COCKTAIL

GDAL, OTB and QGIS in a virtual environment for resource constrained collaboration

Last update: Nov 16, 2022

Cocktail is a collaboration utility to facilitate collaborative remote sensing data collection, analysis and evaluation under constrained resources. Cocktail is designed to facilitate resource-constrained collaborative inquiry on optical satellite assets, unlike the EOreader that also processes SART constellation data.

Collaboration is supported because Cocktail is a lightweight package that will run an even modest cloud-based Linux systems (16CPU and 32GB RAM[1]). Cocktail will vary parameters of a classifier in batch mode and store the classifier performances for each setting so you can collaboratively can find the best configuration. Cocktail keeps track of the large collection of parameters across the individual processing modules in a single file such that remote collaborators can reliably replicate a workflow. Moreover, Cocktail can be linked with an external cloud storage provider to move assets out of the compute environment for low-cost storage. The code repository implements a solution for this issue using pCloud. Other providers can likewise be attached.

Cocktail is designed as a compendium to other GIS investigation environments, combines GDAL, OTB and GQIS elements, and is intended to be deployed remotely. The combination requires some attention.
The Orfeo machine learning library and the translator library for raster and vector geospatial data formats GDAL do not play well with the open source geographic information system QGIS outside of a resource intensive GUI environment. Cocktail allows you to setup a low-cost virtual computer that can support GDAL, OTB and QGIS functionality and how to apply various classifiers including vector support machines, random forests, neural networks and various band math operations across GDAL, OTB and QGIS.

Private sector, high-resolution, daily-updated satellite assets have become a significant resource for remote sensing operations. A case in point is PlanetScope (PS), currently operating the largest collection of small Earth-imaging satellites. Cocktail facilitates the combination of free and commercial satellite assets to monitor an area of interest with a low-cost system, and then switches to a high-resolution asset only if a condition of concern or interest has been detected and not adequately understood in the first data set. Not unlike the tip and cue concept in which one monitors an area of interest with one sensor and then 'tips' another complementary sensor platform to acquire another image over the same area, Cocktail allows you to tip and cue based on economic constraints.

Here is a brief description of the steps required to get started:

1 - VM
Create a virtual computer.
Choose Ubuntu 18.04 LTS, with at least 16GB of RAM (test system: 8CPUs and 32GB RAM)

2 – Get the files from GitHub

git clone https://github.com/realtechsupport/cocktail.git

3 - Install QGIS
Make the file basics+qgis.sh executable.

chmod +x basics+qgis.sh

Run that file
sh basics+qgis.sh

[1] Also tested with only 8GB of RAM. Rather slow; otherwise fine.

4 - Install OTB in a virtual environment with conda
Make the file conda-packages.sh executable.

chmod +x conda-packages.sh

Run that file

sh conda-packages.sh

5 - Create a conda environment with defined dependencies.
(OTB 7.2 requires python 3.7, for example)
Create a conda environment with the settings in the environment.yml file.

conda env create -f environment.yml

6 - Customize
Add other libraries to the OTB environment as needed

conda install -c conda-forge pillow

(Enable the OTB environment to install geojson and geopandas.
The geopandas install may take some time…)

conda install -c conda-forge geojson
conda install -c conda-forge geopandas
conda install -c conda-forge sentinelsat

7 – Setup and check the directory structure
After cloning the repository, your directory structure should look like this:

```
|-- miniconda2
|-- cocktail
        |-- code
        |-- data
                |-- auth
                        pcloud_auth.txt
                        sentinel_auth.txt
                        planet_auth.txt
                |-- collection
                |-- rasterimages
                        sentinel2
                        landsat8
                |-- rawsat
                |-- vectorimages
                        |-- roi
                settings.txt
                somefile.geojson
                colormap.txt
                colormap.qml
        |-- results
        |-- setup
        README.md
        LICENCE
        Install+Use.pdf
```

Then adjust the paths to reflect your current installation. From the setup folder run the adjust_datapaths script:

python3 adjust_datapaths.py

Follow the prompt. All paths will be adjusted to your current installation and the settings.txt file will be updated accordingly.

Put analysis-ready raster and vector files (including shapefiles) into the data/collection directory. Use the settings.txt file (in the data directory) to set your file names, process preferences and classification parameters. The content of this file is parsed and passed to the classification steps.

Change the fake username and passwords in the auth folder according to your needs. PC.txt is for pCloud, sent.txt for sentinel and planet.txt for Planet Labs. Put data from sentinel and planet into the collection directory. Add .geojson references files (for sentinel and planet downloads) and colormaps as needed to the data directory.

