eo-grow Earth Observation framework for scaled-up processing in Python

eoresearch@sinergise.com

Configurability

- Pipelines are written in a configurable way
- Writing pipelines is very intuitive and designed for non-developers
- eo-grow offers automatic validation of configuration to catch early mistakes
- Possibility to provide custom variables at runtime

{json} Pydantic $ click_

Scalability

- Processing part of the pipeline is not affected by area size
- Pipelines can be run on computing clusters
- Integration with cloud storage options is seamless

RAY docker Clustered System

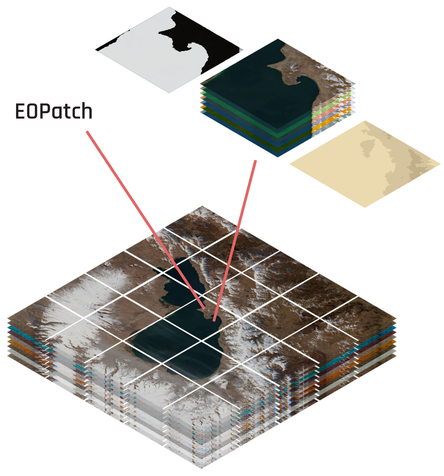
Flexibility

- Effortless switching between local executions and running on a cluster
- Adaptation of the pipeline to a new area-of-interest or time interval is effortless
- Re-using pipeline configurations for increased productivity

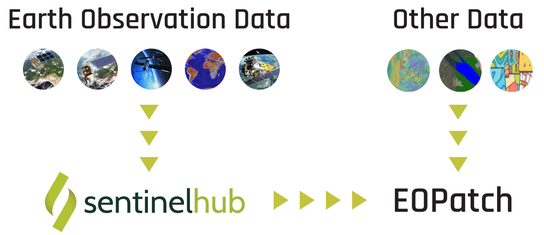
aws jupyter Copernicus Data Space Ecosystem

Problem definition

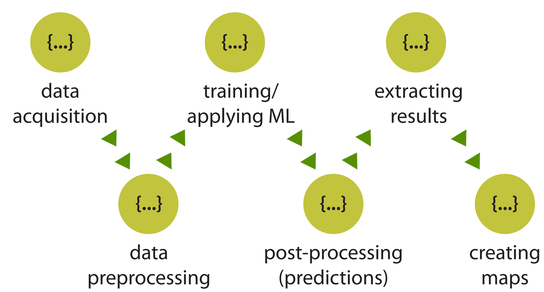
- Define the time interval of your choice
- Determine the area of interest to be processed
- Define the size and resolution of the basic processing unit (EOPatch)
- Prepare the processing pipelines for your specific tasks



Data acquisition



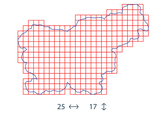
Processing



LULC processing example

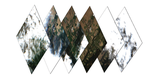
Splitting the AOI

The area of Slovenia is split into 25 × 17 equal parts, which amounts to 300 EOPatches of about 1000 × 1000 px at a 10 m resolution.
The splitting choice depends on the amount of available resources.



Adding info to the EOPatches

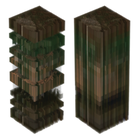
eo-learn uses the sentienlhub-py package to download six Sentinel-2 L1C bands (B2, B3, B4, B8, B11, B12) for each patch. The S2-cloudless cloud detector is used to obtain the cloud probabilities and cloud masks. Normalized difference indices (such as NDVI and NDWI) are calculated. Official land cover data is added in the raster form of 10 classes (cultivated land, forest, grassland, shrubland, water, wetlands, tundra, artificial surface, bareland, snow and ice).



Stack of true color images of a random area in the AOI.

Temporal interpolation

Due to the non-constant acquisition dates of the satellites and irregular weather conditions, missing data is very common.
To tackle this, we perform data interpolation and estimate the pixel values at arbitrary dates.



Temporal stack with missing data due to clouds before temporal interpolation (left) and after (right).

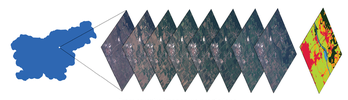
Spatial sampling

A negative buffer with a disk size of 1 pixel is performed in order to remove pixel-wide structures and pixels on the borders between classes.
By randomly sampling the time-series of pixels (40k per patch) we select a subset which is used in ML model training.

Effects of applying the erosion task on a random part of the reference map of the AOI.

ML model training

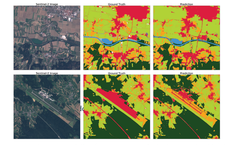
The training pixels and labels are accumulated over all patches.
The data sample is split into the train (80 %) and test (20 %) samples, where the former is used to train the model, while the latter is used for the model validation. A Light Gradient Boosting Machine (LightGBM) is used as a machine learning model.



From satellite imagery to predictions.

Predicting the LC

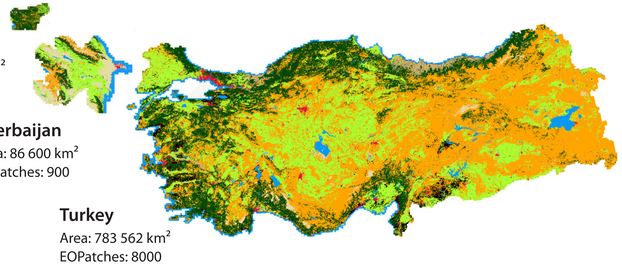
The trained ML model is plugged back into the pipeline, where it predicts the land cover for each patch.
When all patches are processed, GIS related Python libraries are used to stich together the land cover map for the whole AOI.



Sentinel-2 L1C, ground truth and prediction for an area near Celje (top) and the Ljubljana airport (right).

Flexibility and scaling

From small to large countries, or even continents, eo-grow is capable of processing large amounts of data, strongly utilizing cluster computing to efficiently distribute the workload. A processing pipeline is designed to run on a single patch, and the user then applies it over the whole area, regardless of size.



Slovenia
Area: 20 273 km²
EOPatches: 250

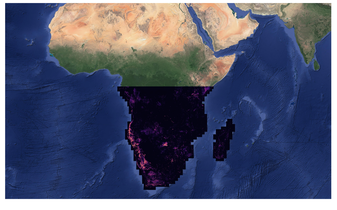
Azerbaijan
Area: 86 600 km²
EOPatches: 900

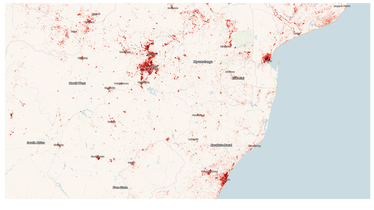
Turkey
Area: 783 562 km²
EOPatches: 8000

Sharing is encouraged

Use case 1: Africa Built-up area detection

With eo-grow it is possible to detecting built-up areas over the full continent of Africa at 120 m resolution. This was a done as a proof-of-principle to showcase the capabilities of the library and possible extensions, such as continuing with a drill-down approach, or using the application for continuous monitoring purposes.





Use case 2: Ukraine Field Delineation

We used eo-grow to automatically determine the borders of agricultural fields from satellite images, based on the similarity of spatial, spectral, and temporal properties of pixels belonging to the same parcel. A U-Net-based deep neural network model was employed over Ukraine for years 2016-2022, and the results facilitated further research into how the war is affecting the agricultural landscape in Ukraine, and consequently also global food supplies.

The project has received funding from European Union's Horizon 2020 Research and Innovation Programme" under the Grant Agreement 101004112.