Deep Learning Bottom-of-Atmosphere Correction and Cloudless Vista_S2-L2A

ClearSKY Vision

ClearSky Vision x ESA Network of Resources

Company Introduction

ClearSky Vision is a startup located in Denmark with a vision of increasing data availability and usability in earth observation. That has resulted in two technologies that are further explained to the right ->

Our focus has also been on developing products using Sen2Cor, however, our deep learning algorithms are versatile, and we are now testing this on other data types and sources.

Rapid BoA

A deep learning approach to escalate bottom-of-atmosphere correction in earth observation by estimating the final images instead of doing the actual calculations. It can, compared to Sen2Cor, produce the image x180 times faster (in 10 seconds) with 99.7% accuracy.

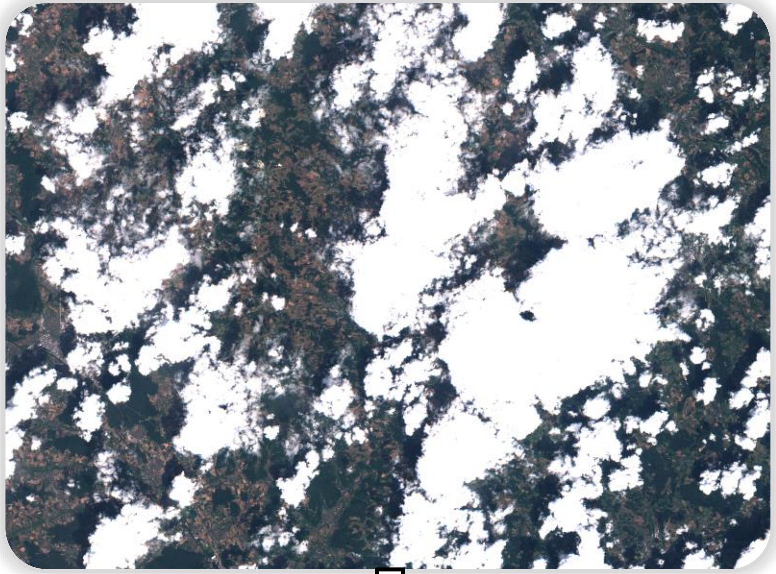
Cloudless Sentinel-2

A deep learning approach to remove clouds and shadows from Sentinel-2 imagery by combining satellite data from numerous constellations into one daily, cloud-free image. The deep learning model is primarily developed for agricultural monitoring and uses long time-series of images to estimate daily changes on the ground.

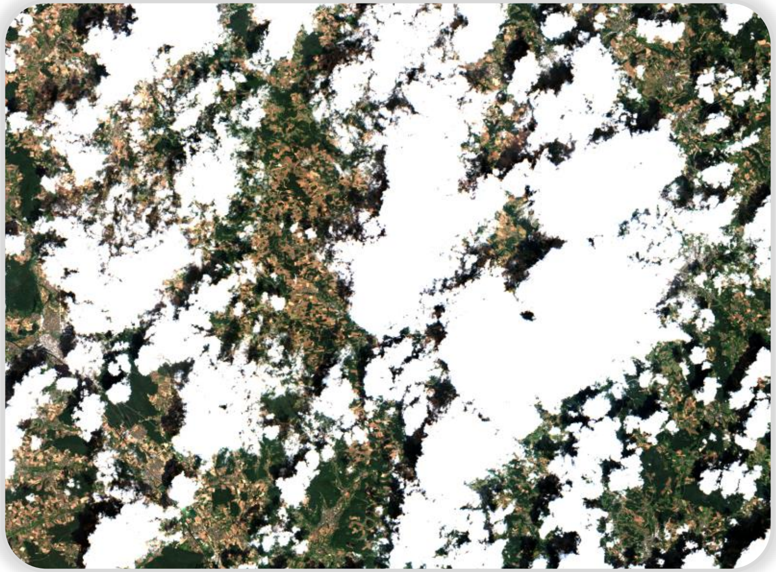
Project Introduction

- Doing bottom-of-atmosphere (BoA) correction is a necessity which on Sentinel-2 imagery is usually done with Sen2Cor. However, this project will explore “Vista_S2-L2A” on the Food Security Platform (TEP).
- The goal is determining the feasibility of accelerating the BoA algorithm with deep learning models.
- The NoR grant is necessary to get access to the data.
- Due to the huge amounts of data needed for the cloud removal algorithm, and some problems getting access to enough data from Vista (we need long time-series of all data for the model, even fully clouded images, which Vista does not produce), the cloud removal algorithm has not yet been trained on Vista_S2-L2A. However, the Rapid BoA algorithm brings the possibility one step closer.

