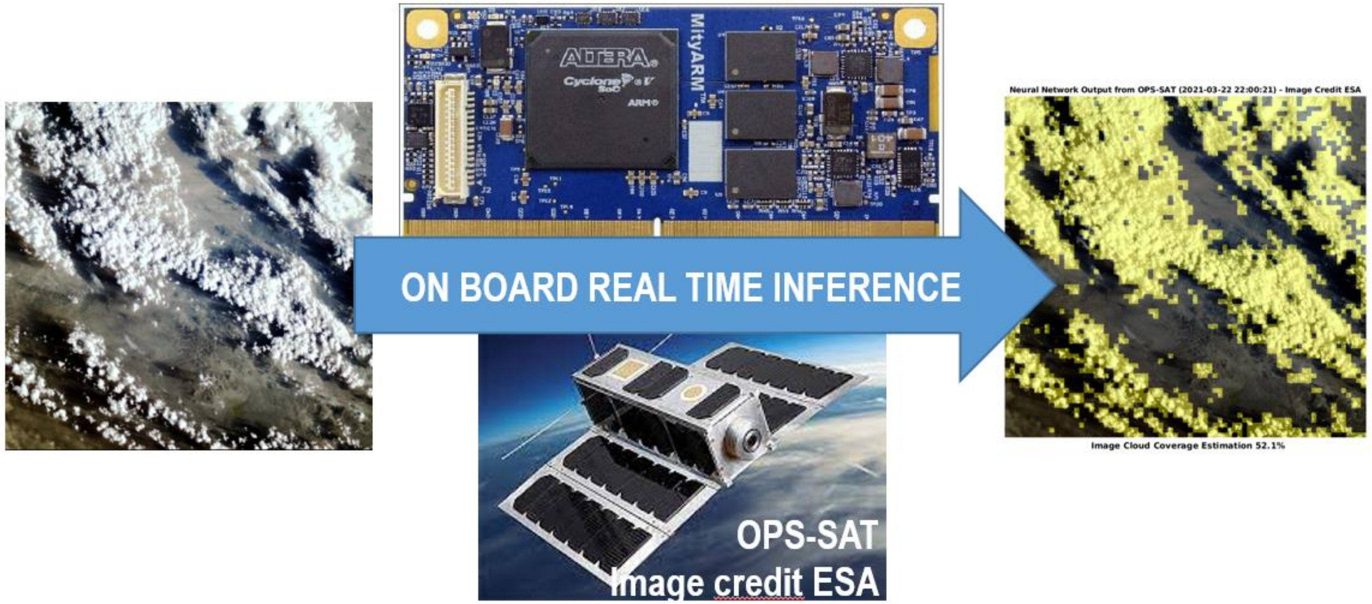
In-flight OPS-SAT images processing by AI

image credit: ESA

Agenda.

- Experiment context
- OPS-SAT use case
- Deployment workflow on SoC
- In flight inferences results
- Conclusion and perspectives



FRENCH INSTITUTES OF TECHNOLOGY

About us.

Project team “Autonomous and Reactive Imagery Chain” (CIAR) from French Institute Research & Technology Saint-Exupéry

Main Work Package:

- Image data base
- ANNs coding
- Deployment on electronic chips
- Performance on use cases

AI in Embedded Systems

FRENCH INSTITUTES OF TECHNOLOGY

OPS-SAT use case.

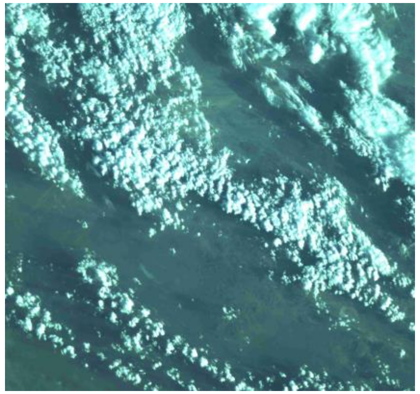
Goal:

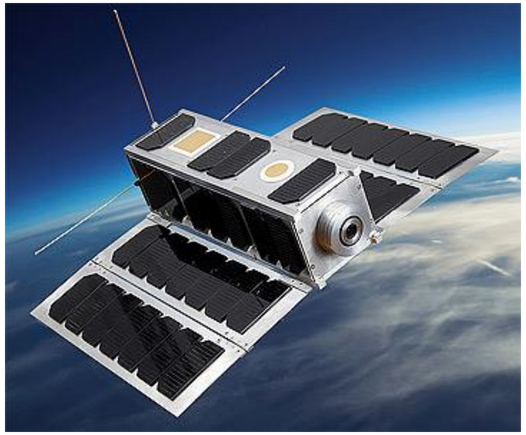
- RGB Cloud Segmentation in real time
- with “tiny” ANNs on Cyclone V SoC FPGA