8 - Test

Adjust the parameters in the file data/settings.txt to work with your data directory.

a) Test QGIS

Run the qgis test in the base environment.

python3 qgis_test.py

You should see a list of operations supported by QGIS.

b) Test OTB

Activate the OTB environment.

conda activate OTB

Run the otb test

python3 otb_test.py

You should see a list of operations supported by OTB.

c) Update settings

Update the settings and file parameters (in settings.txt) corresponding to paths on your VM.

nano settings.txt

Then verify that all is ok

python3 settings_test.py

You can now use QGIS in the base installation and enable the conda environment to access OTB functionality. You can also move across environments to access libraries from either environment with python scripts as outlined below (see vector_classify_top.sh):

```
echo "Starting OTB-QGIS pipeline...\n\n"
conda run -n OTB python3 /home/code/otb_code_A.py
python3 /home/code/qgis_code_A.py
conda run -n OTB python3 /home/code/otb_code_B.py
python3 /home/code/qgis_code_B.py
```

Since the directory structure you setup will be identical in both the QGIS and OTB environments, intermediate filed produced will be available to processes running in either environment, allowing for data to be shared at no extra computational cost. All settings across the script modules are stored in the 'settings.txt' resource file and imported to the individual modules.

After installation and testing you can perform operations as such (provided you have already prepared your data accordingly.)

If you are using pCloud to move results from the VM to the pCloud server, install the library in the base and in the OTB environment with the correct python version (here 3.7).

```
pip install pcloud

conda activate OTB
python3.7 -m pip install pcloud
```

Working with Cocktail

All examples below assume that you set your input and process parameters in the settings.txt file.
Make sure to adjust the path to the settings.txt file listed in each of the scripts to suit your setup.

> Place all your assets – raster images and shapefiles – into the collection directory <

Get free satellite imagery from the European Space Agency
Perform steps 1-4 once.

1) Create an account with Copernicus Open Access Hub (https://scihub.copernicus.eu/dhus)
You will need a login and pswd to access sentinel2 files
2) Generate a geojson file for area of interest (https://geojson.io/)
Save the file as sat_roi.geojson and place a copy on into your data directory.
3) Put the ESA credentials in a file in the auth folder.
4) Generate a shapefile with a specific area of interest (ROI) to clip the asset to a specific area.
Place that file in the collection directory.

Repeat this next step periodically (Sentinel2 has a combined revisit frequency of 5 days).
You can also call it via a cron job.

4) Run the script:

python3 sentinel2_getdata.py

Enter choices at the prompt
Inputs: bandselection, now, uuid

If the second input is empty, the end date in settings.txt will be used.
You can cue on the sentinel asset and simply get the latest available summary image (TCI) to see if the data is useful.

To do this, type

tci now

If you have setup pCloud (or someother provider) and the TCI passes a simple statistical test (see code for details), that TCI will be saved to pCloud. If that image suggests good data, get all bands with these inputs:

all now

Alternatively

tci

Will get the last TCI between dates set in the settings.txt file

all

Will get all bands between dates in the settings.txt, and

71d7edef-292c-4f4e-af40-72c3f0e9eb64 (sample)

Will get all bands of a unique identifier (uuid)

Sentinel's SAFE format with the JPEG2000 images will be converted to geotif (.tif) will be clipped with the specified region of interest specified in the settings.txt file, compressed (.zip) and saved to the collection directory (area2_0726_2021_sentinel2.zip, for example).

If you selected 'tci', and the operation was successful, you should see a message like this:

```
Image passes test. Percentage of non-black pixels: 66
Sending to pCloud..
2022-07-07 20:09:46,697 - pycloud - INFO - Using pCloud API endpoint: https://api.pcloud.com/

Uploaded:
['/home/blc/cocktail/data/rawsat/S2B_MSIL1C_20210706T021609_N0301_R003_T50LKR_20210711T143413.SAFE/GRANULE/
L1C_T50LKR_A022623_20210706T022834/IMG_DATA/T50LKR_20210706T021609_TCI_roi.jpeg']

Adding log entry...
Deleting temporary files to save storage space.
```

You should also have the reduced size .jpeg image in your cloud storage as a reference.
Depending on the area you select and your cloud coverage threshold, you might not find any good data within a given time window. Vary your settings. Also, ESA limits the data you can download. Caveat.

Once you do get a good test image, download and save the complete package (often several hundred MB) by running

python3 sentinel2_getdata.py
with the prompts
all now

… as mentioned above. You can use the result for band operations or classifications (see below).

However, the resolution of the sentinel data may be insufficient for your needs – in that case you would seek an alternate source such as Planet Labs. Finding viable free data (tip) can will reduce the number to times you grab non-free higher resolution data from a private provider (cue). Use the same ROI file and the date of the downloaded TCI to get data only on dates where you area is clearly visible and accessible. That is how the tip and cue concept is translated in this context.

Get satellite imagery from Planet Labs

If you have an account with Planet Labs, add the API key to a file in the auth folder. To access the Planet Labs imagery, you need that key. Then run the two scripts below:

python3 planet_get_previews.py

Get thumbnail size previews of a scene based on AOI, start date, number of days and cloud cover. Follow the prompt to enter an AOI (the geojson file), a target date, the number of days to extend the search and the maximum cloud cover acceptable. If you have external storage enabled (pCloud), the thumbnail images will be copied there and you can preview them. To download the actual image, do this:

python3 planet_place_order.py

This script will download the geotif image, based on the image id identified in the preview, collected with the planet_get_preview module above. At the prompt, enter the image ID and the AOI. The image and metadata will be saved to the rasterimages directory. A copy of the geotif file will be transferred to external storage, if enabled.

