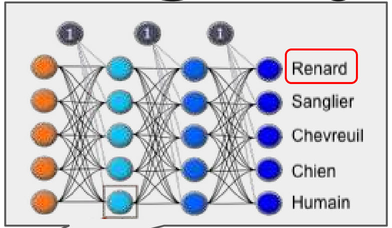
Remark: Indexing @ Home: human or automatic ([DeepFaune])

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

Wildcount


Renard
Sanglier
Chevreuil
Chien
Humain
Wildcount
R(enard): 2
LoRa
kerlink
latency ~1-2 sec.

Wildcount : AI @ Edge



R(enard): 2

latency ~1-2 sec.

Fix embedded inferences

latency ~1-2

Wildcount’s Aims

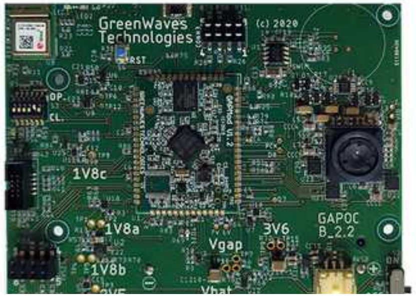
- Internet of Things endpoint for
  - counting of wild fauna in natural and protected areas
  - counting of (anonymous) users of natural and protected areas
- Technical requirements
  - Low power / Long battery lifetime
  - Long range transmissions of counters (in “real time”)
  - Transmissions in white zoom (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 LoRa

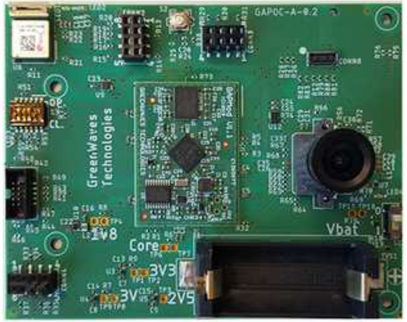
Prototyping with Greenwaves Gap8

GAPPoc (Greenwaves Technologies)

- Visible-light sensor (GAPPoc A) *grey level*
- 80x80 IR sensor (GAPPoc B)
- MobileNet V1-V2 networks (Grid5000)
    - trained with 3000 wildlife photos (from hunting cam)

\+ LoRaWAN modem (Microchip RN2483)
\+ PIR motion sensor, SDCard
\+ RTC, TH, Magnetometer, Accelerometer, GNSS (option)
\+ 2 x NCR18650 battery pack (6600 mAh)
\+ X-NUCLEO-LPM01A Powershield (*energy consumption*)