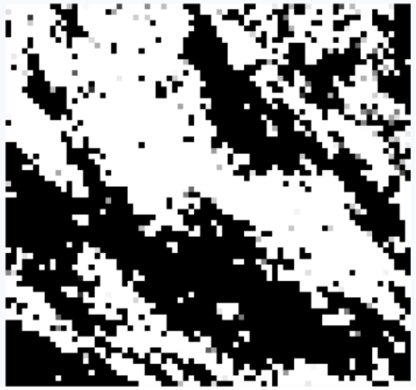
Two firsts in Europe:

- Deep learning on-board processing of an image on FPGA
- Remote updating, from an ESA ground station, of an ANN on board the OPS-SAT satellite

Full Resolution OPS-SAT Image 2044*1932pix







Onboard Segmentation Map 73*69pix From the FPGA

image credit: ESA

FRENCH INSTITUTES OF TECHNOLOGY

OPS-SAT images caracteristics

- **OPS-SAT IMS-100 Camera**
  - Based on ST200 star tracker
  - Radiation tolerance up to 12.5 kRad
  - Size 30 x 33.5 x 41.4 mm^3
  - Mass 48.8g, power 650 mW
  - Bayer sensor with pixel size : 2,2µm
  - Focal length 25mm, Diagonal FOV : 21°

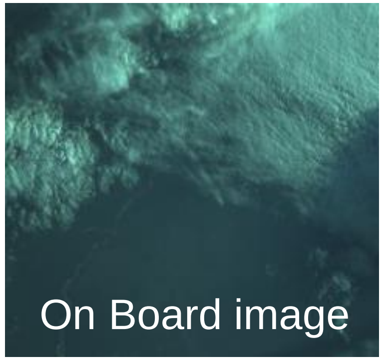
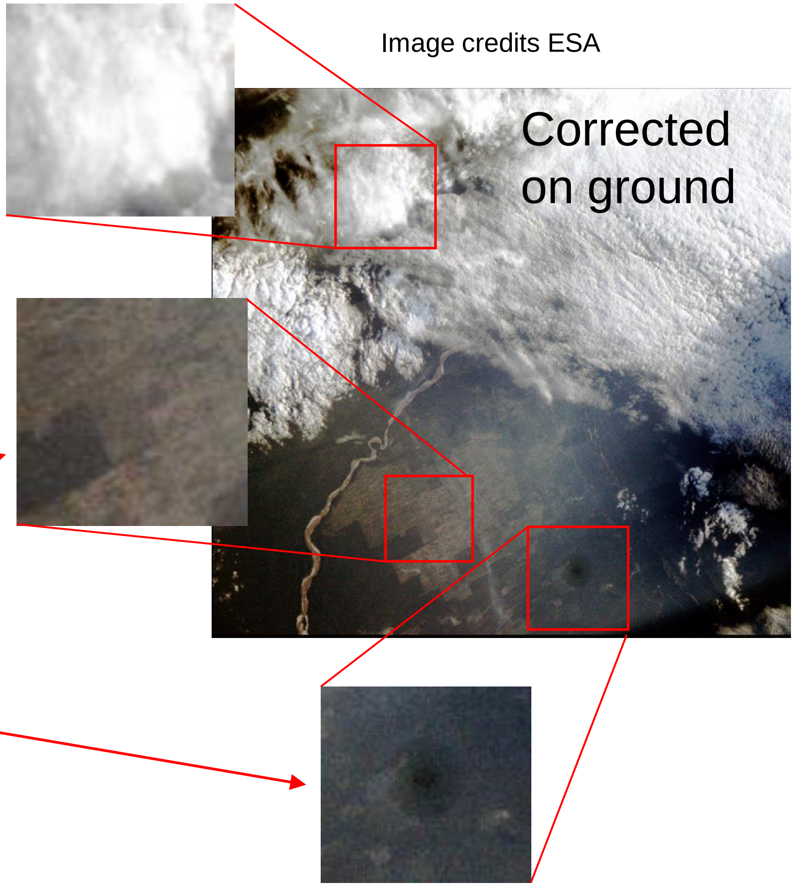


