AI4 copernicus

Reinforcing the AI4EU Platform by Advancing
Earth Observation Intelligence, Innovation & Adoption

Overview of bootstrapping services

Michele Lazzarini, SatCen

Technical Workshop, 9th March 2023

Aim of bootstrapping services

"Making resources available and maintain them for Open Calls"

The bootstrapping resources include a set of services and resources made available from the Security, Agriculture, Energy and Health communities for AI4Copernicus open calls winners.

The development of these bootstrapping services aimed to reduce the time and resources of the bidders in different tasks as data access (EO and ancillary data), pre-processing, labelling datasets, ML algorithm definition. The AI4Copernicus consortium support allows to address open calls winner's effort on the development of innovative services based on AI.

Bootstrapping Services overview

Resource	Domain	Origin
Sentinel-1 GRD pre-processing	Security/ General	AI4Copernicus
Sentinel-1 SLC pre-processing	Security/ General	AI4Copernicus
Sentinel-2 pre-processing	Security/ General	AI4Copernicus
Sentinel-1 Change detection– Amplitude Change Detection and Multi-temporal Coherence	Security/ General	AI4Copernicus
Sentinel-2 Change Detection	Security/ General	AI4Copernicus
Vector data of human features	Security	AI4Copernicus
Deep network for pixel-level classification of S2 patches	Agriculture/General	AI4Copernicus
TimeSen2Crop	Agriculture	AI4Copernicus
Harmonization of pre-processed Time Series of Sentinel-2 data	Agriculture	AI4Copernicus
Long Short-Term Memory Neural Network for NDVI prediction	Agriculture	AI4Copernicus
Long Short-Term Memory Neural Network for Sentinel-2 for crop type classification	Agriculture	AI4Copernicus
Pre-Trained Long Short-Term Memory	Agriculture	AI4Copernicus
Energy datasets	Energy	External references
Probabilistic downscaling of CAMS air quality model data	Health	AI4Copernicus

Resources milestones

- M12: 1st version of generic and domain-specific services and resources in place → December 2021
- M20: Final version of generic and domain-specific services and resources in place → August 2022
- Services are maintained during open calls

Some practical answers

- How will you access the resources?
  - Resources have been packaged as dockerized applications. We have in the roadmap the integration of the resources in AI4Experiments.
  - Docker registry: https://harborai4c.cloudferro.com/
  - User/Password will be provided to winners.

- How to fine-tune/customize the services?
  - We receive operational needs of users to evolve the services accordingly.
  - Some services can be customized directly by users (source code available in the docker images).

Some practical answers

- Who is responsible of each bootstrapping resources?
  - Each domain is responsible to provide and maintain a specific resource, but the resources are available for all the domains/projects.
  - If you need further information for any service:
    1. Read technical documentation
    2. Ask to AI4Copernicus team
       - https://ai4copernicus-project.eu/contact-us/
       - info@ai4copernicus-project.eu

    3. During the project you will be in contact with the responsible of the services for support

AI4 copernicus

Reinforcing the AI4EU Platform by Advancing
Earth Observation Intelligence, Innovation & Adoption

Security Services

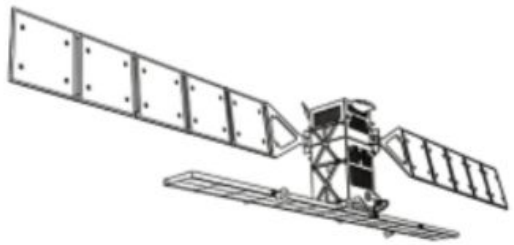
Omar Barrilero, SatCen

Technical Workshop, 9th March 2023

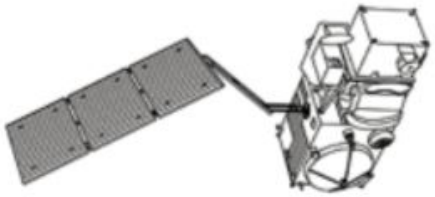
Security Services

Resource	Summary
Sentinel-1 GRD pre-processing	S1 GRD product in native format to terrain corrected calibrated backscatter.
Sentinel-1 SLC pre-processing	S1 SLC product in native format to terrain corrected calibrated backscatter.
Sentinel-2 pre-processing	S2 product in native format to common resolution for all the bands. The process allows to apply a land/sea mask and a cloud mask.
Sentinel-1 Change detection– Amplitude Change Detection and Multi-temporal Coherence	Pairs of S1 SLC products in native format to generate coherence, ACD, MTC
Sentinel-2 Change Detection	Computes (and classifies) the changes using as input a pair of S2-L2A products by using the Change Vector Analysis approach.
Vector data of human features	SatCen has pre-processed and ingested several OSM data layers and can provide the data as a service in the scope of the project.

Input data and tools/libraries



Sentinel-1
SAR



Sentinel-2
Optical

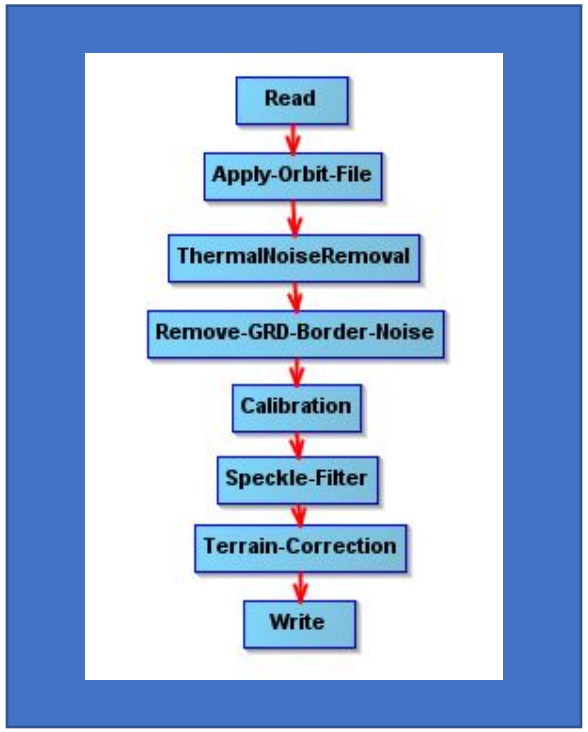
- SNAP is the common software platform and host for the Sentinel Toolboxes and others
- Graph processing Tool (GPT) allows to execute SNAP in batch-mode from command-line. Possibility to create complex workflows.
- Snappy: Allows to access the SNAP Java API from Python

- Geospatial Data Abstraction Library
- Open source library for reading and writing raster geospatial data formats
- Command line utilities for data translation and processing

Sentinel-1 GRD pre-processing

S1 SM/IW GRD

Polarization
Speckle
AoI
Resolution
Projection
Format



terrain-corrected
calibrated backscatter

```
S1-GRD-preprocess --input "VALUE" [--calibration "VALUE"] [--polarization "VALUE"]
[--speckle "VALUE"] [--AoI "WKT"] [--resolution "VALUE"] [--projection "VALUE"]
[--output_format "VALUE"] --output_path "VALUE"
```

