### Contrasts and Comparisons between Mosaic and KELP Architecture

Xi Sun V00025479

Ting Liu V00045129

Zheng Yin V00915261

In the original architecture and implementation of the Phase I of our project KELP, we took advantage of Arbutus supercluster as a storage and a preprocessing and post-processing platform of Sentinel 2 imagery data. We relied on the cloud-based platform Google Colab for resource computation to calculate the NDVI which will be used to indicate the distribution of sea vegetation.

In Phase II of the project, we will consider if our stakeholders have proprietary data sets and algorithms for their specified purposes, like the iceberg detection, wildfire protection, etc. Especially, when the data sovereignty comes to some special communities, a centralized database will become insufficient and no longer credible. We need to rethink our architecture from technical and societal perspectives.

In the first part of Phase II, we worked on the mosaic platform which was developed by EarthDaily and can provide satellite imaging, data services and geoanalytics. We accessed a proprietary data set from Earth Daily Analytics and used Earth Daily Mosaic to calculate the NDVI based on the same AOIs as our project.

Mosaic can be run as a Jupyter Notebook service locally. We access the hyperlink of <https://mosaics-preview.earthdaily.com/home> to fetch the raw data into the notebook and run the computation to demonstrate the NDVI values and histogram map of selected area.



Figure 1: Area of Interest (Tofino, BC, Canada)

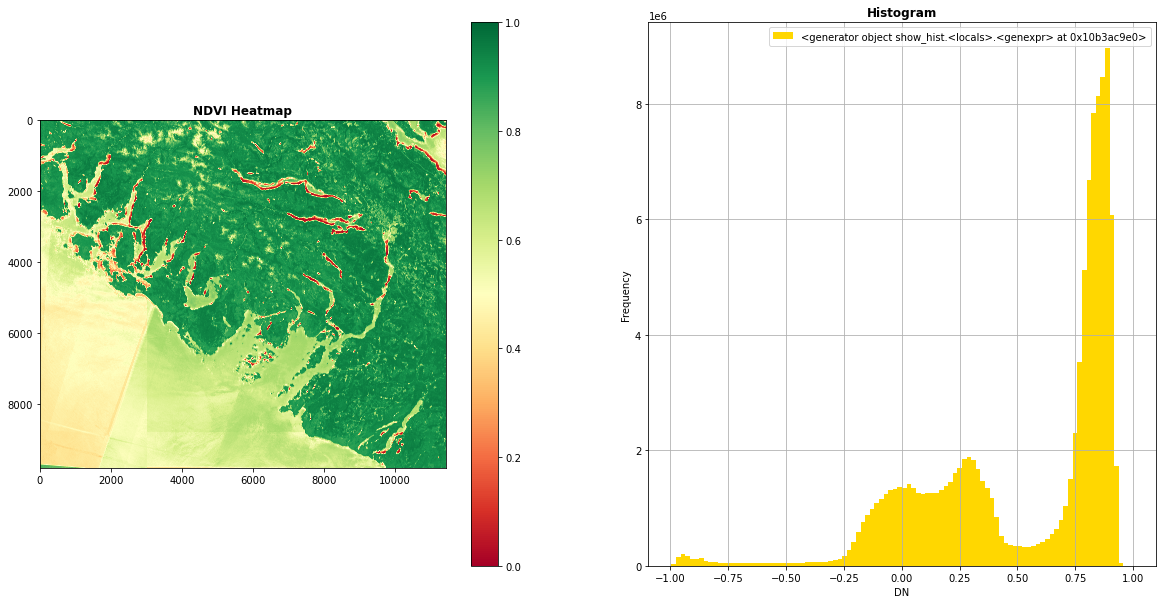


Figure 2: NDVI Heatmap calculated by Mosaic, the red area indicates the distribution of sea vegetation

In our scene-by-scene processing of Sentinel-2 data, we chose the similar area from the Sentinel 2 data storage in Arubus and calculated the NDVI on Google Colab.