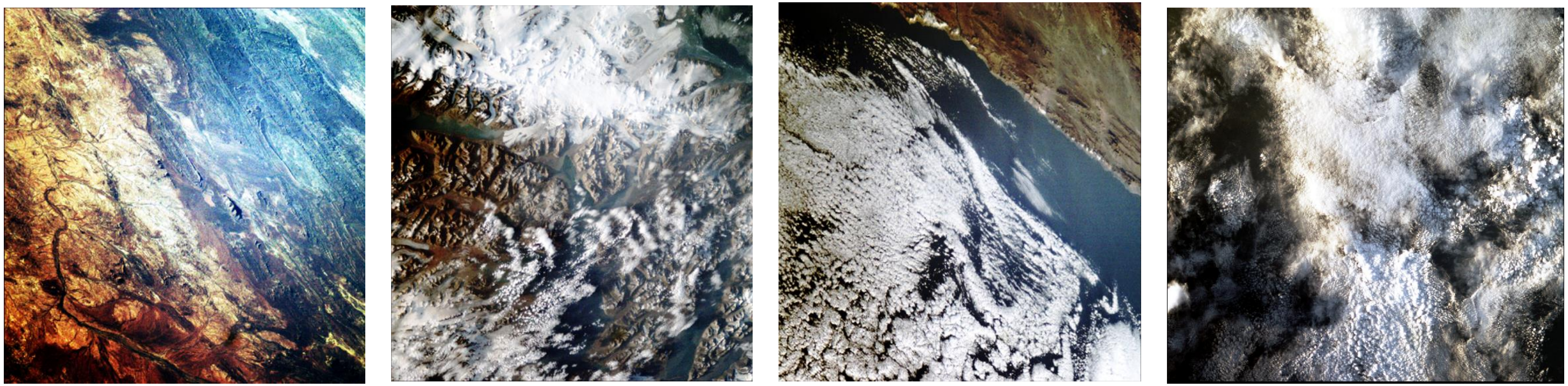
Image credits ESA

- **OPS-SAT image**
  - Size 2048 x 1944 x 8bits/pixel
  - Ground sampling : 53m @altitude 600km
  - Raw radiometry :
    - Saturations for high radiometric level areas
    - Poor SNR on scenes at low Sun elevation angle
    - Sensor defects not corrected (dust on optics)
  - No radiometric correction applied



OPS-SAT images dataset

- **Training data :** 19 training images from the first OPS-SAT acquisition campaign on August 2020



Test data : 23 test images from theOPS-SAT acquisition campaign on March 2021

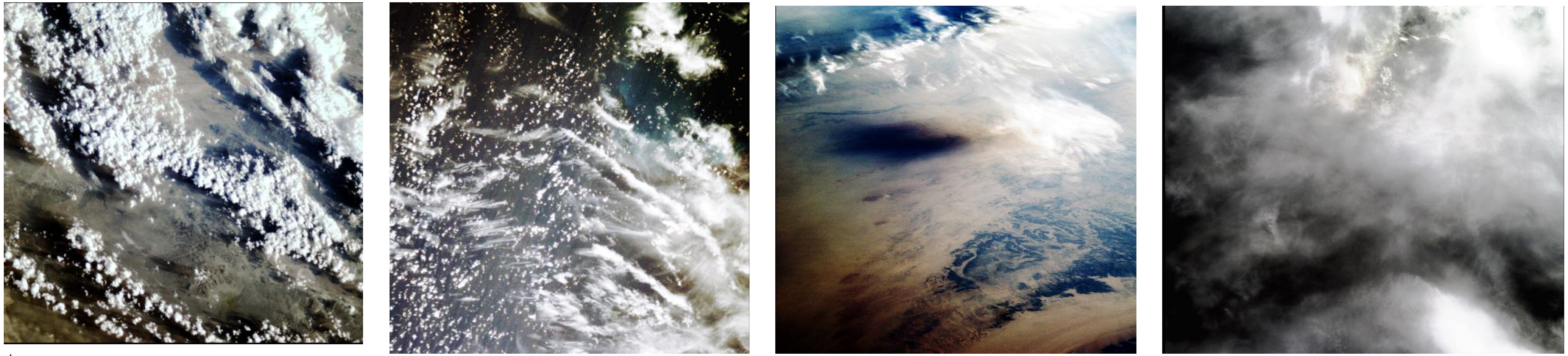


Image credits ESA

FRENCH INSTITUTES OF TECHNOLOGY

OPS-SAT image labelling

Semi manual labelling by an image processing expert :

- Binary cloud mask cloud/no_cloud labelling at pixel level to produce ground truths for training and performance metrics assessment
- Automatic 28x28 patching and patches annotation wrt. cloud coverage from 0% to 100% by step of 10% used for classification algorithms training (CNN, SNN)

