Land use land cover classification in Sentinel-2 satellite imagery as a first step towards COVID-19 use cases

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 machine learning models can be employed to meet many challenges. One challenge that is particularly relevant in our time is that of the global Covid-19 pandemic. An important aspect of this challenge is monitoring human activity in different types of areas, as well as traffic between different areas. On one hand, increased activity in areas corresponding to hospitals and health care institutes, for example, in the right circumstances can forecast a new wave of viral infection. On the other hand, monitoring human activity in areas associated with commerce and entertainment can be used as an approximation of how strictly certain lockdown measures are observed. Monitoring activity in residential areas can also provide important insights related to the ongoing global pandemic.

Background

Telecom data can provide an approximation for human activity in certain areas, or provide information about traffic between areas. For this information to be more useful for decisions and predictions regarding, for example a global pandemic, it is important however, that one can tell what type of area this information corresponds to. As an example, in case of a lockdown, an increased human presence during working hours is expected in a residential area, but not in commercial or industrial areas. Thus it is important that one can identify the role (land use) of different areas. One possible approach for land use classification is by the use of satellite information of the given area. Despite the long and rich history of land use classification [1,2] and detailed urban land use land cover classification [3,4], we identified room for improvement for several types of areas in Sweden. Therefore, in this Task, we want to put a stronger focus on the classification of different types of areas from Swedish satellite imagery.

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 classifying areas of 500 meters by 500 meters into different categories. Here, the appropriate categories to be used is to be identified by each team. For this they can take use of the scene classification product of Sentinel-2 [5] that is also available in the Open Data Cube, as well as any outside resources that are available to the public (i.e. other teams also have the opportunity to use).

2) Competing teams train and evaluate their models for the classification of 500 meter by 500 meter areas.

Task Description and Data

Data to be used: satellite imagery data of Sweden that will be available to be queried through the OpenDataCube environment.

Expected outcome and evaluation metric: the technical part of 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
2. The validity and strength of argument provided for the choice of land use classes, and utility of their solution
3. The adequacy of the datasets created for the task
4. The technical difficulty and originality of the presented solution
5. The quality of the presentation
6. 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 aspects 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] Ünsalan C., Boyer K.L. (2011) Review on Land Use Classification. In: Multispectral Satellite Image Understanding. Advances in Computer Vision and Pattern Recognition. Springer, London

[2] Gavade, Anil & Rajpurohit, Vijay. (2019). Systematic analysis of satellite image-based land cover classification techniques: literature review and challenges. International Journal of Computers and Applications. 1-10. 10.1080/1206212X.2019.1573946.

[3] Chapa, Fernando & Hariharan Sudha, Srividya & Hack, Jochen. (2019). A New Approach to High-Resolution Urban Land Use Classification Using Open Access Software and True Color Satellite Images. Sustainability. 11. 10.3390/su11195266.

[4] Cai, G., Ren, H., Yang, L., Zhang, N., Du, M., & Wu, C. (2019). Detailed Urban Land Use Land Cover Classification at the Metropolitan Scale Using a Three-Layer Classification Scheme. *Sensors (Basel, Switzerland)*, *19*(14), 3120. https://doi.org/10.3390/s19143120

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