Prototyping with Greenwaves Gap8

3 units with “IP67” enclosure

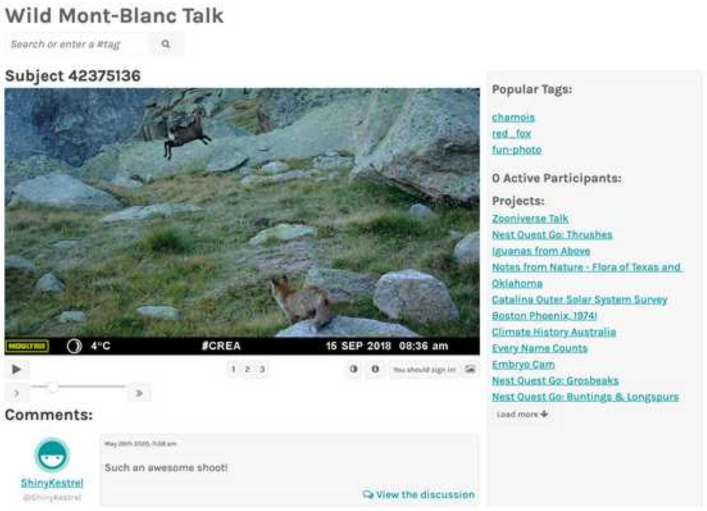
Firmware

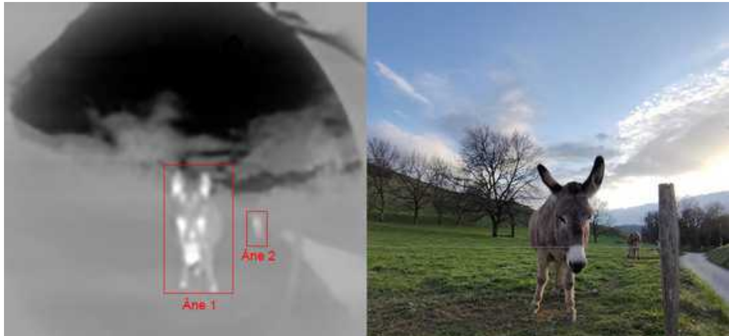
- PulpOS/FreeRTOS (SDK Greenwaves)
- Google MobileNet
    - dégradé : 2 couches en moins
    - trained with 3000 wildlife photos (from hunting cam)
    - day and night (IR flash)

Labeled image databases for training

Fauna
- Crea Mont Blanc Zooinverse
- Bing images (but too perfect !)
- Photothèque (Parc des Ecrins)
- 30 hunting cameras (Parc des Ecrins)
- 80x80 IR images

Users
- ~~pièges photo de bivouac (comptage de tentes)~~
- ~~ski de randonnée dans la combe de Laurichard~~

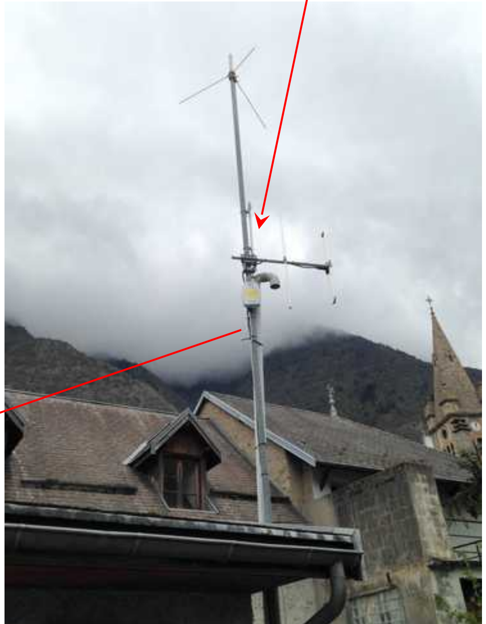




Demonstration Architecture

Field test case(s)

- Parc National des Ecrins (Entraigues, secteur Valbonnais)
  - Fauna, Users (July ~~2021~~ 2022)
    - Kerlink iStation Gateway @ Maison du Parc
  - PGHM Le Versoud
    - “outdoor” security (to be confirmed)
- Next
  - Ecole de Physique des Houches ?
  - ONF Chartreuse ?
  - PIA Equipex Terra-Forma (2021-2029)
    - ~50 labs, Zones Atelier
  - Deep Faune Database
    - ~1.5 millions images



Next prototypes

- Brand new MCUs for Embedded (Very Low Power) AI
  - M5 Stack / ESP32 Cam
  - Maix Speed (YOLO)
  - Sony SPresence (FOMO)
  - Arduino Nicla Vision (STM32 H7)
  - RPI Pico
- Image sensors
  - OmniVision sensors (Night / Day)
  - IR Illuminator
  - Thermal Image Sensor (MLX90640)