OPS-SAT RGB image

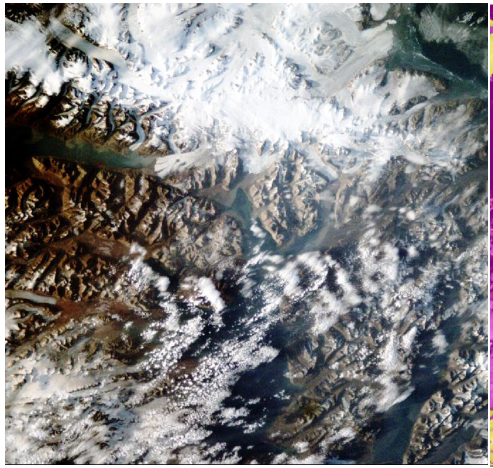
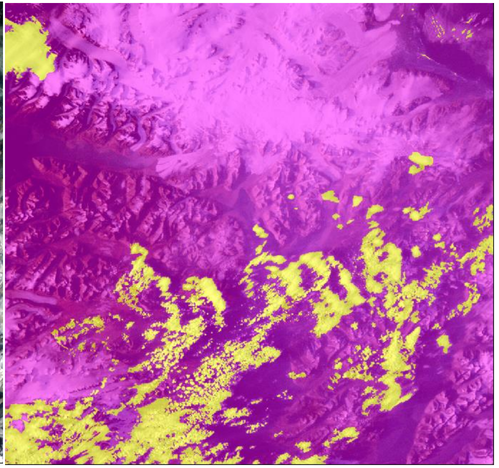


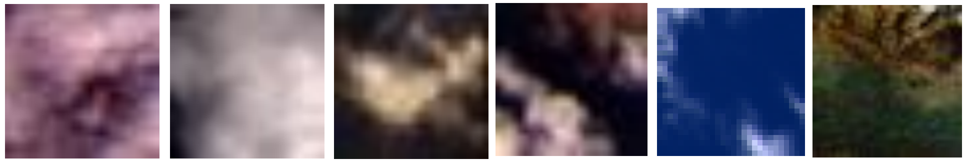
Image credits ESA

Binary Cloud/No_cloud ground truth



Patching 28x28

Examples of cloud coverage patch labelling



Patch Label	100%	90%	50%	20%	10%	0%

Labelling rules

- Drop shadows over clouds are considered as cloud
- Transparent clouds (cirrus, haze) showing visible ground features are annotated as no cloud
- Snowy and ice are annotated as no cloud

FRENCH INSTITUTES OF TECHNOLOGY

OPS-SAT « tiny » Neural Networks.

Formal Neural Network CNN

- LeNet-5 architecture: C5.3.1,R,P2,C5.5.1,R,P2,D10,R,D2,SM
- 1440 parameters
- Classification of 28*28 patchs (5037 patches in full 2048*1944 image)
- « Best paper » of conference On Board Payload Data Compression (OBPDC 2020)

Full Spiking Neural Network

- LeNet-like architecture without pooling : C6.3.1,R,C6.3.1,R,D10,R,D2,SM
- 2666 parameters
- Classification of 28*28 patchs (5037 patches in full 2048*1944 image)
- PhD thesis at university of Cote d'Azur (December 2020)

Full convolutional Neural Network FCN

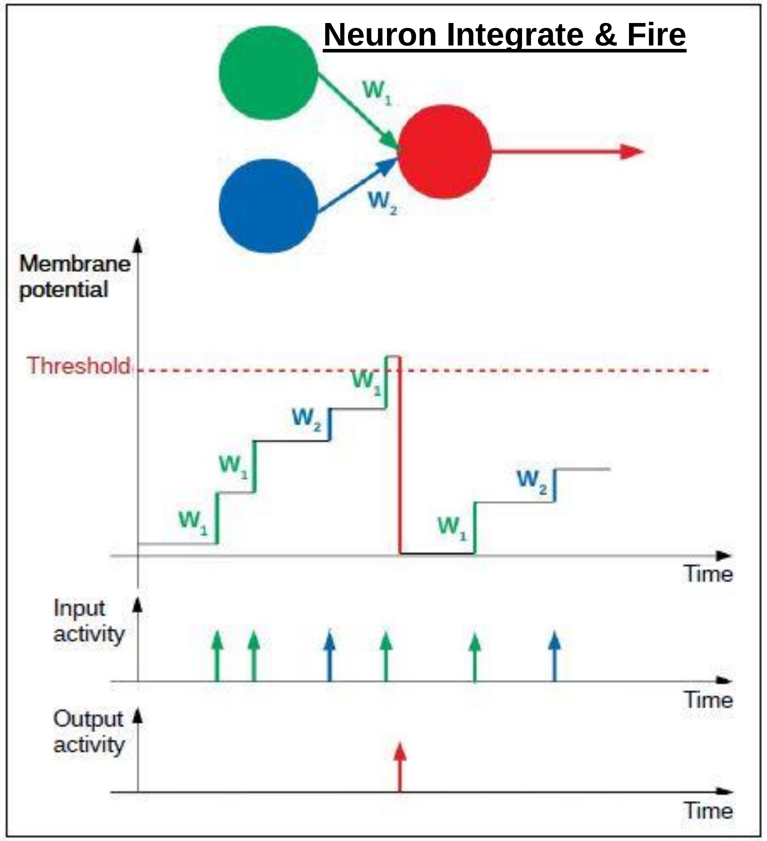
- "Tiny" semantic segmentation*: C3.5.1,R,P2,C5.5.1,R,P2,C1.1.1,SG
- 644 parameters (No Fully Connected)
- Output of the 28*28 patch inference is a 4*4 features map
- Publication selected at International Conference on Computer Vision (ICCV 2019)