Sentinel-1 SLC pre-processing

Polarization

Speckle

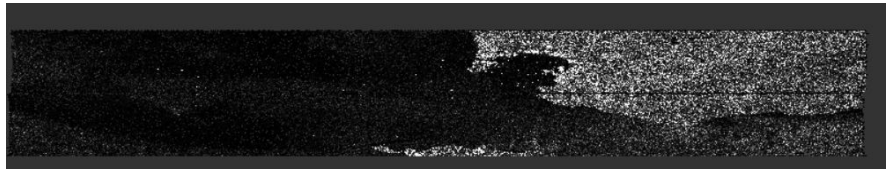
AoI

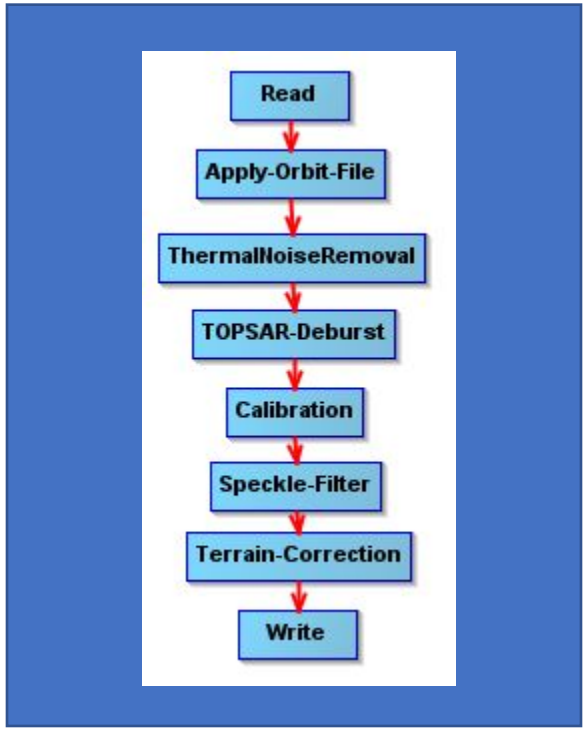
Resolution

Projection

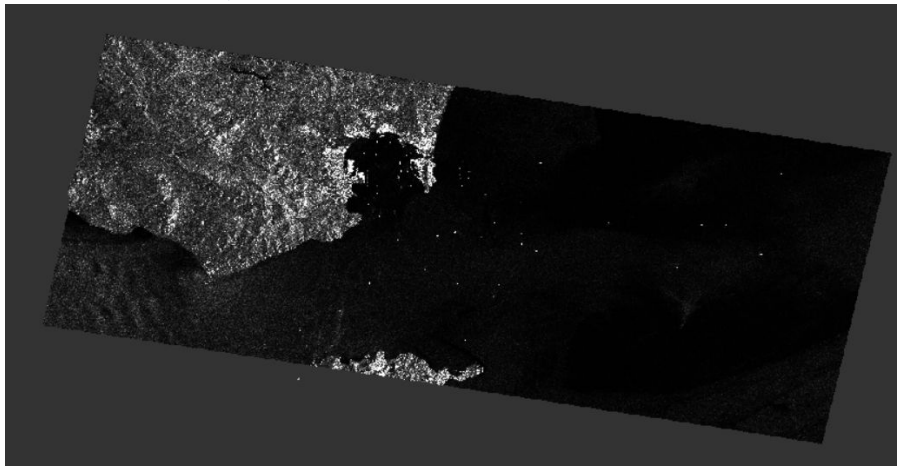
Format

S1 SM/IW SLC



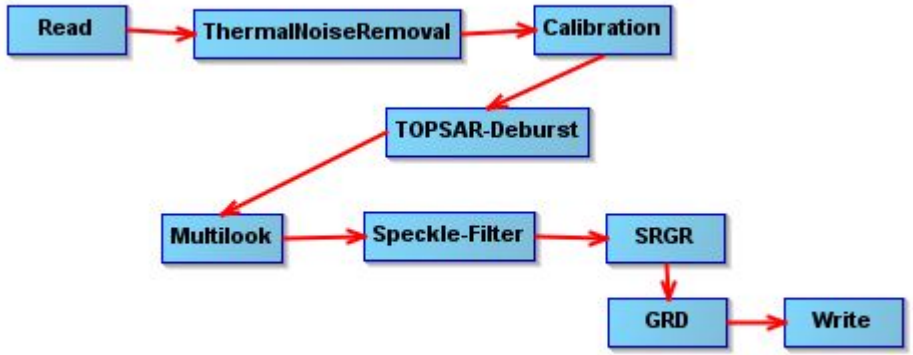


terrain-corrected
calibrated backscatter

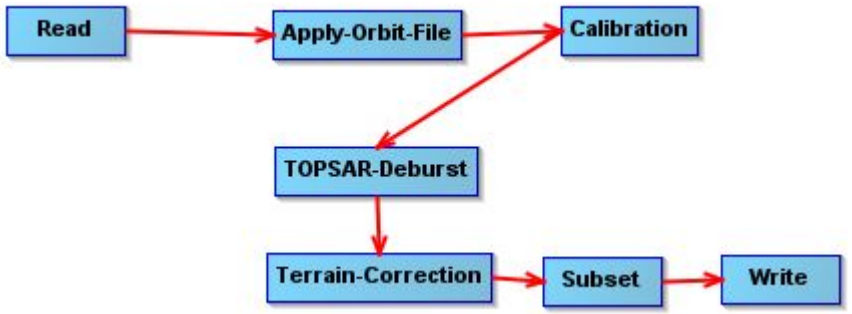


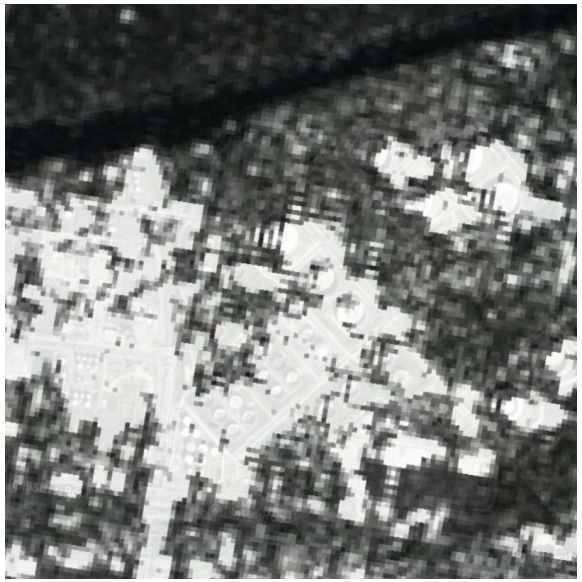
```
S1-SLC-preprocess --input "VALUE" [--calibration "VALUE"] [--polarization "VALUE"]
[--speckle "VALUE"] [--AoI "WKT"] [--resolution "VALUE"] [--projection "VALUE"]
[--output_format "VALUE"] --output_path "VALUE"
```