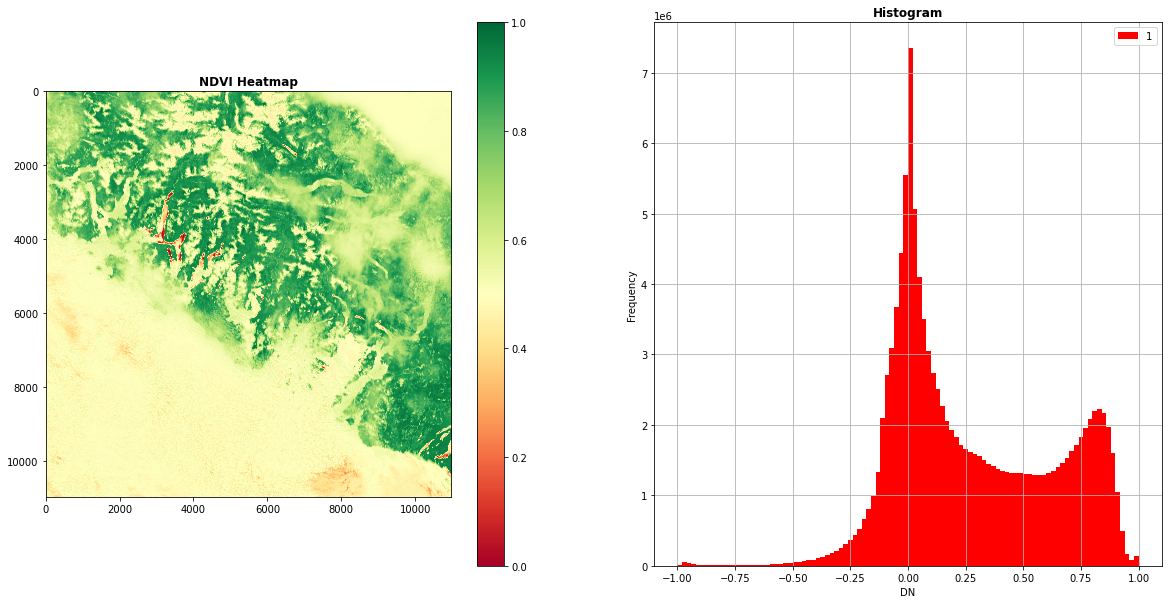


Figure 3: NDVI Heatmap calculated by KELP, the red area indicates the distribution of sea vegetation

The two heatmaps look very similar, But due to the different data sources, there are still some differences in the processing procedures and final results. Now we are going to do some contrasts and comparisons.

### 1， Data sources

In Earth Daily Mosaic, the data source was stored in Amazon AWS as a private repository, the data has to be fetched using a token. The sample data was the 2018 and 2020 surface reflectance data which can be compared side by side on Earth Daily Mosaic webpage. We can get the hyperlink from the webpage and setup in jupyter notebook to fetch the data by manually selecting the AOIs.

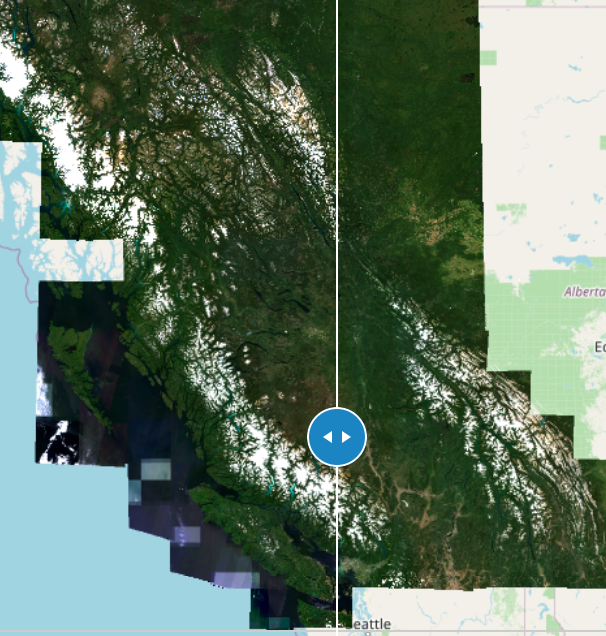


Figure 4: 2018 (left) and 2020 (right) surface

reflectance data side by side comparison

In our KELP project, the data source is sentinel2 imagery which has been pre-downloaded from the ESA hub. Unlike Mosaic, the AOIs have to be defined before the downloading. However, we can specify the date of the data, so we can calculate and look up the NDVI on different dates and compare the changes over a year.