FRENCH INSTITUTES OF TECHNOLOGY

Zoom on Spiking Neural Networks.

Spiking neurons model:

- Inspired from the event-driven paradigm of biological brain
- Information is processed in the spiking domain
- Integrate & Fire (IF) neuron model*
  - weighted spikes counted in an accumulator => replace MUL-ACC of formal ANN
  - output spike triggered when threshold raised => replace activation function



* Refer to “Design space exploration of hardware spiking neurons for embedded artificial intelligence” in Elsevier Neural Networks, January 2020

FRENCH INSTITUTES OF TECHNOLOGY

The Zoetrope Genetic Programming (ZGP) algorithm

Principle

- Evolutionary algorithm
- Individuals = mathematical expressions combining input variables and constants
- Evolve through generations via genetic operators (mutation and crossover)
- Survival of the "fittest" individuals with regard to the input data

Advantages:

- Frugality in terms of training data
- Low computational time
- Avoids overgrown and complex models
- Interpretability of models
- Easy deployment and portability on embedded systems

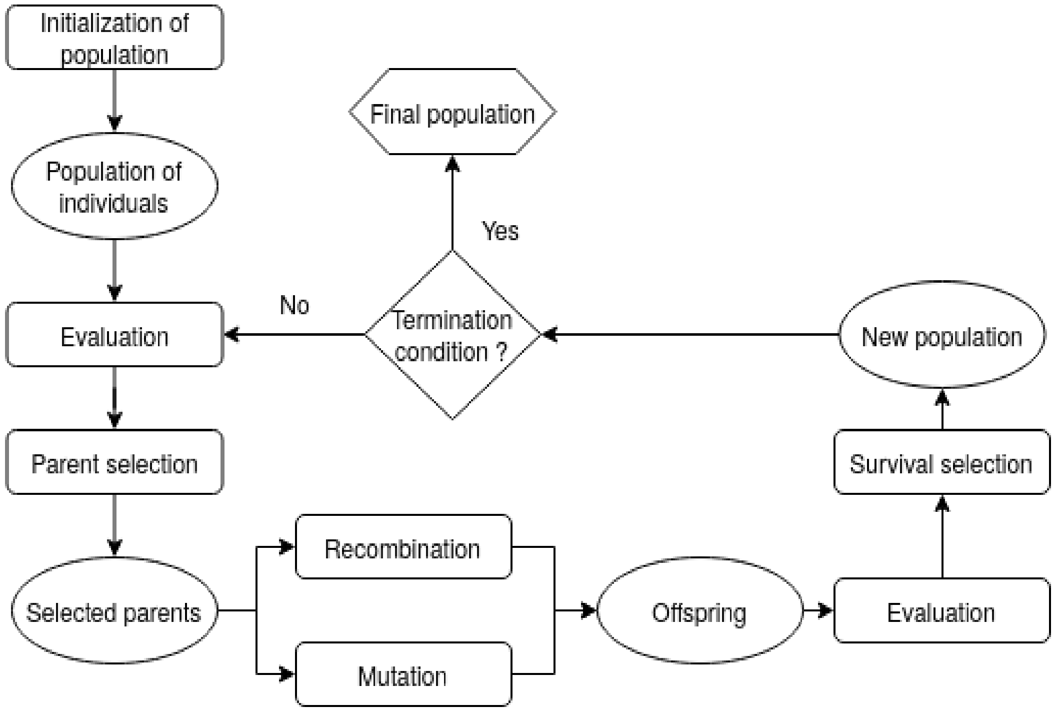


Figure: General evolutionary algorithm process

Two strategies :

- ZGP alone: trained on a subset of pixels from each image from the training set (10000 pixels in total).
- FCN+ZGP: ZGP trained on the segmentation maps resulting from a FCN

A. Boisbunon et al. Zoetrope genetic programming for regression. In 2021 Genetic and Evolutionary Computation Conference(GECCO'21), 2021

ANNs deployment on FPGA.