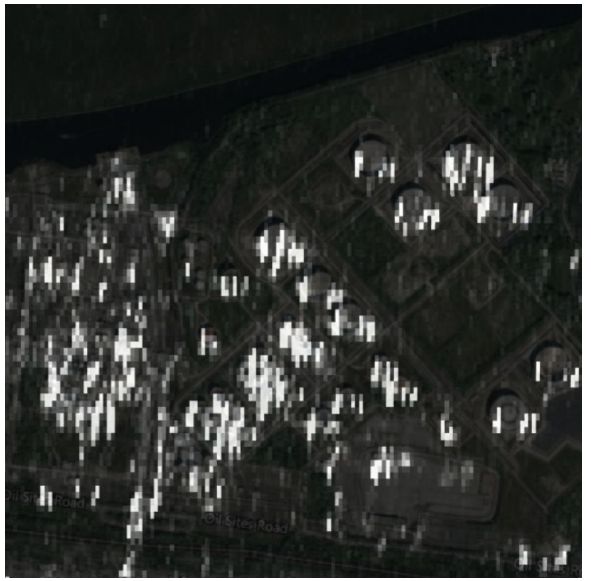
Standard SLC to GRD



Ad-hoc SLC preprocessing







Sentinel-2 pre-processing

S2 L2A

Bands

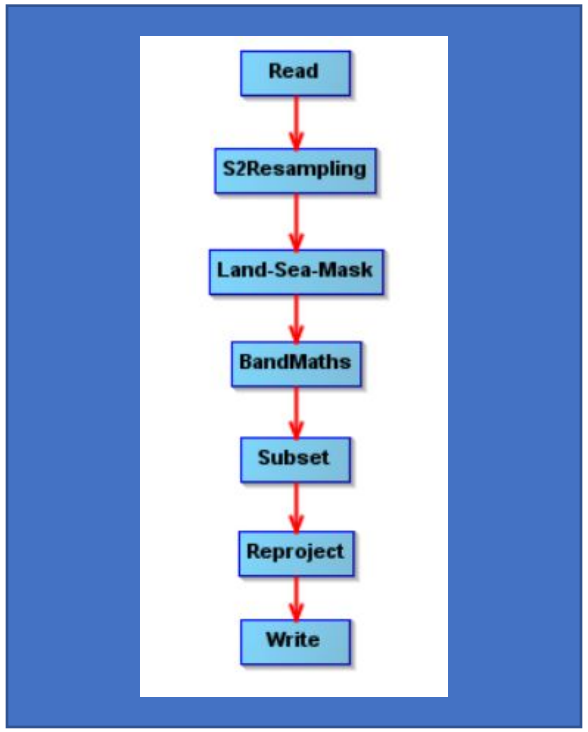
AoI

Resolution

Land/Sea mask

Cloud mask

Projection



Resampled masked reflectances

```
S2-preprocess --input "VALUE" [--bands "XX,XX,XX"] [--landseamask "VALUE"] [--cloudmask "VALUE"] [--AoI "WKT"] [--resolution "VALUE"] [--projection "VALUE"] [--output_format "VALUE"] --output_path "VALUE"
```