Top-of-Atmosphere



Bottom-of-Atmosphere



Project Roadmap & Objectives

- Use The Food Security Platform (TEP) to access thousands of Vista_S2-L2A tiles.
- Train a deep learning algorithm to estimate Vista Vista_S2-L2A images.
- Compare results from Vista_S2-L2A to Sen2Cor and how it affects the end-user application.
- The objective is to determine the feasibility of estimating Vista_S2-L2A with a deep learning model.
- It is also an objective to determine any special hardware requirements for running the algorithms (e.g. GPU).

Objectives: The objective of this project is two-fold and the requested data can be used for both tasks while testing processing capabilities on The Food Security Platform (TEP). Firstly, we will demonstrate that it's possible to do bottom-of-atmosphere (BoA) correction on Sentinel-2 Reflectance at Bottom of Atmosphere / VISTA Algorithm (available on TEP as 'Vista_S2-L2A') using deep neural networks. We estimate that this can improve processing speeds by x100 to x500 while keeping accuracy high. This is inspired by an existing algorithm, developed for another project, that in production as a side effect efficiently fixed incorrect Sen2Cor bottom of atmosphere correction. This is in particular interesting on, important and frequently used algorithms with long processing times like BoA processing algorithms. The results will be avg. pixel error measured against ground truth imagery. We will also present the relevant processing speeds improvements and requirements to run said algorithm (eg. GPU accelerated processing). The results will be made available on TEP as ClearSky Vision demo data, and if possible produced on TEP. It will, furthermore, be measured against data in-sample and out-of-sample, and the project will be finished off by producing a tile unavailable on the platform. This project has the potential capability of greatly reducing required resources for BoA correction on Sentinel-2 imagery by doing it in a fraction of the time (leaving data storage as the final limitation). Not only making it a fast and efficient process but it also makes near-real time monitoring more achievable.

ClearSky Vision has already developed an algorithm for cloudless Sen2Cor imagery (using deep learning and multiple satellites for data fusion). This approach won ClearSky Vision the Copernicus Masters BayWa competition in 2020. It combines Sentinel-1, Sentinel-2, Sentinel-3, and Landsat 8 into daily cloudless Sentinel-2 imagery. This project will further prove, to what degree cloudless results on Vista_S2-L2A will match the accuracy from prior cloudless Sen2Cor imagery tests. The objective is to determine whether this (more complex) processing method will make the cloudless process more difficult or what's more likely, improve the consistency of the output. The results will be made available on The Food Security Platform as ClearSky Vision demo data (10 spectral bands).

Methodology: We will use The Food Security Platform (using API access) to acquire ~20,000 Sentinel-2 tiles processed by Vista. We will start by downloading already processed tiles (~2000 tiles available) to develop a simple prototype to prove it's possible while keeping costs low. Afterwards we will pick strategic geographical places (across EU) to process with Vista's algorithm on TEP (up to 18,000 tiles). We will search our own internal database to identify cloudless Sentinel-2 imagery, and match it against the available data on TEP by date and coordinates. This should yield the data needed for the cloudless development and accompanying tests (this approach needs very clean/cloud-free data). We will, furthermore, download random clouded data needed for BoA correction as well as download Level-1 Sentinel-2 data from Google's storage. The data will mostly be from Bavaria and surrounding areas (much of this data is already processed on TEP and will cost less to use). This data will be used as our ground truth to develop the above mentioned algorithms. If the BoA project is successful the neural network will learn to produce Level-2 data from Level-1 input (in this case Sentinel-2 Reflectance at Bottom of Atmosphere / VISTA Algorithm). If the cloudless project is successful we will have developed a limited cloudless prototype that's expected to work in areas near Bavaria (exactly how narrow an area can be determined afterwards and mostly determined based on available training data).