    https://github.com/CampusIoT/tutorial/tree/master/wioterminal/examples/MLX90640

More information

https://gitlab.com/wildcount

https://campusiot.github.io/

http://edge-intelligence.imag.fr

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

Prendre le pouls de la Terre

Bonus Track

IoT Infrastructure and AI

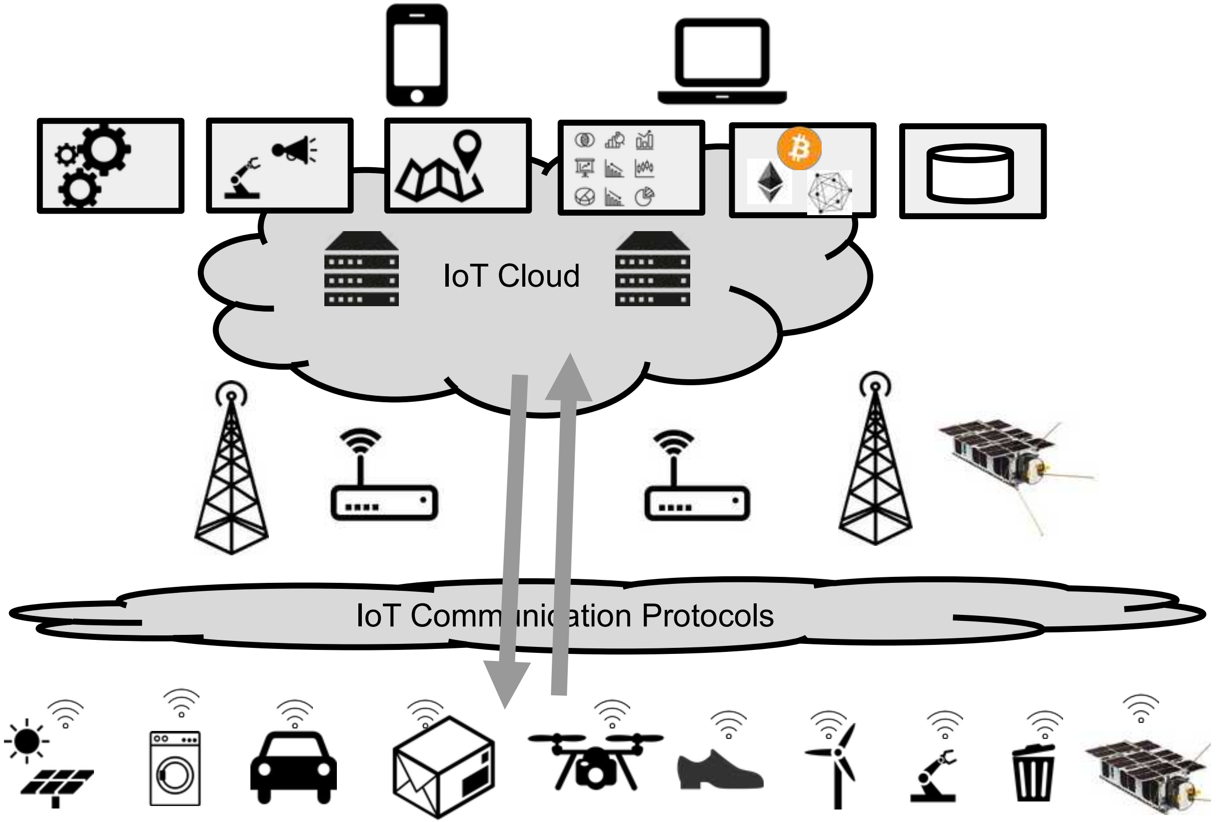
IoT Applications

Cloud infrastructure
(public, private)

Fog/Edge Computing

Communications
- wired/wireless
- IP / No IP
- licensed/free bands

Connected Things
(sensors & actuators)



