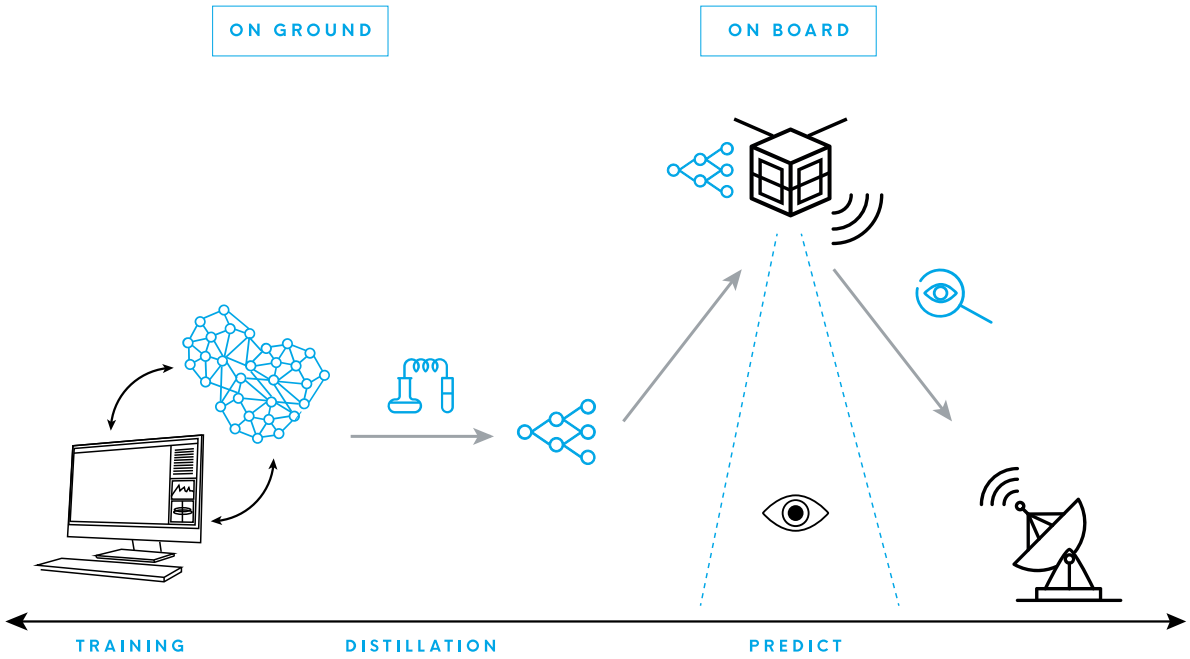
AGENIUM SPACE

esa

→ DISCOVERY



AI at the edge: DeepCube Service IOD/IOV

Executive Summary

Study

ESA AO/2-1816/21/NL/GLC/ov OSIP OPS-SAT EXPERIMENTS CAMPAIGN - STUDIES

Affiliation(s): Agenium Space (Prime)

Activity summary:
OPS-SAT mission is a key element for operational validation in flight (or IOD/IOV) of the DeepCube service of Deep Learning at the edge that our company is preparing with the support of the GSTP programme (GSTP Make, ESA-CNES) and the R&D performed in the CORTEX project (Permanent Open Call EOEP-4, ESA/Phi-Lab, https://esacortexproject.agenium-space.com). The goal of the proposed experiment is to execute on-board the inference of the simplified models defined in those projects. Chosen use case is forest detection (segmentation problem). Our simplified DNN was pre trained on S2 images over Slovenia, these images were processed to simulate OPS SAT images. Then real OPS SAT images were used with the transfer learning method. The Binary Deep Neural Network using convolution was ported on the Cyclone V card using our own custom IP. A lot of efforts have been put to fit the software within the HW constraints (in terms of memory/storage resources) and many tests have been performed with the support of the OPS SAT team to be able to send our package on board. Final results of the experiments are still to be retrieved.

→ THE EUROPEAN SPACE AGENCY

Publishing Date: 09/12/2022
Contract Number: 4000137281/22/NL/GLC/ov
Implemented as ESA Initial Support for Innovation

Executive Summary

The goal of the proposed experiments is to execute on-board the inference of the simplified models implemented in the framework of the DeepCube and Cortex projects (funded respectively under the GSTP programme (GSTP-Make, ESA-CNES) and in the CORTEX project Permanent Open Call EOEP-4, ESA/Phi-Lab, https://esacortexproject.agenium-space.com). This study will contribute to the demonstration/validation (IOD/IOV) of our DeepCube service
The target use case was initially to tackle change detection on board with deforestation in mind. We are able to design an image processing pipeline implementing a Binary Deep Neural Network using convolution. However, due to many hardships about the accuracy image localization (roughly 10km), the possible time allocated for an experiment (1 orbit) and the available memory/storage space in the satellite and time constraints, we were not able to propose a realistic on-board pipeline to assess the change detection problem. None-the-less we are waiting for the results (inputs & outputs) of our on-board segmentation pipeline on the rain forest so as to check on ground that our initial approach to the change detection problem is realistic providing that we get a good product geo-localization.