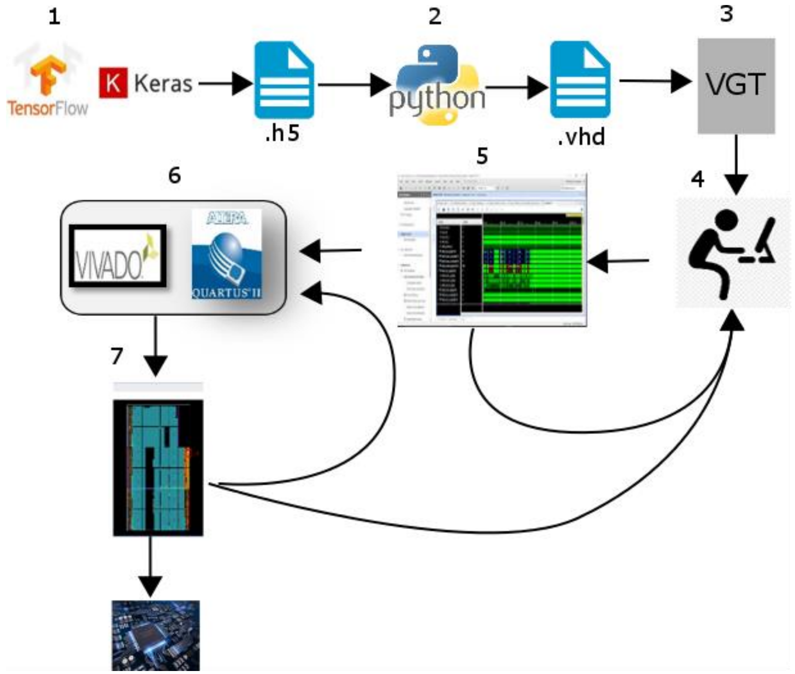
HW target:

MitySOM- -5CSx combines :

- ✓ dual hard-core Cortex-A9
- ✓ 110K Logical Elements
- ✓ one Cyclone V System on Chip
- ✓ 2GB of DDR3 CPU/FPGA RAM
- ✓ 512MB of DDR3 FPGA RAM
- ✓ 48MB of QSPI NOR Flash

Workflow:

1. **ANN train & test**
2. **Weight/biais conversion**
3. **HDL code generation with VGT**
4. **Code modification**
5. **RTL simulation & verification**
   - If an error occurs, return to step 4
6. **Design intégration & test**
7. **Implementation**
   - If an error occurs, return to step 6