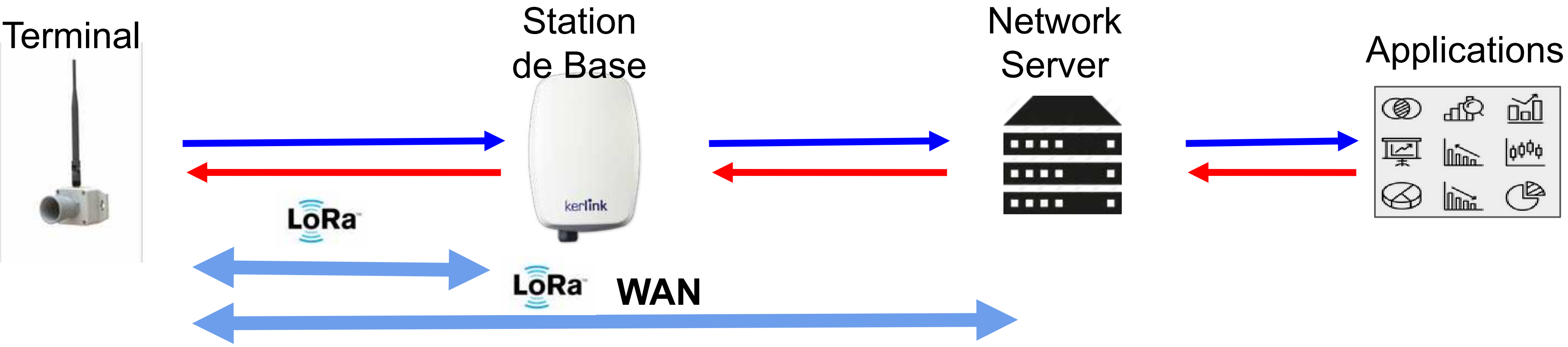
AI @ Cloud

AI @ Edge

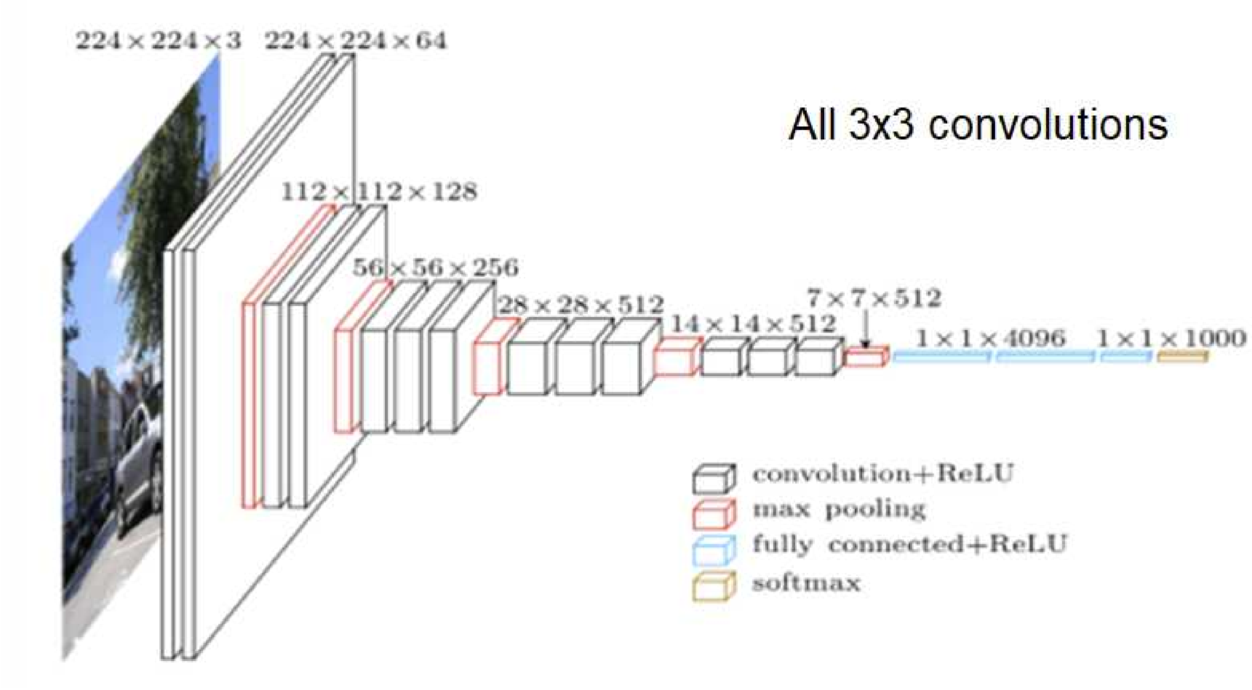
AI @ Extreme
Edge

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)



VGG Network (very deep)



