European Data Handling & Data Processing Conference – EDHPC 2023

Embedded cloud segmentation using AI : Back on years of experiments in orbit on OPS-SAT

E. Kervennic[1], T. Louis[1,2], M. Benguigui[1,3], Y. Bobichon[1], N. Avaro[1,4], I. Grenet[1,5], F. Férésin[1], A. Girard[1]

[1]IRT Saint Exupéry, France – [2]Université Côte d'Azur, CNRS, LEAT, France – [3]ActiveEon, France – [4]Elsys Design, France – [5]MyDataModels, France

INTRODUCTION

Since 2019, IRT Saint Exupéry has been studying the segmentation of clouds embedded in satellites and designing efficient implementations of neural network models for the Cyclone-V FPGA embedded on the CubeSat 3U OPS-SAT launched by the European Space Agency (ESA).

On March 22, 2021, ESA executed experiments developed by IRT Saint Exupéry on board the OPS-SAT satellite. These neural network inferences in the programmable logic part of an FPGA are, to our knowledge, a first in orbit.

Embedded cloud segmentation has several operational advantages: embedded deletion of images that are operationally useless because they are too cloudy, reduction of bandwidth by transmitting only useful information to end users based on cloud cover.

To get the most out of this demonstrator, we have developed different artificial neural network (ANN) architectures and a model generated from a genetic programming algorithm called "Zoetrope Genetic Programming" (ZGP) [1].

This poster aims to summarize the work carried out by IRT Saint Exupéry on the OPS-SAT mission and to present our main conclusions on the results obtained over the years.

COMPETITIVE ALGORITHMIC ARCHITECTURES FOR CLOUD SEGMENTATION

The approach used to perform segmentation with classification neural networks was described in [2]. It involved dividing the OPS-SAT images of 2048x1944x3 pixels into 5037 small 28x28x3 pixel patches. The first dataset created in 2021 from the initial OPS-SAT data originally consisted of 15 images of 2048x1944x3 pixels for training and 10 images of 2048x1944x3 pixels for performance evaluation. Five different algorithms have been trained on this dataset (Table I) and the FCN performed the best.

Table I : Models performance (Float32) on dataset 2021 (First version).

	CNN	HNN	SNN	FCN	ZGP
	Convolutional Neural Network	Hybrid Neural Network	Spiking Neural Network	Fully Convolutional Network	Zoetrope Genetic Programming
Number of parameters	1440	1428	2666	644	-
Layers	2 conv + 2 dense	2 conv + 2 dense	2 conv + 2 dense	3 conv	-
Accuracy (%)	68	69	67	**77**	69
Precision (%)	70	70	69	**75**	65
Recall (%)	47	50	44	**70**	62
F-Score (%)	56	58	53	**72**	63

TESTING THE FCN TOPOLOGY WITH DIFFERENT DATASET CONFIGURATIONS

In 2022, new acquisitions from the OPS-SAT satellite made it possible to extend the dataset to 37 images of 2048x1944x3 pixels. The new dataset is composed of 29 train images and 8 test images.

In 2023, we white-balanced the 2022 OPS-SAT dataset to account for significant solar illumination variability depending on the latitude or on the degree of off-nadir pointing. The effects on NN performance have been described in [3]. The FCN has been trained on these two new databases and performance where compared to the first training on DB 2021 (Table II).

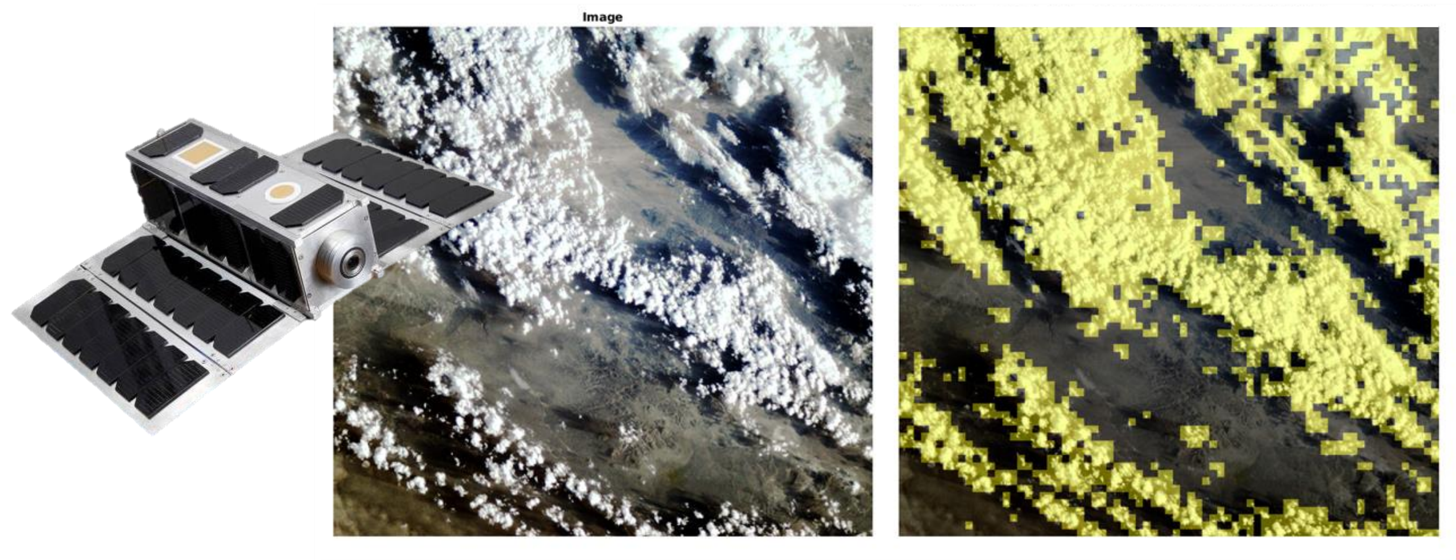
Table II : Float32 performance of the FCN on the 3 dataset versions.