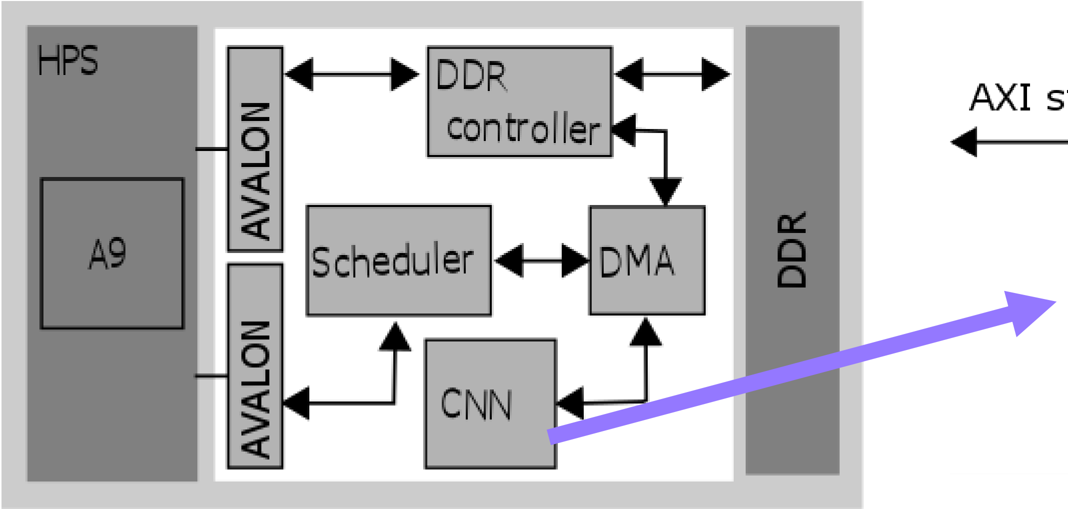


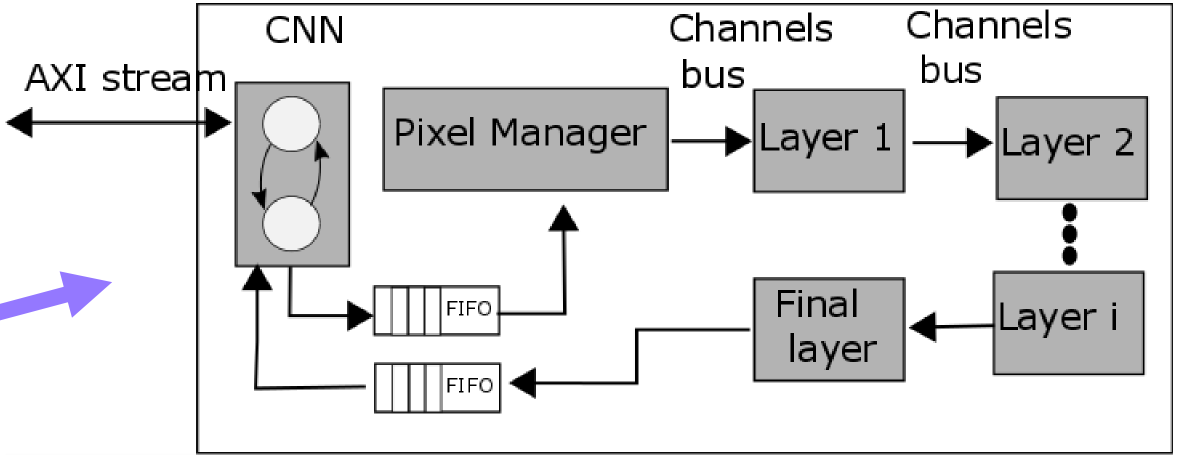
FRENCH INSTITUTES OF TECHNOLOGY

ANNs deployment on FPGA.

Design constraints:

- Pipelined inferences
- Cyclone-V FPGA capabilities





- ➢ Pre/post-processing on core A9
- ➢ Scheduler controlled by software to program the DMA
- ➢ DMA to move data from/to DDR
- ➢ « CNN » main IP to run image inference
- ➢ Communication : Avalon / AXI4 /AXI stream

- ➢ Final state machine manages the AXI stream IF
- ➢ Pixel manager which performs pixel re-organisation
- ➢ Layers interconnection with enabling signals
- ➢ Final layer stores results into FIFO

FRENCH INSTITUTES OF TECHNOLOGY

ANNs deployment on FPGA.

- Inférence consumption < 1,8 W (Quartus estimator)

ANN architecture	FCN 8.11	CNN 8.9	SNN 8.8
Number of parameters	644	1440	2666
Logic cells occupation %	92%	69%	31%
Zeroded W&B by quantification	0/614	809 / 1440	15 / 2666
Clock frequency (MHz)	100	100	100
Latency per 28x28 patch (µs) ↓	24.99	24.84	312.0
Power estimation (mW)	1750	1599	1062

FRENCH INSTITUTES OF TECHNOLOGY

Preliminary results: Upload process.

Software
Java code and libraries:
4 x 826kB=3.3MB IPK
Upload time: 4 x 1 min 8 sec.

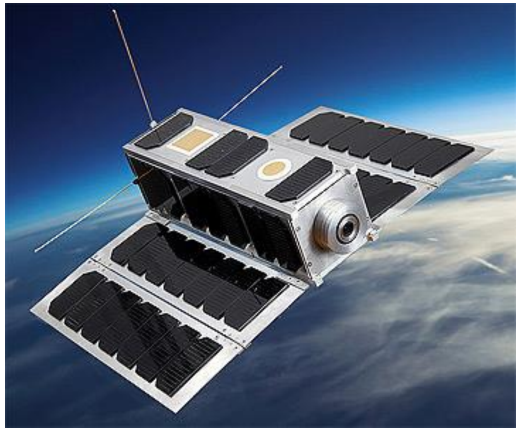


image credit: ESA

Band S : Average measured data rate 12,818kB/s.

Firmware:
FPGA Bitstream, preloader, bootloader
One 1.5MB IPK
Upload time: 1 min 46 sec.

image credit: ESA

fit | FRENCH INSTITUTES OF TECHNOLOGY

In-flight results.

Throughput:

Full OPS-SAT image (2048x1944) < 156ms (>25Mp/s)

AI algorithm	Hardware	FPGA time (ms)	Total time (ms)
CNN 8.9	FPGA + CPU	155	4 370
FCN 8.11	FPGA + CPU	156	12 106
ZGP	CPU	NA	32 411

Power consumption:

ANN architecture	HPS PDU (W)	FPGA PDU (W)		Total (W)
		Average	Estimation	
CNN 8.9	2,41 ± 0,11	1,68 ± 0.09	1,59	4,09 ± 0,17
FCN +ZGP	2,61 ± 0,08	1,61 ± 0,05	1,75	4,22 ± 0,08

FRENCH INSTITUTES OF TECHNOLOGY

In-flight results.