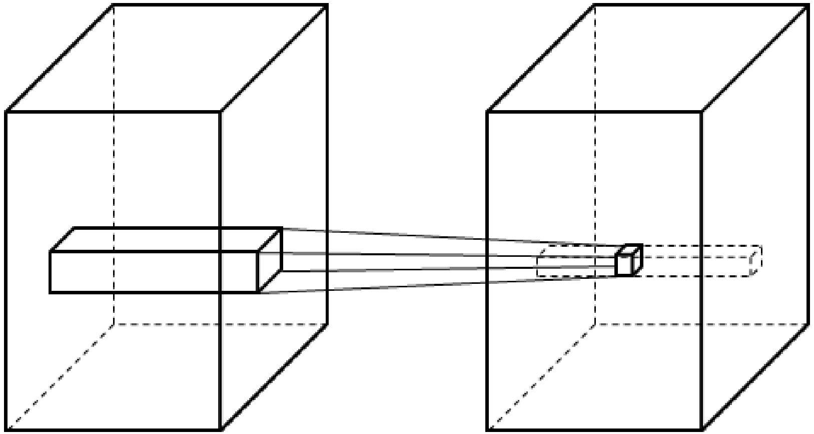
All 3x3 convolutions

Simonyan and Zisserman, Andrew: *Very Deep Convolutional Networks for Large-Scale Image Recognition*, CVPR 2014.

MobileNet

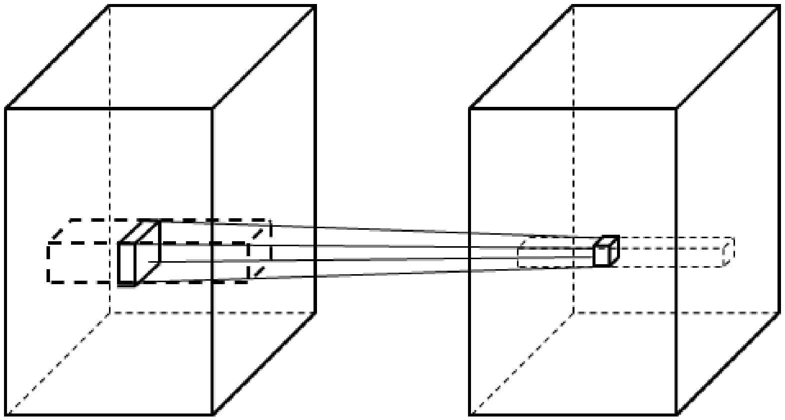
- Variant of VGG with less layers
- Use depth-wise separable convolutions, sequences of:
  - depth-wise (within map) convolutions
  - pixel-wise (1x1) convolutions, actually matrix-vector products
- First layer with “full” convolutions, 13 separable convolutions (two layers each), global pooling, a final fully connected layer, and a softmax layer
- ReLU as the activation function everywhere except for the last layer
- Batch Normalization everywhere (the main trick for training very deep networks, kind of mean and variance normalization at the batch level)
- Resolution changes with strides

“Full” convolutional layers



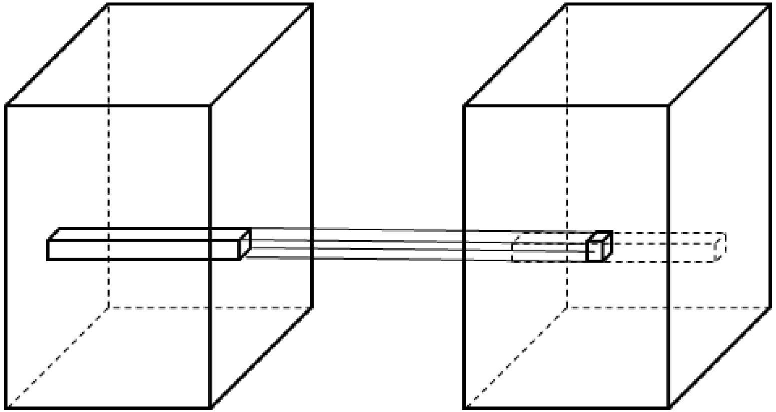
- Each map point is connected to all maps points of a fixed size neighborhood in the previous layer
- 4D kernel

Depth-wise convolutional layers



- Each map point is connected the maps points of a fixed size neighborhood of the same map in the previous layer
- 3D kernel : within map convolutions, 2D times #maps
- Same number of maps in input and output

Pixel-wise convolutional layers



- Each map point is connected to all maps points corresponding to the same pixel in the previous layer
- 2D kernel per pixel matrix-vector multiplication
- Parallel “all to all” between pixel vectors.