In spring 2022, the 4-band Dove satellite configuration was decommissioned and replaced with the 8-band SuperDove configuration. Requests for assets with dates newer than April 2022 will access the SuperDove set. The 8-band datasets will be saved as one, large (in excess of a GB, depending) multiband geotif image.

Certainly this tip and cue approach is a poor-man's approach to minimizing the use of private industry sensing assets, and it may miss some opportunities as the Sentinel revisit frequency is 5 days compared to the 1day rate of Planet Labs. Yet for many applications, such as monitoring change in vegetation growth or urban development, this difference might not matter.

Clip a collection of satellite image raster bands to a specific region of interest

1) Get your choice of satellite imagery (Sentinel2, Landsat, Planet) and, convert to .tif (see above), zip the individual bands together to one file (say sat_bands.zip)
1b) Or just identify a single image / band that you want to clip.
2) Create an ROI file that is contained within the extent of the satellite imagery. Geojson.io is a useful tool for this step (say sat_clip.zip).
3) Move the sat_bands.zip and sat_clip.zip files to the collection directory
5) Update the settings.txt file with the name of the sat_clip.zip file
6) Activate OTB environment:

conda activate OTB

6) Run the python script

python3 otb_clip.py

Follow the prompts to input either your zipped raster bands or an single band. If you operate on a single input, enter the location of that file at the prompt.
Inputs will be clipped with the file specified in the settings.txt file. If you input a raster band collection, all resultant bands will be zipped and saved in the rasterimages directory.

Normalized Difference Built-up Index (NDBI) on Sentinel and Landsat series

1) Activate OTB environment

    conda activate OTB

2) Place a zipped folder with all bands available into the collection directory. Or just use the sentinel2_getdata.py for sentinel2 assets.
(Naming convention: area_monthday_year_satllite,zip > area2_0726_2021_sentinel2.zip)

3) Run the program

    python3 otb_ndbi.py

Follow the prompts to enter a satellite source. The script will perform the Normalized Difference Built-up Index (NDBI) for urban development on the asset.

NDBI difference map on Sentinel and Landsat series

1) Same requirements here as items 1 and 2 above.

2) Run the program

    python3 otb_difference_ndbi.py

Follow the prompt to enter two zipped satellite asset collections from different times and the same source, collected with the script sentinel2_getdata,py. The script will perform the Normalized Difference Built-up Index (NDBI) for urban development on each asset, calculate the difference between the results and then overlay that difference data on the NIR band of the newer image. Change condition between images thresholded at midpoint between 0 and 255 - 128).

In both of the NDBI operations, the final image will be sent to external storage, if enabled in the settings.

Pixel based classification (SVM or RF)

These routines assume you have created adequate training samples in a corresponding shape file.
Update the settings.txt file if you want to change SVM or RF hyperparameters..

1) Put all inputs (raster images and shapefiles) in the collection directory.

2) Enable the OTB environment:

    conda activate OTB

3a) Run the program

python3 otb_raster_classify.py

Follow the prompts to set the input .tif image and the classifier choice (Random Forest [rf] or Support Vector Machine [libsvm])

area2_0612_2020.tif rf

3b) Run the related script that generates Haralick texture features derived from the image information and concatenates that information with the input image. Texture information can improve classification results on some images.

python3 otb_raster+texture_classify.py

Follow the prompts to set the input .tif image and the classifier choice (Random Forest [rf] or Support Vector Machine [libsvm]) as above.

4) You can also vary the parameters of the classifiers specified in the settings file and batch execute either rf or libsvm classifiers for each of the possible combinations with a "script of scripts". Edit the file otb_vary_raster_classify.py to set the parameters you want to include the approach (use texture or not) and the classifier (rf or libsvm).
Then run the script:

python3 otb_vary_raster_classify.py

This script pauses after each iteration to ensure that the timestamps (1minute resolution) generated for the image results are unique. Results from all permutations of the selected parameters under the chosen classifier will be saved to the results folder. If you installed pCloud and enabled uploading, these results will be saved to that environment.

Object based classification (ANN)

These routines assume you have created adequate training samples in a corresponding shape file.

1) Deactivate the OTB environment first

conda deactivate

2) Open the settings.txt file and change the parameters “raster image” and “pointszipfile” (the zipped vector shape file) to your data.

3) Run the program:

sh vector_classify_top.sh

This script runs all the processes across multiple otb and qgis scripts to perform the object-based classification.

Enabled by default in the settings.txt file, summary statistics (precision, recall, f1-score) of the classifier performance across all categories are calculated, saved for reference (and transferred to the cloud, if enabled). This is important for comparing classifier results in remote collaboration settings.