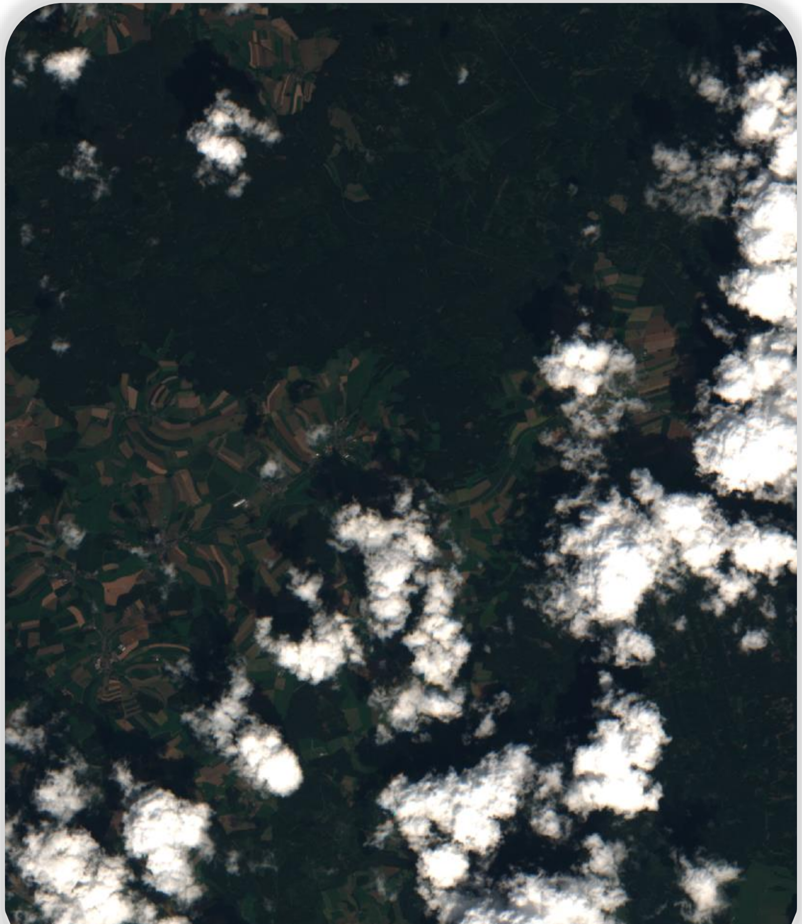
We estimate that this project requires 2,000,000 sq. kms. of imagery (the project is flexible on dataset size but more processed data equals better results). All the Copernicus imagery needed for the cloudless algorithm (as input) is already located on ClearSky Vision's servers. This funding request is about gaining access to the already 2,000 available tiles while producing up to 18,000 new Sentinel-2 tiles using Vistas algorithm, limited data storage, and a premium user subscription for interacting with support for when the algorithm is ready to produce data on TEP. Using deep learning for this purpose requires millions of small images (128x128 px), and this is why we are applying for 4,900 euro worth of credits for Sentinel-2 data processed by Vista including a premium user subscription.

Visual Results

4 hours to produce

10 seconds to produce

Level-1c



Vista_S2-L2A

Rapid BoA

Accuracy Results

Value Range: 0-12000

- The model has been validated on out-of-sample data in close proximity to the training data from Central Europe (~99.85% accuracy).
- The error rate is consistent with Sen2Cor results, and the images have been produced equally fast. It seems like Sen2Cor and Vista_S2_L2A is a comparable problem for the deep learning model.
- The spectral band error in red and near infrared results in 0.25% error when measured in NDVI. Meaning it's on an average 0.005 NDVI values off the true target which is negligible for most applications.
- The deep learning algorithm can be run on cheap cloud GPUs, making it TEP capable.

Pixel Error	B2	B3	B4	B5	B6	B8	B8a	B11	B12
Percent	0.13%	0.13%	0.12%	0.14%	0.13%	0.15%	0.18%	0.07%	0.06%
Values	15.6	15.8	14.1	16.4	16.1	18.3	21.6	8.0	7.0

Conclusion

1

The initial results shows that the data can be produced equally fast with Vista_S2-L2A which makes it x720 faster.
The initial accuracy results also show that the error rate is on par with the baseline from Sen2Cor even though the problem is more "difficult".

2

The societal benefits are increased data availability and overall faster data delivery. This has the capabilities of providing researchers with data that could otherwise be difficult to obtain or at least slow for the researcher to obtain.

3

The overall results are promising as the resulting error in NDVI are very low. This project proves for us that the deep learning algorithm has the capabilities to solve a wide range for computational problems in earth observation as it can do Vista_S2-L2A as well as Sen2Cor.