Sentinel-1 Change Detection

Pair of S1 SLC

Calibration

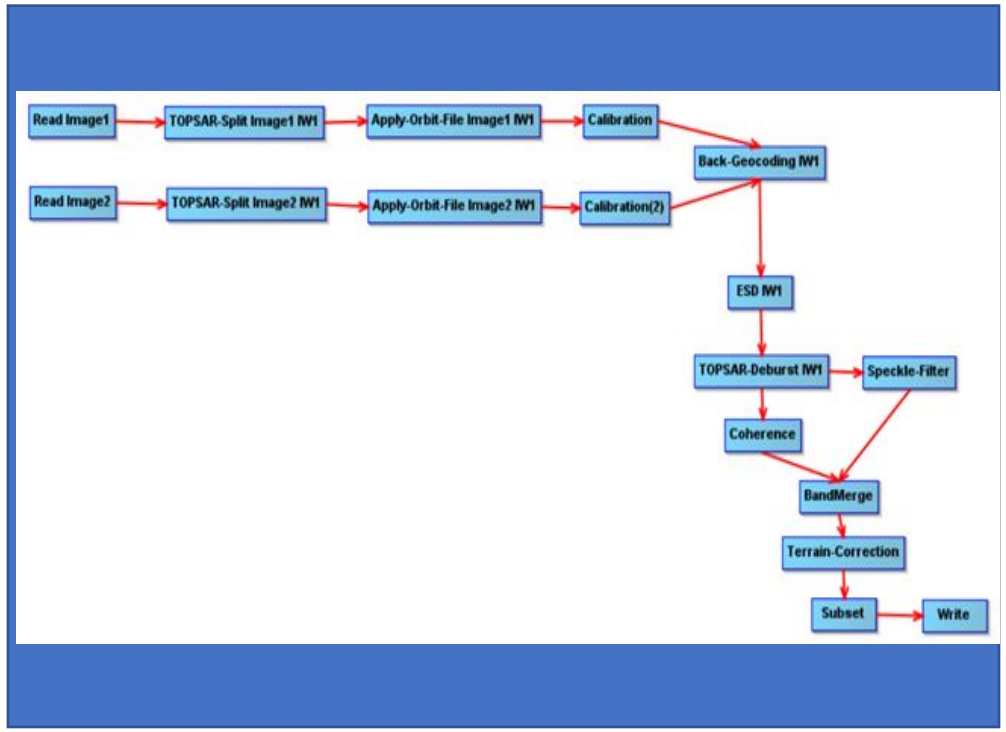
AoI

Resolution

Land/Sea mask

Polarization

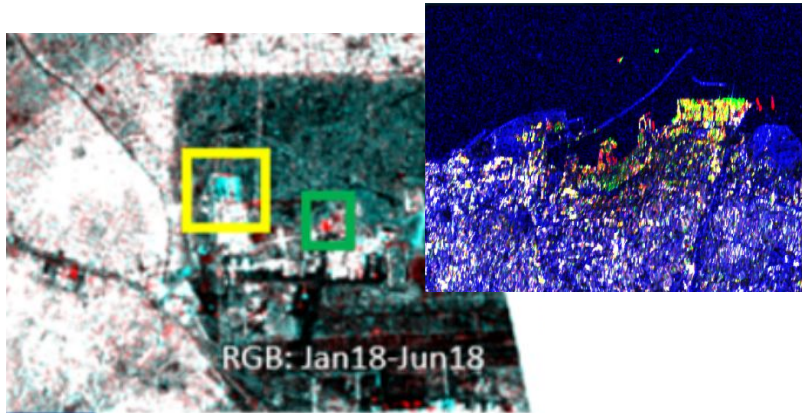
Projection



Coherence

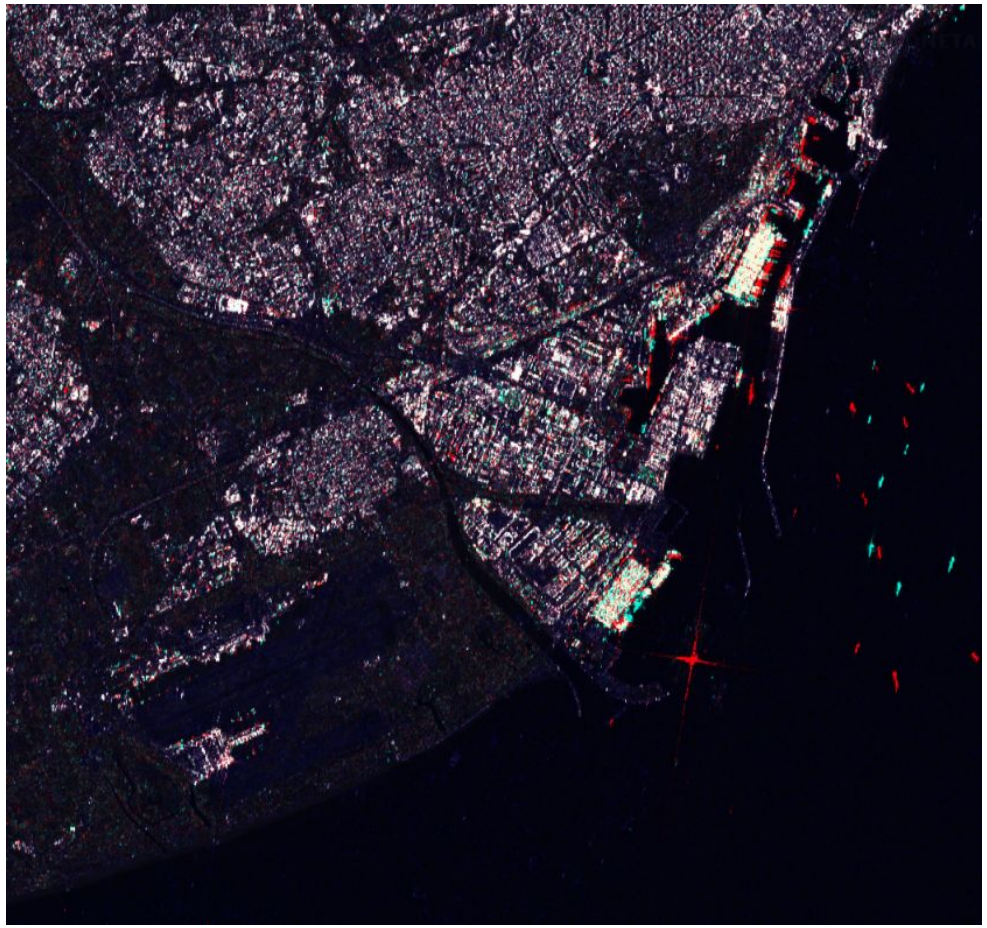
ACD

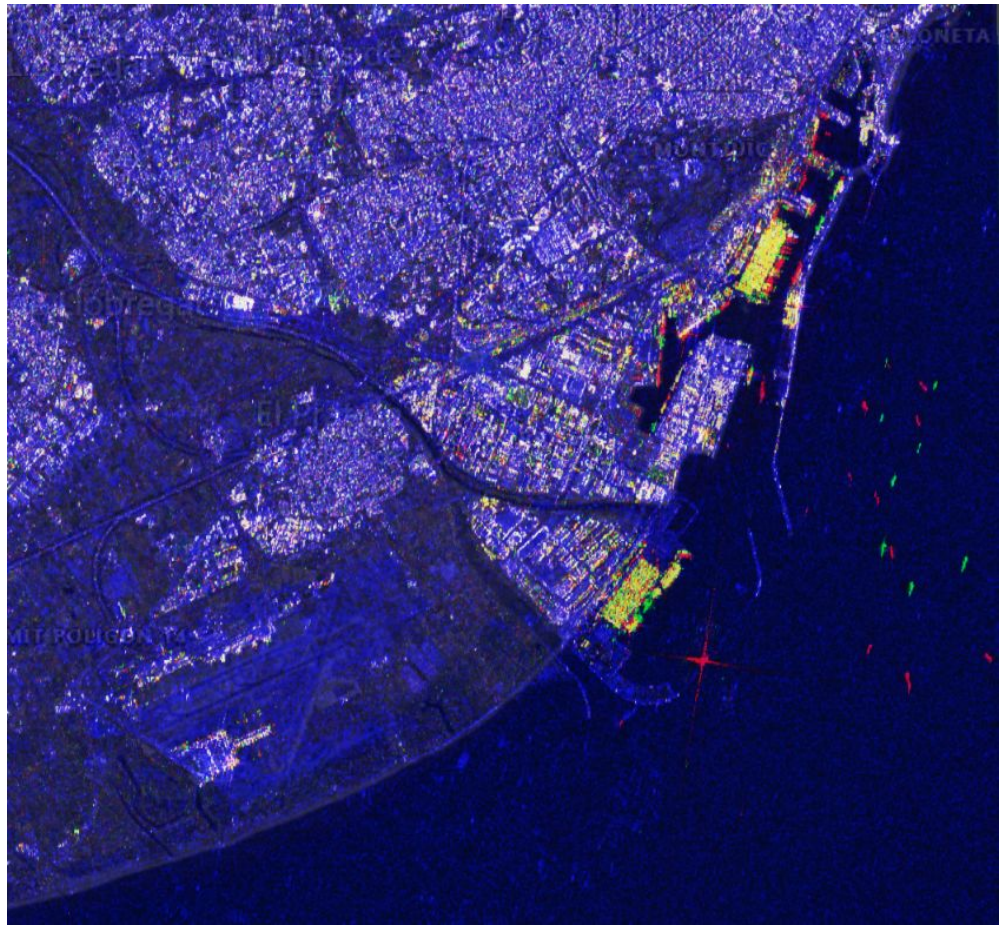
MTC




```
S1-CD --input1 "VALUE" --input2 "VALUE" [--calibration "VALUE"] [--polarization "VALUE"]
[--landseamask "VALUE"] [--speckle "VALUE"] [--AoI "WKT"] [--resolution "VALUE"]
[--projection "VALUE"] [--output_format "VALUE"] --output_path "VALUE"
```

