Sinergise Shares Example EO Data for Free in Order to Bring Closer the Fields of Remote Sensing and Machine Learning

In the past half year, [Sinergise](#) has published a series of blog posts on [Medium](#) on the topic of land use and land cover (LULC) classification based on data from Sentinel-2 imagery. The blog posts very thoroughly -- yet intuitively -- describe the full process of how to start using satellite data in order to extract useful information using machine learning. In addition to the explanations in the blog posts, they also provided an extensive code example in the form of a [Jupyter notebook](#), which anyone can see and execute. This makes the field of machine learning more approachable for people from the remote sensing community, while simultaneously introducing the huge potential of Earth observation data to data scientists and machine learning experts with little to none remote sensing expertise.

First Obstacles in the Field

The problem of a large amount of high spatial resolution data at high revisit frequency has already been noted by Sinergise in one of their previous articles, where they expose the lack of frameworks able to automatically extract complex patterns in such *spatio-temporal* data. This means that in order to start producing valuable information, one must first deal with the creation and implementation of various tools which are used to obtain this information in an efficient and a reliable way, resulting in the loss of valuable hours which could be better spent on the value extraction itself.

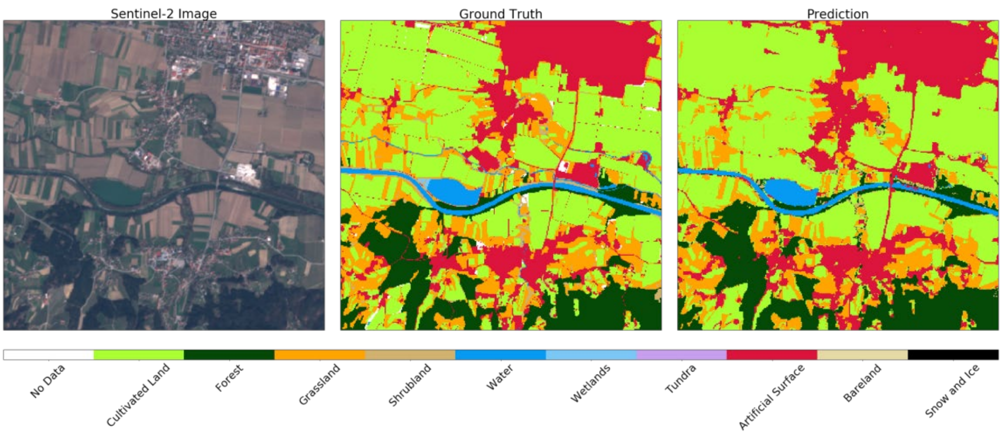

Figure 1: Sentinel-2 image (left), ground truth (centre) and prediction (right) for a random EOPatch in the selected region.

Sharing the Free Dataset

In order to help speed up the learning process of starting to explore the data, Sinergise has shared an *eo-learn* compatible Sentinel-2 dataset for the whole of Slovenia 2017 in [their last blog post](). The dataset has been obtained using [Sentinel Hub services]() and is available in the form of about 300 of so-called *EOPatches*, where each patch represents an area of about 1000 x 1000 pixels at 10 m resolution. Each patch contains the values of 6 Sentinel-2 bands for the year of 2017, which are most commonly used in machine learning approaches regarding LULC classification. Additionally, the cloud masks obtained with the s2cloudless(link) cloud detector are added, including some other information, such as the count of valid pixels in the given time period, where a valid pixel counts as a successfully detected, non-cloudy pixel.

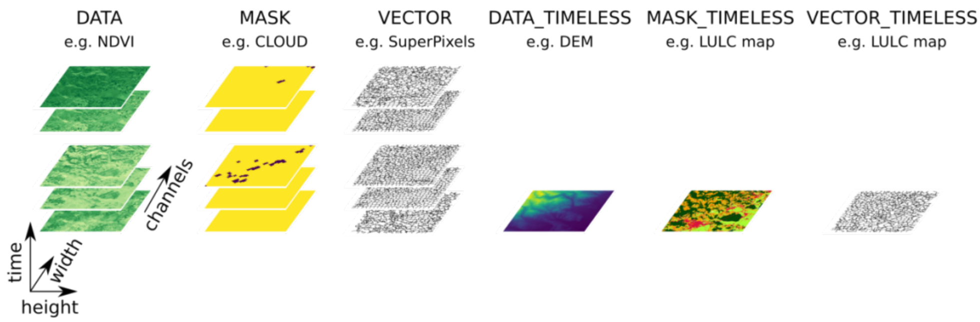

Figure 2: Structure of an EOPatch (from the eo-learn introduction blog post).

Improving the Classification Results

The focus of Sinergise's last LULC blog post was about experimenting with datasets and improving the classification results, showing how to easily do so with their *eo-learn* package. They conducted a series of experiments that studied, for example, the effects of cloudy data on the classification result, the differences among various resampling choices after the temporal interpolation, or simply checking for classification improvements using deep learning algorithms. A more detailed description of two of the experiments is shown below.

Cloud experiment

Clouds are a nuisance in the world of EO, especially when working with machine learning algorithms, where you want to detect the clouds and remove them from your dataset in order to perform a temporal interpolation over the missing data. Four experiments in combinations involving cloudy scene filtering and taking the cloud masks into account when performing linear interpolation were studied. The conclusion was that the clouds do not seem to affect the overall performance much. This might be due to the fact that the reference map quality is very high and the model is able to correctly determine the land cover label based on just a few observations. However, this might just be the case for this particular region and the results probably would not generalise well for all areas.