Inferences metrics:

CNN degraded by zeroed weights and biases with good precision
SNN with potential overfitting then impacted by spikes encoding
FCN is clearly the best ANN with Fscore over 70%
ZGP provides homogenous results with very good generalization

Metrics	Trainset in Float32 (α)			Testset in Float32 (β)						Testset on FPGA/Cyclone V (σ)					
ANN (arithmetic)	Fscore	Precision	Recall	Fscore	(β-α)	Precision	(β-α)	Recall	(β-α)	Fscore	(σ-β)	Precision	(σ-β)	Recall	(σ-β)
CNN (8.9)	58	74	47	56	-2	70	-4	46	-1	50	-6	76	6	37	-9
CNN/SNN* (8.8)	62	79	51	53	-9	69	-10	44	-7	67	14	62	-7	72	28
FCN (8.11)	62	81	51	72	10	75	-6	70	19	72	0	75	0	69	-1
ZGP (FP32)	60	62	58	63	3	65	3	62	4	63	0	65	0	62	0

*SNN results on ground cause not yet not uploaded

Conclusion and perspectives.

Successful demonstration of on-board image processing by AI on FPGA:

- **High througput (25Mpixels/s)**
- **Low consumption (>2mW)**

New opportunities:

- To download only the useful information to end users
- To alert in near real time from space
- To change the mission by uploading new ANN codes

OPS-SAT limitations:

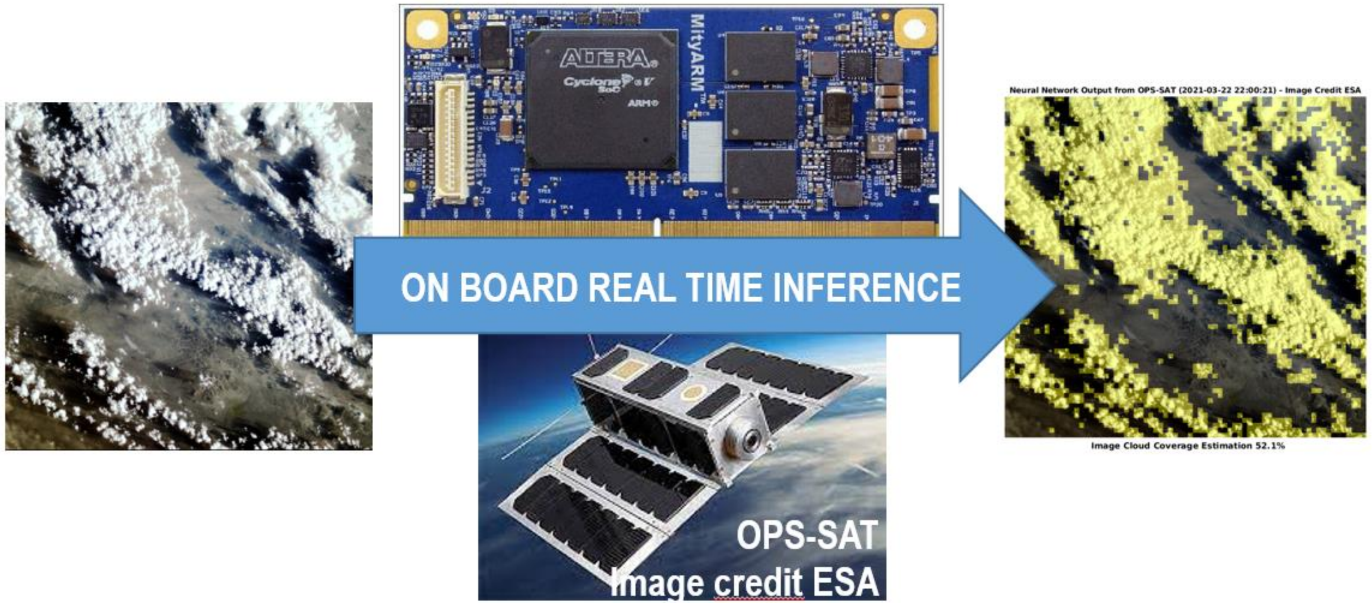
- Clouds detection in RGB
- Images quality without correction
- Limited and unbalanced train dataset
- FPGA number of logical elements
- Quantification losses

Metrics improvements:

- Images correction before inference
- Improved training dataset
- Additional layers/filters on SNN
- Custom optimized CNN
- Optimised quantifying method
- ZGP training on more pixels

FRENCH INSTITUTES OF TECHNOLOGY

Thank you for your attention



Any question please?

FRENCH INSTITUTES OF TECHNOLOGY