ACD and MTC examples





Sentinel-2 Change Detection

Pair of S2 L2A

Bands

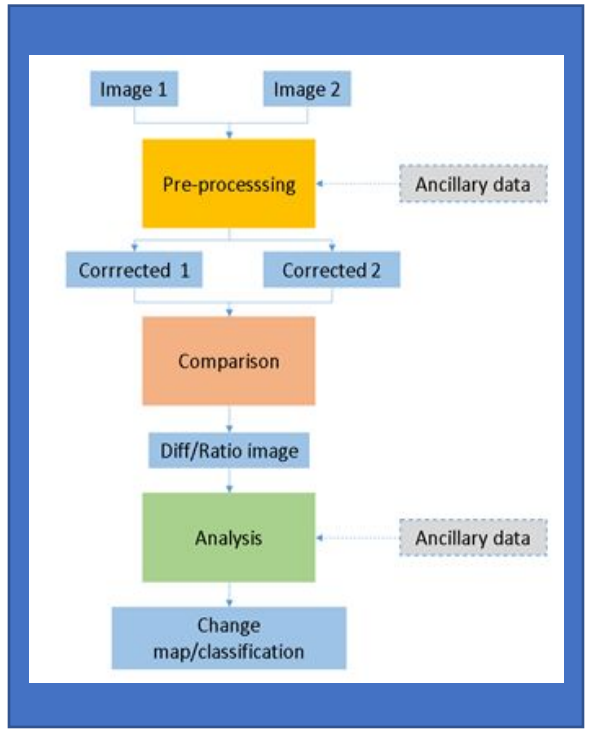
AoI

Resolution

NumberOfClasses

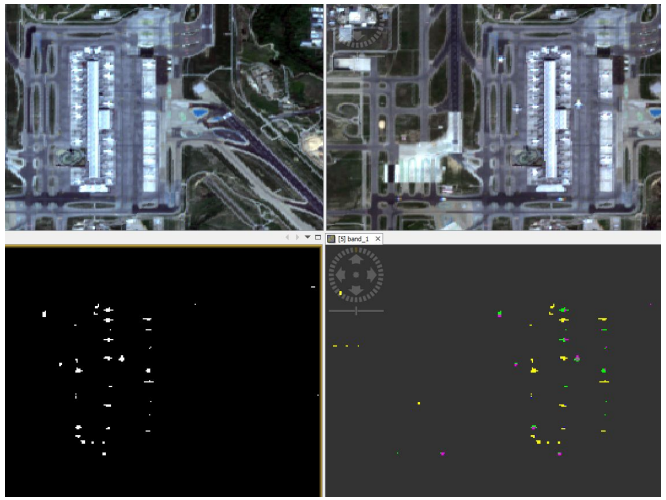
LevelOfConfidence

Cloud mask



CVA

Changes



```
S2-CD --input1 "VALUE" --input2 "VALUE" [--bands "XX,XX,XX"] [--AoI "WKT"] [--resolution "VALUE"] [--projection "VALUE"] [--numberClasses "VALUE"] [--referenceVector "VALUE"] [--levelConfidence "VALUE] [--output_format "VALUE"] --output_path "VALUE"
```

Vector data of human features

- Based in OpenStreetMap (OSM)
- The access to OSM is not always easy:
  - limitations in the servers and APIs
  - data structure is not the preferred by Security domain.

- SatCen has pre-processed OSM data layers and can provide the data as a service in the scope of the project

AI4 copernicus

Reinforcing the AI4EU Platform by Advancing
Earth Observation Intelligence, Innovation & Adoption

Bootstrapping Services: Agriculture

Lorenzo Bruzzone, UniTN

David Hassan, Thales

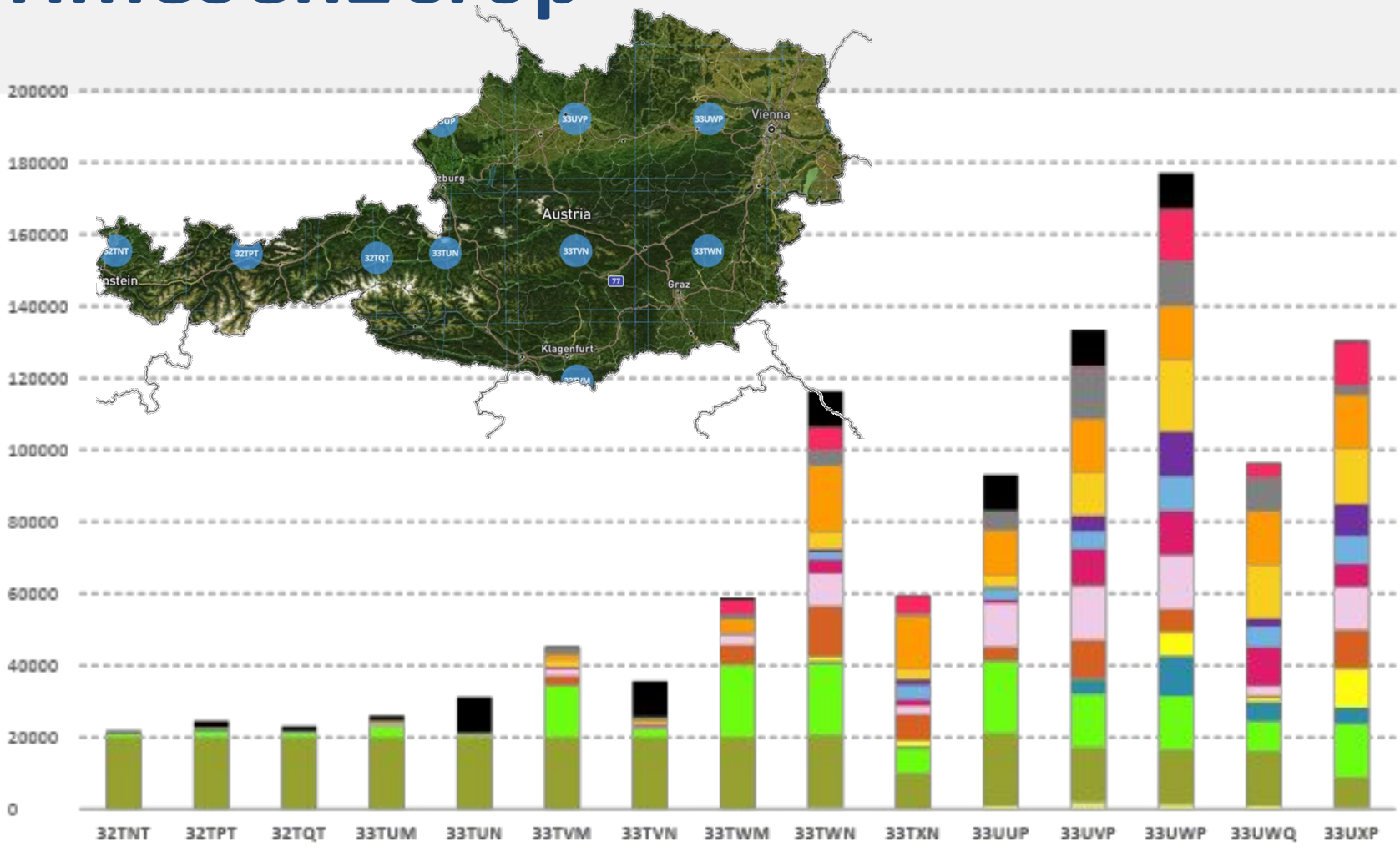
Technical Workshop, 09 March 2023

Agriculture Bootstrapping Services

Resources & Services

- ✔ TimeSen2Crop
- ✔ Harmonization of pre-processed Time Series of Sentinel-2 data.
- ✔ Pre-Trained Long Short-Term Memory Neural Network for Sentinel-2
- ✔ Training of the Long Short-Term Memory Neural Network.
- ✔ Inference using the Pre-Trained LSTM.
- ✔ Training of the Deep Network for pixel-level classification of S2 patches.

TimeSen2Crop



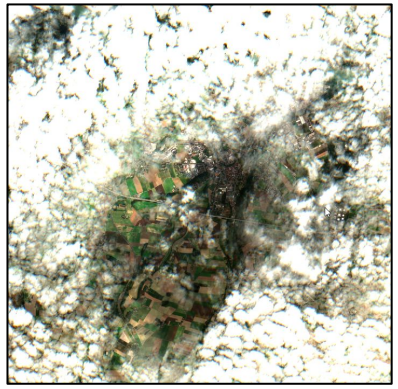
Crop Type	#Samples
Legumes	7098
Grassland	263202
Maize	144315
Potato	27768
Sunflower	22636
Soy	61134
Winter Barley	74060
Winter Caraway	1270
Rye	46191
Rapeseed	37978
Beet	30491
Spring Cereals	77242
Winter Wheat	112326
Triticale	54106
Perm. Plantations	46102
Other Crops	66712
Total	1.072.631

[1] G. Weikmann, C. Paris and L. Bruzzone, "TimeSen2Crop: A Million Labeled Samples Dataset of Sentinel 2 Image Time Series for Crop-Type Classification," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 14, pp. 4699-4708, 2021, doi: 10.1109/JSTARS.2021.3073965.