For Earth Daily Mosaic, we have no idea if we can select the specific date of the satellite imagery. But from the samples, we didn’t see any configuration to fetch the daily data.

### 2， Data transmission

The data in Earth Daily Mosaic was downloaded from AWS as a whole. All bands including BLUE, GREEN, RED, NIR were downloaded into a file, the source is a Cloud-Optimized Geotiff, so we can use GDAL to automatically choose the most appropriate subsampled overview level for the specified output area and only read the regions specified by the selected AOI. Because the source data contains too much information, the data transmission time was very long. It took 570 seconds to download the area shown in Figure 1.

In our KELP architecture, we only downloaded the needed files into the computation environment. In this case, we just need band 4 and 8 JPEG files to calculate NDVI, so we specified the file suffix before downloading source data from Arbutus to Google Drive storage. So the transmission time of the select area as shown in Figure 1 was significantly reduced. However, it also depends on how many dates of imagery we are supposed to process. For one day processing, the downloading time was about 20 seconds. For more dates, the time increased linearly.

### 3， Data Processing

The formulas of NDVI calculation were the same on both platforms.

In Mosaic, the data was downloaded as geotiff files and read by GDAL to calculate the NDVI of the selected area. The data processing time was 37 seconds.

In KELP, we use the rasterio package to read the JP2 files, and then calculate NDVI. The NDVI was formatted to a geotiff file with their original geoinfo. The total processing time of two tiles was 72.18s. So in terms of the processing time, there is no obviou performance gap between two platforms.

By the way, in this comparison, we did not consider the power difference of the local machine and Google Colab virtual machine.

### 4， Data Result

In Mosaic, after the NDVI was calculated, we can not see the further interaction with the Mosaic platform. For example, in the Mosaic example, it provides 2018 and 2020 surface reflectance data, but we do not know how to recall our results to the webpage and compare the results calculated from these two data sources. We are looking forward to further instruction of Mosaic.

In our KELP platform, the results would be uploaded back to Arbutus. In Arbutus, we use SNAP to post-process the result for further analysis. For example, the result file contains NDVI values of all pixels. When we just need to investigate NDVI on the sea, the other values would be annoying. So we have to filter these values out, we can use the SNAP tool to mask all lands, just render the values with water covering.

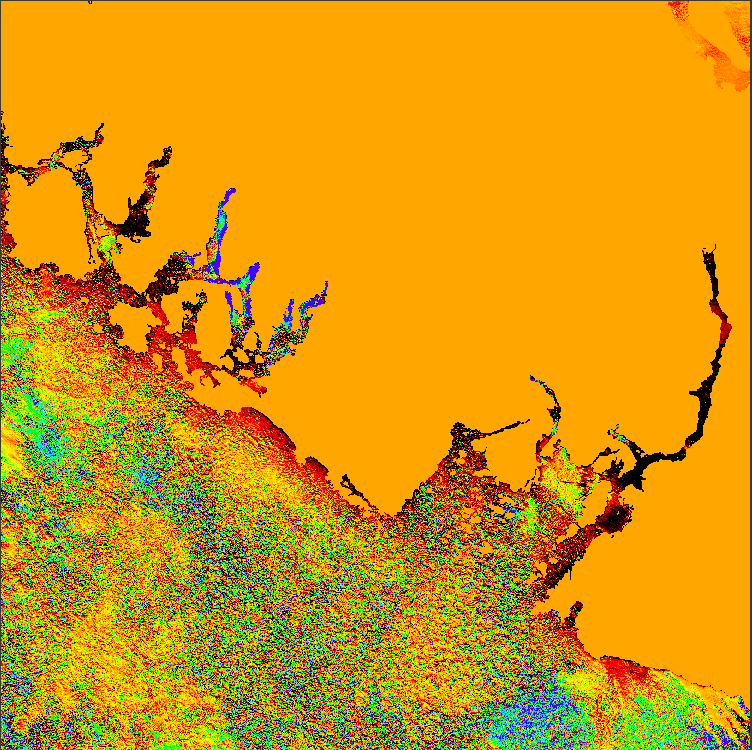
****

Figure 5: Land Mask of NDVI Result

Next, we will figure out more features and applications of both Mosaic and our KELP platform. Also, we have to think about how to optimize and use these two platforms for proprietary data sets and algorithms for the next part of Phase II.