Depth-wise separated convolutions

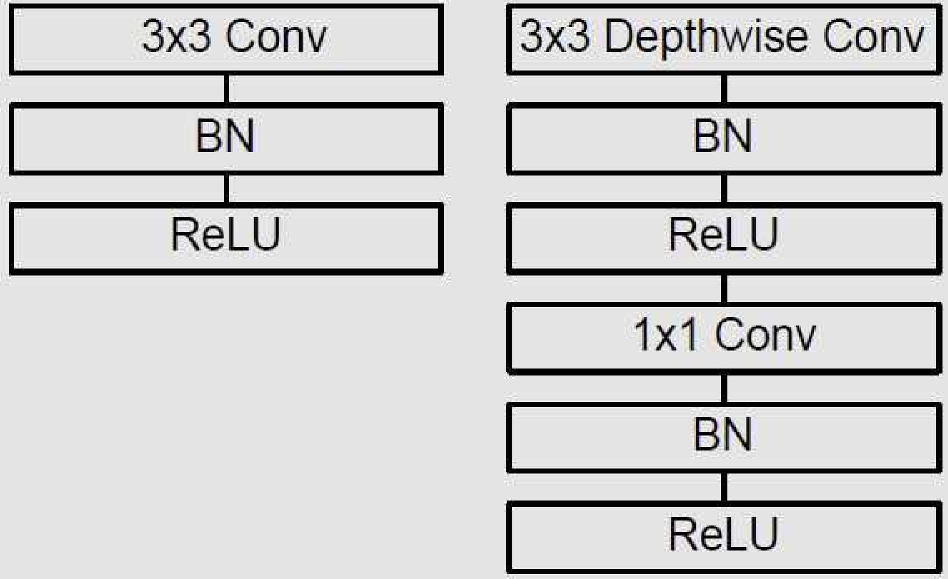


Figure 3. Left: Standard convolutional layer with batchnorm and ReLU. Right: Depthwise Separable convolutions with Depthwise and Pointwise layers followed by batchnorm and ReLU.