S2 Tile Harmonization

8th September

11st September



13rd September

18th September

21st September

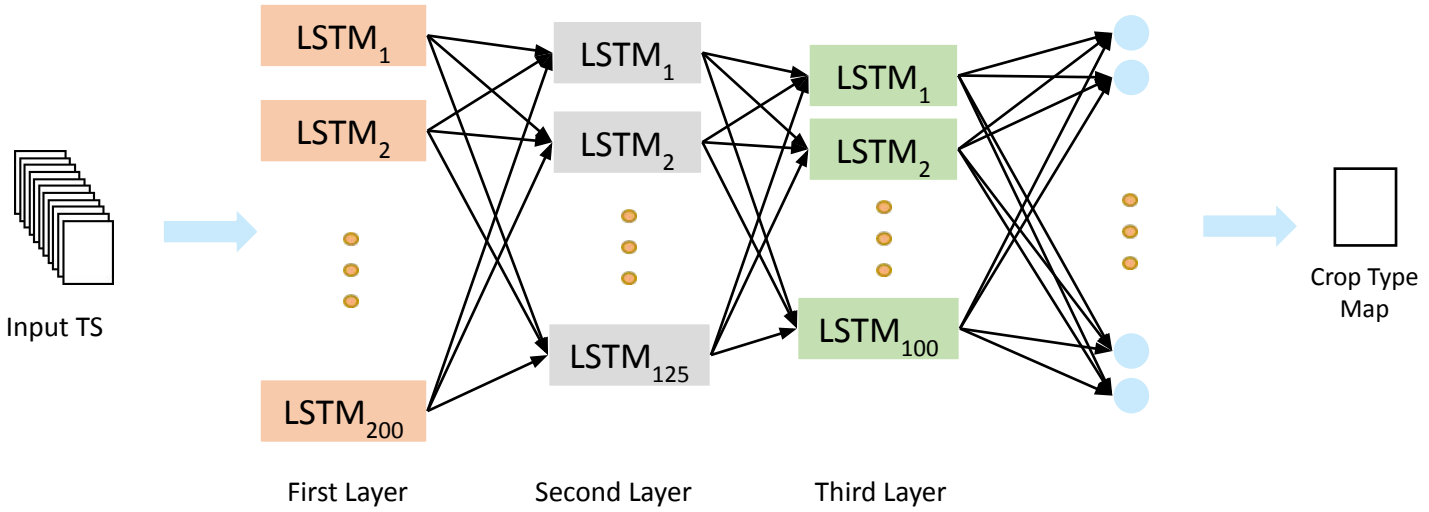
28th September

Monthly Composite (September)

- ✔ The service aim to harmonize Sentinel-2 Time Series (TSs) through a monthly composite approach to create temporally homogeneous time series.
- ✔ Allows processing of different TSs length.
- ✔ Mitigates the presence of clouds in the scene.

Pre-Trained LSTM

- ✔ Long Short-Term Memory (LSTM) exploits the temporal context to discriminate the spectral signatures of the targets and obtain accurate classifications.
- ✔ The network has been trained on the TimeSen2Crop database, balancing the loss function considering the prior probability of each different crop type.



[2] C. Paris, G. Weikmann, L. Bruzzone, "Monitoring of Agricultural Areas by using Sentinel 2 Image Time Series and Deep Learning Techniques", *SPIE Remote Sensing Conference, 21 - 24 September 2020.*

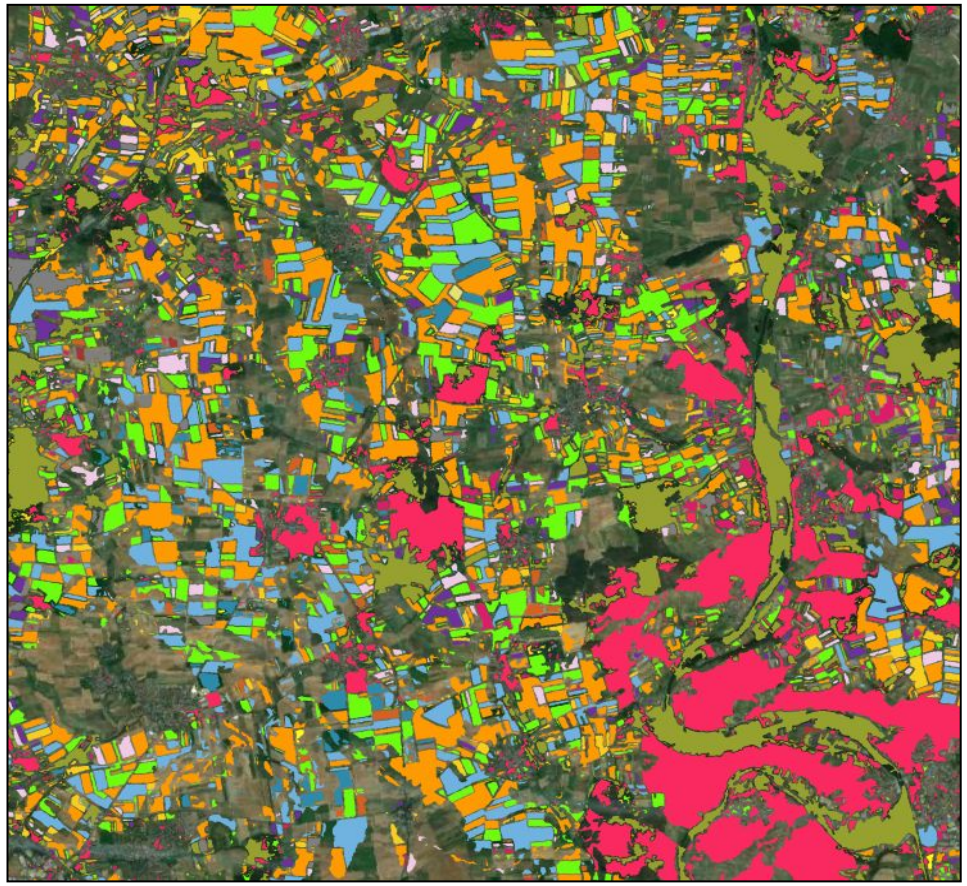
LSTM Training

- ✔ The network can also be trained from scratch, either using the TimeSen2Crop dataset or with a dataset created by the user.
- ✔ A modified version for the prediction of NDVI values has been made available and can be trained using the dataset defined by the user.
- ✔ The parameters of the network can be modified in accordance with the user requirements.
- ✔ The retrieved architecture is stored and can be used directly on the LSTM Inference service to perform the classification.