One of the first step of the study consist in selecting images to train the DNN.
Agenium Space has a 40Tb dataset of optical images (Sentinel 2) over France and the matching forestry essence maps produced by IGN. These images at a 10m resolution can be downsampled to match OPS-SAT images resolution.
10% of this dataset was used to pre-train a Deep Neural Network to extract forest segmentation maps from OPS-SAT acquisitions in real time. In order for the DNN to learn how to process an OPS-SAT image based on this pre-training dataset (gathering Sentinel 2 images), Agenium Space has developed a 4 steps algorithm to simulate the characteristics of an OPS-SAT image including colorimetry, spatial resolution and debayering effects. The binary neural network trained on this database to detect clouds and forest could reach high scores (Accuracy > 80%) on a holdout set of 25x25 km².

Second step consists in preparing the SW on ground. The processing platform on-board OPS-SAT consists of the Intel Cyclone V SoC, providing both a hard processor system (HPS/CPU) (dual-core ARM Cortex-A9) and a Cyclone V FPGA. This is a big advantage as it allows the FPGA to focus on accelerating the computing operations, while the HPS handles the operation scheduling and sequencing. Today, multiple frameworks exist to accelerate deep learning (DL) algorithms on FPGAs. These include the use of High-Level Synthesis tools (HLS) such as FINN (Xilinx), HLS4ML or PipeCNN (OpenCL-based). However, these tools present some limitations: Intel has its own tool, called OpenVINO, for performing inference on different targets, including FPGA. Unfortunately, the Cyclone V is currently not supported.
Thus, we decided to implement our own solution, without the use of external libraries/frameworks, making Agenium Space more flexible in terms of hardware targets for inference.

Using a custom module means that we have full control on how the resources are used. There is no overhead related to the conversion of high-level code to HDL for example. On the other hand, this also means that the implementation will be more complex because everything needs to be built from scratch. To avoid issues related to resources and obtain interesting performances we decided to build an accelerator for Binary Neural Networks (BNN).

A convolutional neural network contains different operations such as convolution, activation functions or scaling. In order to ensure the success of the project while respecting the time constraints, the design was separated into several steps. The design flow can be summarized as follows: for each execution, the CPU loads the data in memory shared with the FPGA and sends an execution instruction. The FPGA then processes the data, writes it back to the memory and indicates that it has finished the current execution.
The sequence of operations is:

- When the experiment is started, the model parameters are loaded.
- It then sends a request to the camera to obtain an image. The amount of memory available on the SEPP is not enough to process the whole image directly.
- It is therefore sliced into tiles of 512 pixels by 512 pixels, which are processed sequentially.

- Each tile is loaded into the DDR, along with the weights of the current convolution layer.
- An instruction is then sent to the FPGA to start its operations.
- The software then processes the output from the FPGA and saves the results, for each tile.
- The output tiles are not reconstructed on-board to reduce memory usage.

This package was tested on ground, on the MitySOM board and on the FlatSat. Several areas can still be optimised. After OPS-SAT, Agenium Space will continue to improve the design by offloading some operations currently being done in the HPS on the FPGA.

The algorithm also needs to be prepared on ground. Since our accelerator only supports Binary Neural Networks (BNN), we adapted our algorithm and retrained a BNN model for this experiment. For the training set, we used modified Sentinel 2 data and OPS-SAT images. The modified S2 images comes from our labelled Forest segmentation database on which we apply a down-sample to obtain a 50-metre resolution image "OPS SAT like".
Our model has first been trained on S2 images then, retrained on OPS-SAT images. Then, we tested it on OPS-SAT images.

The algorithm predicts a segmentation map with 3 channels corresponding to 3 classes:
- background
- forest
- cloud

We determine the class predicted for each pixel by taking the maximum value predicted among the three channels.

Thorough testing in collaboration with the OPS SAT Experimenter support team has been led, allowing to successfully generate a package to be sent on board the satellite. As this report is written we are still expecting the results.

As a conclusion the OPS SAT study has been very fruitful for Agenium Space : although we do not have yet any on board results, it helps us enlarge our experience through the implementation of a binary neural network inference algorithm on the Cyclone V SoC to perform forest segmentation on-board the OPS-SAT satellite. We have also developed our own custom inference IP.
This project adds the Cyclon V FPGA, and the accompanying Intel software (Quartus, Platform Designer, Embedded Design Suite, …) to our toolbox. The Cyclone V, being a relatively small FPGA, encouraged us to simplify/optimize our design on a very low-level. This is very important for "at-the-edge" applications.

Finally, once the on-board results are available and are compared to the expected results, Agenium Space will have successfully completed another IOD/IOV for its Deepcube Service.