	DB 2021	DB 2022	DB 2023
Accuracy (%)	77	**81**	80
Precision (%)	75	**85**	84
Recall (%)	70	**77**	74
F-Score (%)	72	**81**	79

DEPLOYING THE FCN ON RELEVANT HARDWARE TARGETS

Three hardware targets have been selected to deploy the FCN trained on the dataset 2023, based on their ability to accelerate neural networks inference:

- SoC FPGA : Critical Link MitySoM of OPS-SAT,
- ASIC : Google Coral,
- ASIC : Intel Neural Compute Stick 2.



Segmentation and hardware performance on these three targets are in Table III and Table IV, respectively.

Table III : FCN's inference performance on the different hardware targets.

Hardware target	Baseline	Cyclone V		Myriad X	Edge TPU
Deployment toolchain	TF	VGT	VHDL generic library	OpenVino	TF Lite
Quantification	Float32	FxP8.11	FxP11.13	Float16	Int8
Accuracy	80	**79**	**79**	74	74
Precision	84	**84**	**84**	68	69
Recall	74	74	74	**93**	91
F-Score	79	**79**	78	**79**	**79**

Table IV : FCN latency and power consumption for 28x28pixels patches inference.

Hardware Target		Cyclone V		Myriad X	Edge TPU
Deployment toolchain		VGT	VHDL generic library	OpenVino	TFLite
Quantification		FxP8.11	FxP11.13	Float16	INT8
Resource usage (%)	ALM	**80%**	90%	-	-
	DSP	**0%**	99%	-	-
Latency over number of patches (µs)	1	25	**24**	1 999	889
	5037	125 875	**121 241**	10 071 381	4 479 102
(1/Throughput) over number of patches (µs)	1	25	**24**	252	889
	5037	125 875	**121 241**	1 267 863	4 479 102
Mean power consumption (W)		**1,4**	1,5	2	2
Energy consumption for one patch (µJ)		**35**	37	503	1 778

We also implemented the ZGP model on both a CPU and FPGA and computed the latency for inference (Table V).

Table V : ZGP latency for 2048x1932pixels inference on CPU or in Programmable Logic (PL).

	Latency for one image of 2048x1932px (s)
On-ground Cyclone V CPU	14,91
On-board Cyclone V CPU	32,64
Programmable Logic	**2,65**

CONCLUSION

We studied algorithmic performance of tiny neural network topologies, selected the most efficient topology and trained it on different dataset versions. We identified a gain of around 10 points in precision over the years.

Deploying the network on OPS-SAT's FPGA in orbit as early as March 2021 (a space premiere as far as we know) allowed for the inference of large satellite images in less than 126ms. And the deployment of the same neural network topology on ASIC targets (Google Coral, Intel Neural Compute Stick 2) shows that the processing throughput on FPGA is 10 to 36 times higher than on manufacturer ASIC at an equivalent power consumption of 2W.

The FPGA deployment also maintains the algorithmic performance of the FCN model trained in float32 arithmetic while a larger performance divergence is identified on ASIC.

REFERENCES

[1] Boisbunon et al. – Zoetrope Genetic Programming for regression – GECC 2021.

[2] Feresin et al. – In space image processing using AI embedded on system on module – OBDP 2021.

[3] Feresin et al. – On-board images processing using IA to reduce data transmission – OBPDC 2020.

Contact :
Erwann KERVENNIC
erwann.kervennic@irt-saintexupery.com