Parameter	Valid Values	Default Value
epochs: number of epochs to update the internal model parameters	Any number >1	100
batch_size: number of samples used to update the internal models	Any number >2 and power of 2	64
class_weights: flag used in the training to give different weights based on the a-priori probability of each class.	0 (not used) - 1 (used)	0
learning_rate: number describing the step size.	Any number > 0 and < 1	1e-3
dropout: probability of dropping out each unit.	Any number > 0 and < 1	0.3
val: string pointing to the path where a validation set is stored.	path to folder containing the .npy data	None

LSTM Inference

- ✔ The Pre-Trained LSTM or the network built by the user can be used to perform tile classification.
- ✔ The user can specify a different number of time sequence if a time series different than 12 monthly composites must be classified (the network must be trained accordingly).
- ✔ A crop mask can be given as input to perform the classification only on a sub-portion of the original image.
- ✔ The final crop type maps are stored in a .tif format and the posterior probabilities can be retrieved.



Deep Network for pixel-level classification of S2 patches

- Allow user to train pixel-level segmentation models on Sentinel 2 (S2) images.

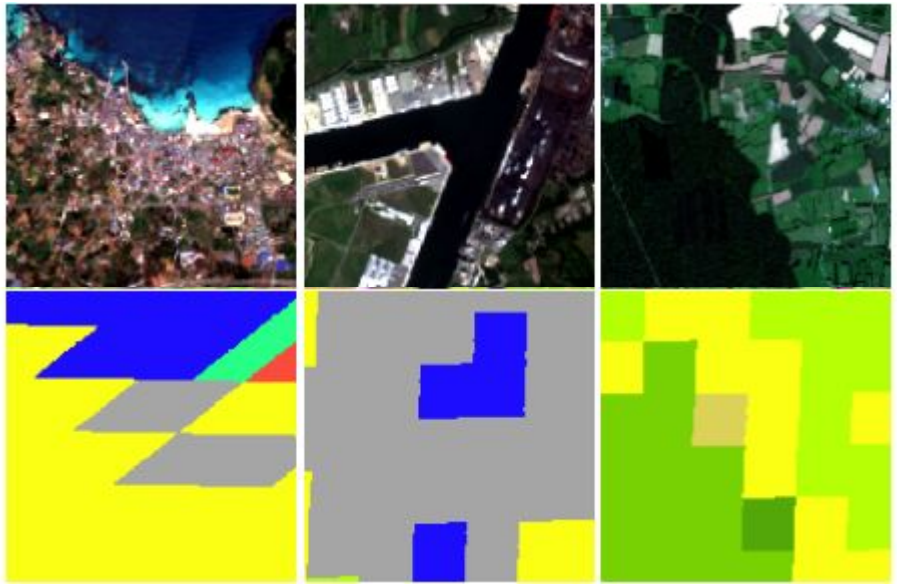
- The goal is to detect one of/a combinaison of S2 classes : crop types (corn, sunflower, wheat, etc), land cover (urban vs natural, water vs land), road extraction (road vs other).

- Service is a docker image usable in an environnement compatible with docker.

Image Segmentation

- Pixel level segmentation : each pixel of the input multispectral image is associated with its corresponding label according to the training data and the selected labels from the user.