MobileNet base architecture

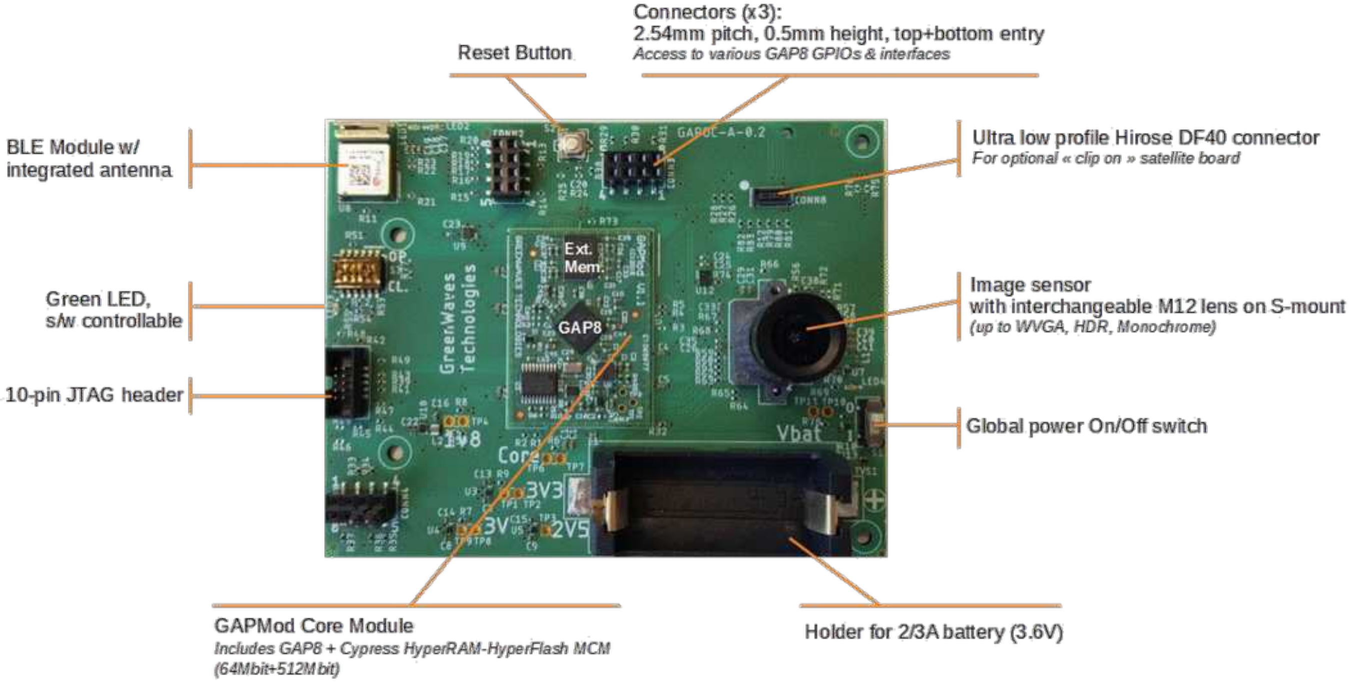
Table 1. MobileNet Body Architecture

Type / Stride	Filter Shape	Input Size
Conv / s2	$3 \times 3 \times 3 \times 32$	$224 \times 224 \times 3$
Conv dw / s1	$3 \times 3 \times 32$ dw	$112 \times 112 \times 32$
Conv / s1	$1 \times 1 \times 32 \times 64$	$112 \times 112 \times 32$
Conv dw / s2	$3 \times 3 \times 64$ dw	$112 \times 112 \times 64$
Conv / s1	$1 \times 1 \times 64 \times 128$	$56 \times 56 \times 64$
Conv dw / s1	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 128$	$56 \times 56 \times 128$
Conv dw / s2	$3 \times 3 \times 128$ dw	$56 \times 56 \times 128$
Conv / s1	$1 \times 1 \times 128 \times 256$	$28 \times 28 \times 128$
Conv dw / s1	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 256$	$28 \times 28 \times 256$
Conv dw / s2	$3 \times 3 \times 256$ dw	$28 \times 28 \times 256$
Conv / s1	$1 \times 1 \times 256 \times 512$	$14 \times 14 \times 256$
5× Conv dw / s1 Conv / s1	$3 \times 3 \times 512$ dw $1 \times 1 \times 512 \times 512$	$14 \times 14 \times 512$ $14 \times 14 \times 512$
Conv dw / s2	$3 \times 3 \times 512$ dw	$14 \times 14 \times 512$
Conv / s1	$1 \times 1 \times 512 \times 1024$	$7 \times 7 \times 512$
Conv dw / s2	$3 \times 3 \times 1024$ dw	$7 \times 7 \times 1024$
Conv / s1	$1 \times 1 \times 1024 \times 1024$	$7 \times 7 \times 1024$
Avg Pool / s1	Pool 7×7	$7 \times 7 \times 1024$
FC / s1	1024×1000	$1 \times 1 \times 1024$
Softmax / s1	Classifier	$1 \times 1 \times 1000$

MobileNet

- Small VGG with depth-wise separable convolutions
- Two global hyper-parameters
  - α : scale for the “thickness” of the stack of feature maps
  - ρ : scale of the input images
- Both applied on all layers (except first and last)
- Both the number of parameters and the number of operations (strongly correlated with the processing time) depends approximately quadratically with both α and ρ, allowing for a wide range of accuracy versus speed and memory compromises
- The number of layers could be modified also but is less good in the compromise

Greenwave GapPOC boards



GapPOC-B
Lynred ThermEye infrared camera