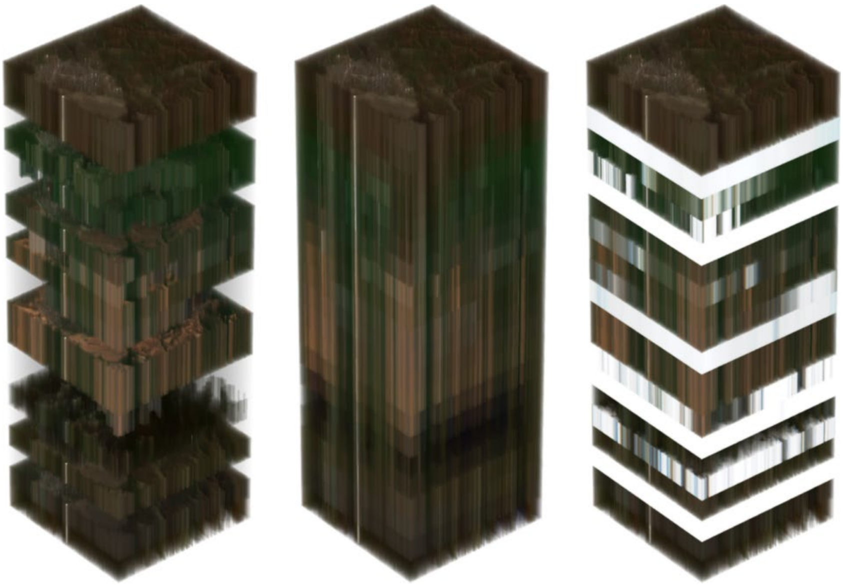

Figure 3: A visual representation of a temporal stack of Sentinel-2 images over a randomly selected area. The transparent pixels on the left imply missing data due to cloud coverage. The stack in the centre represents the pixel values after cloudy scene filtering and temporal interpolation with cloud masking (best case scenario), while the stack on the right shows the case without cloudy scene filtering and no cloud masking performed during interpolation (worst case scenario).

Resampling experiment

The choice of temporal resampling after the interpolation is not obvious. On one hand, you want a relatively fine grid of sampled dates in order not to lose valuable data, but at some point, all available information is taken into account, so including more sampling dates should not improve the result further. On the other hand, you are constrained by computing resources. Decreasing the interval step by a factor of 2 doubles the amount of time frames after the interpolation, and therefore increases the number of features that are used in the classifier learning. The question is if the improvement of the result, in this case, is large enough to justify the increased usage of computing resources. The results showed that the amount of required computing resources approximately doubles, due to the increased number of training features, while the increase in overall accuracy (OA) and the weighted F1 score is only less than one per cent. Such improvements are too small to be visible in a proper application, so changing the original choice of resampling is not necessary.

To learn more about *eo-learn* and how to use it for LULC classification we recommend following the links below:

- Introducing *eo-learn*
- Part 1
- Part 2
- Part 3
- Jupyter Notebook
- Shared Dataset

***

If you find the described experimentation with the provided EO data too complex, you can explore the classification prepared by [European Space Agency (ESA)] in [EO Browser].

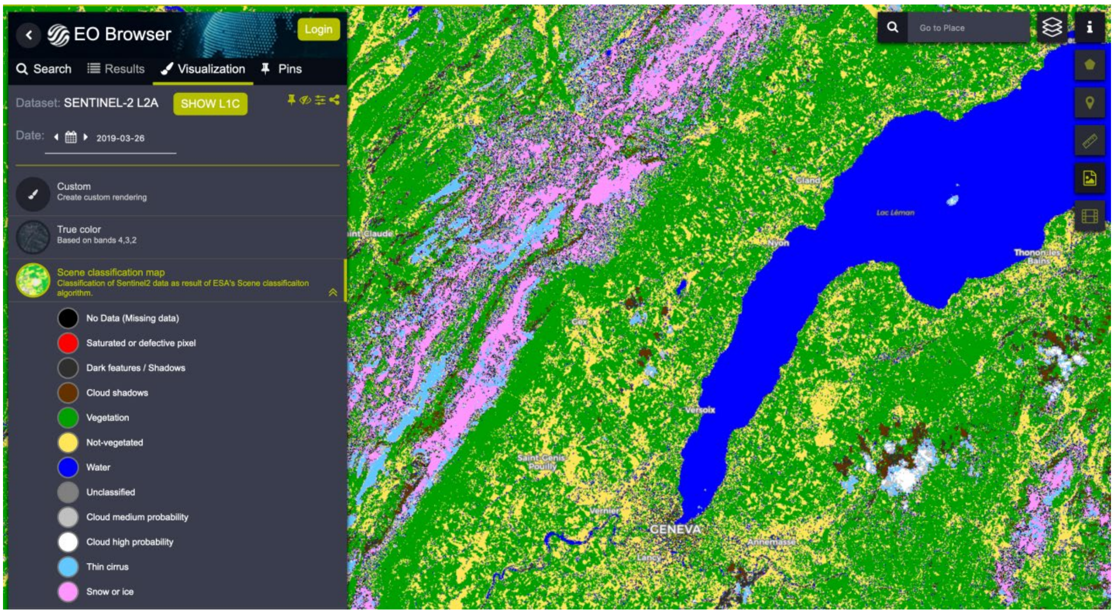

Figure 4: Scene classification map of Sentinel-2 data as a result of ESA's Scene classification algorithm explored in EO Browser.

You can also simply use the custom option to write your own custom script. With this intuitive research you will be able to expose important details of your observation. While you are already exploring this possibility, don't miss a chance to participate in the Sentinel Hub Custom Script Contest. You still have time to submit your scripts until May 5, 2019 and get a chance to win attractive prizes. See their web page and blog about the contest for more details.