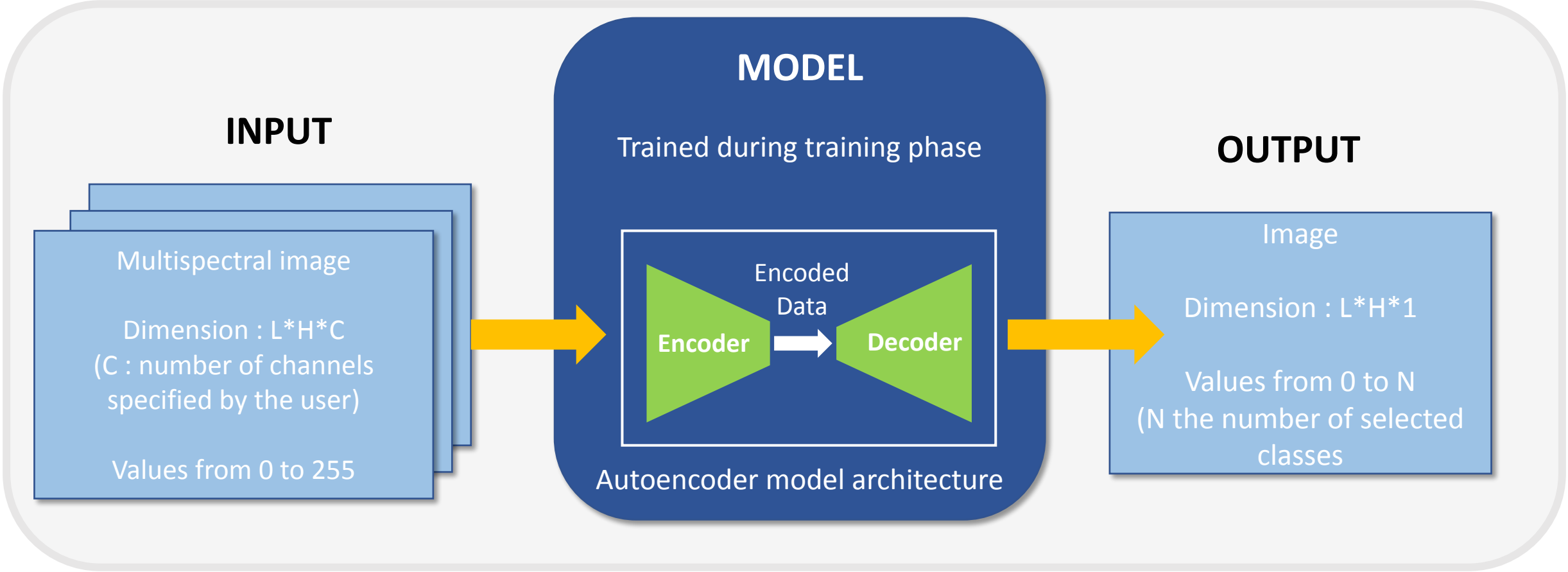
Segmentation



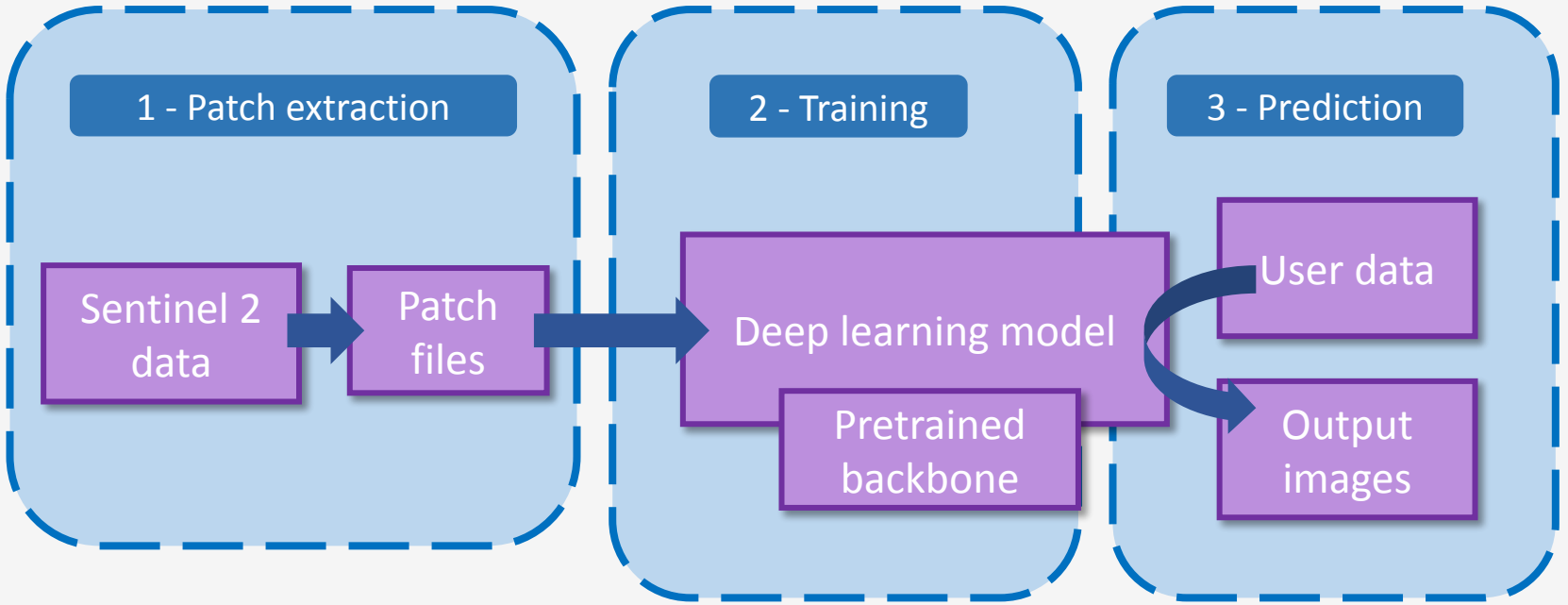
Ref : SEN12MS – A CURATED DATASET OF GEOREFERENCED MULTI-SPECTRAL SENTINEL-1/2 IMAGERY FOR DEEP LEARNING AND DATA FUSION

Image Segmentation

Example: segmentation for one image



Docker Service Description



- All three steps can be run separately
- User can provide custom training data if the input format remains correct
- User can retrieve trained models and predicted data
- Various number of parameters allow user to fine-graine the training process

Additional Information

Among available parameters :

- Paths for input images and input groundtruths
- Paths for output predictions and models
- Selected process (patch extraction, training, prediction or all)
- Use of pretrained backbone, output model format
- Various parameters for the training

For testing purpose, we provide a pre-trained model on SEN12MS data.

Reinforcing the AI4EU Platform by Advancing
Earth Observation Intelligence, Innovation & Adoption

The health bootstrapping service - super-resolution of CAMS air quality data

Mohanad Albughdadi, Mihai Alexe and Vasilis Baousis
European Center for Medium-Range Weather Forecasting (**ECMWF**)
Bonn, Germany / Reading, UK
[firstname.lastname@ecmwf.int]

AI4Copernicus Technical Webinar - March, 2023

Overview

Objective:
Address current public health and air pollution / quality challenges using EO.

Topic:
Probabilistic **downscaling (super-resolution)** of **CAMS air quality** (AQ) and **atmospheric composition** (AC) model output.

Models:
Generative adversarial networks (GANs): capture aleatoric uncertainty in the data

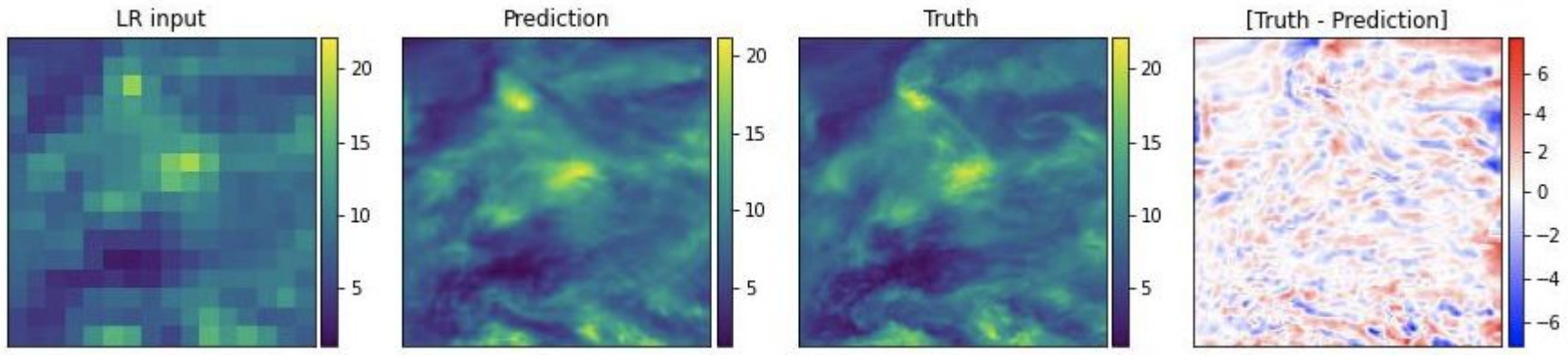
Inputs: Coarsened AQ CAMS-regional, ERA5, EAC4 (CAMS global reanalysis)
Outputs: Super-resolved AQ fields (at CAMS-regional resolution, ca. 10km over Europe)

An example

Super-resolution of PM2.5 fields over Europe.

LR input == coarsened CAMS-Regional data (ca. 80km)
Prediction / ground truth: CAMS-regional data resolution (ca. 10km)

The GAN is able to generate realistic high-frequency content



How to use the service

Two ways to use the service:

1/ Through a **Jupyter-lab instance** (browser-based) running on the K8s container. Example notebooks are available under `/notebooks`.

2/ Using the **command line tools**:
Data downloads (NB: you will need a CDS / ADS user account)

Pretraining

```
$ ai4cop-cams-pretrain --help
usage: ai4cop-cams-pretrain [-h] --model {srgan} --config
CONFIG

optional arguments:
  -h, --help       show this help message and exit

required arguments:
  --model {srgan}  Super-resolution model
  --config CONFIG  Model configuration file (YAML)
```

https://github.com/mishooax/ai4cop-health-cams

Training

```
$ ai4cop-cams-train --help
usage: ai4cop-cams-train [-h] --model {srgan,unet,xnet,swin} --config
CONFIG [--pretrained-generator]

optional arguments:
  -h, --help            show this help message and exit

required arguments:
  --model {srgan,unet,xnet,swin}
                        Super-resolution model
  --config CONFIG       Model configuration file (YAML)

optional arguments:
  --pretrained-generator
```

Inference

```
$ ai4cop-cams-predict --help
usage: ai4cop-cams-predict [-h] --model {srgan} --config CONFIG

optional arguments:
  -h, --help       show this help message and exit

required arguments:
  --model {srgan}  Pre-trained super-resolution model
  --config CONFIG  Model configuration file (YAML)
```

Thank You!

AI4 copernicus

Any Questions?

DEMOKRITOS
NATIONAL CENTRE FOR SCIENTIFIC RESEARCH

HELLENIC REPUBLIC
National and Kapodistrian University of Athens
EST. 1837

ThalesAlenia Space
a Thales / Leonardo company

INSEAD
The Business School for the World®

THALES

ECMWF

CloudFerro

UNIVERSITÀ DI TRENTO

SatCen
European Union Satellite Centre

equinor

BLUE
sight conseil

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101016798.