Detection of snow/ice and clouds in Sentinel-2 satellite imagery

Introduction

The availability of large amounts of satellite imagery data through the European Copernicus project and open source platforms like the OpenDataCube (ODC) greatly bolsters the opportunities to apply classical machine learning and deep learning algorithms. These models can be employed to meet many challenges. One challenge that is particularly relevant in our time is that of climate change and global warming. An important aspect of this challenge is monitoring the effects of climate change using the aforementioned satellite data. Thus change of snow cover in itself is already an interesting phenomenon to follow and report as a complement to other measurements which allow us to monitor the indicators of climate, such as surface temperature, ocean heat, or sea levels. More importantly, however, it has been widely researched how the change in snow cover influences surface energy level, water balance, thermal regimes, and vegetation and thus plays a major role in the ecological, and hydrological systems, as well as the climate of Arctic and other regions [1,2,3]. Moreover, the extent of snow cover, has more direct economic effects too. For example in road cleaning costs, and in architecture [4]. In order to accurately track the changes in snow cover, it is an important step to be able to identify snow in satellite images, and also to distinguish between a snow-covered area and an area occluded by clouds. This inspired the launch of this challenge.

Background

The detection of cloud and snow has had a key role in multiple remote sensing applications [5], and has provided a difficult challenge, due to the similar spectral characteristics of the two [6]. This, and the various implications of snow cover listed in the previous section lead us to set this challenge for the Copernicus Hackathon, where participants should demonstrate the utility of Copernicus data and the Open Data Cube environment on a challenging and relevant problem.

Goal

The challenge has two aspects:

1) Competing teams use the Sentinel-2 satellite imagery data available of Sweden to construct a data set that is appropriate to train and evaluate models for detecting snow/ice or cloud in multispectral satellite images For this they can make use of the scene classification product of Sentinel-2 [7] that is also available in the Open DataCube, as well as any outside resources that are available to the public (i.e. other teams also have the opportunity to use). It is important to note that the scene classification product has many errors is differentiating between clouds and snow cover, and thus it should not be used as ground truth in its present condition.

2) Competing teams train and evaluate their models for classification, which is divided into two tasks:

2a) The first part is a simple classification problem. For a given multispectral satellite image of an area in Sweden, the task is to identify, whether the image contains clouds, snow/ice, or neither. For this decision, the model can make use of any of the spectral bands of Sentinel-2 [8]. Teams can also decide whether further information (e.g. geographical coordinates of the area, time of the satellite image taken) should be used. Justifying the use (or disregard) of this data is also an important part of the classification problem.

2b) For the second part, the goal is a more fine grained classification, where the trained model also has to accurately predict the location of cloud and snow patches in the area of interest.

Task Description and Data

Data to be used: satellite imagery data of Sweden that will be available to be queried through the OpenDataCube environment. The organisers may also provide access to other datasets related to the tasks. However, these will be provided as is, without taking any responsibility for their content, or preparing any specific scripts facilitating their use.

Expected outcome and evaluation metric: the task would be evaluated on the following criteria (in order of decreasing importance)

1. The degree of accuracy to which the models trained perform the classification task of 2a) and the detection task of 2b)
2. The adequacy of the datasets created for the task
3. The technical difficulty and originality of the presented solution
4. The quality of the presentation (this includes the validity and strength of the argument teams provide for different decisions, such as error metric used, as well as the validity and strength of the use case presented for the detection task)
5. The size of outside packages invoked (here smaller is better: the solution not relying on big third party packages are preferable)

Important note: as repeatability and reproducibility are crucial questions in science, it is important that each team saves its final models for prediction (the one their performance should be measured on) in a dedicated folder. Furthermore, code should be written in a way that it enables organizers to evaluate said models on new data, or the data given during the hackathon, in the latter case the output should be the same as reported by the team. This also means that the use of third-party libraries should be clearly stated. This requirement is not only to ensure repeatability, but also to contribute to the integrity of the competition. The organizers reserve the right to exclude teams from the final evaluation that do not fulfil these requirements.

Requirements

- Experience with programming (Python)
- Basic knowledge of image processing methods
- Familiarity with some Open Data Cube examples for querying and visualising data (https://datacube-core.readthedocs.io/en/stable/user/guide.html) – we will provide example scripts to work with the Copernicus data in the Swedish Space Data Lab environment.

References

[1] Callaghan, T.V., Johansson, M., Brown, R.D. et al. Multiple Effects of Changes in Arctic Snow Cover. AMBIO 40, 32–45 (2011). https://doi.org/10.1007/s13280-011-0213-x

[2] Da Ronco, P., De Michele, C., Montesarchio, M., Mercogliano, P., 2016. Comparing COSMO-CLM simulations and MODIS data of snow cover extent and distribution over Italian Alps. Clim. Dyn. 47, 1–23. http://dx.doi.org/10.1007/s00382-016-3054-2.

[3] Jungqvist, G., Oni, S.K., Effect of Climate Change on Soil Temperature in Swedish Boreal Forests, 2014, https://doi.org/10.1371/journal.pone.0093957

[4] Croce, Pietro & Formichi, Paolo & Landi, Filippo & Mercogliano, Paola & Bucchignani, Edoardo & Dosio, Alessandro & Dimova, Silvia. (2018). The snow load in Europe and the climate change. Climate Risk Management. 20. 10.1016/j.crm.2018.03.001.

[5] Y. Zhan, J. Wang, J. Shi, G. Cheng, L. Yao and W. Sun, "Distinguishing Cloud and Snow in Satellite Images via Deep Convolutional Network," in IEEE Geoscience and Remote Sensing Letters, vol. 14, no. 10, pp. 1785-1789, Oct. 2017.

[6] Han, Ling & Wu, Tingting & Liu, Qing & Liu, Zhiheng. (2019). A Novel Approach for Cloud Detection in Scenes with Snow/Ice Using High Resolution Sentinel-2 Images. Atmosphere. 10. 44. 10.3390/atmos10020044.

[7] Sentinel-2 Scene Classification: https://earth.esa.int/web/sentinel/technical-guides/sentinel-2-msi/level-2a/algorithm

[8] Sentinel-2A Satellite Sensor: https://www.satimagingcorp.com/satellite-sensors/other-satellite-sensors/sentinel-2a/