Using Jupyter to Create a Community for Satellite Imagery Analysis and Sharing

DigitalGlobe
Sean Gorman PhD.
@seangorman
Our Jupyter Journey
What are we going to cover?

1. Our problem
2. Why Jupyter
3. Accessibility
4. React UI/UX
5. Scaling
6. Community
1. The Problem
DigitalGlobe collects 3,000,000 km² a day and has a 17 petabyte archive of imagery.
Current WorldView constellation is in sun-sync orbits

Future constellation mixes sun-sync and mid-latitude orbits for greater revisit
2. Jupyter
```python
def extract_photo(msg):
    recs = [rec for rec in msg['entities']['media'] if rec['type'] == 'photo']
    output = recs[0]
    output['screen_name'] = msg['user']['screen_name']
    output['text'] = msg['text']
    del output['sizes']
    del output['indices']
    return output
```
Get Task Definition

Get a Task Definition from the Task Registry

https://geobigdata.io/workflows/v1/tasks/task_id

PATH PARAMS

| task_id | String identifier of the task. For example, Getting_Started |

Get a Task by Name and Version Number

Retrieving a task by name returns the latest registered version of that task. To get a specific version of a task, include the version number in the request.

The request will look like this:

https://geobigdata.io/workflows/v1/tasks/name:version

- Note: The task name must match the task registry or the request will return a 404 error.
Why Jupyter

- Muhammad must go to the mountain
  - Don't reinvent the wheel
  - Go to where the users are
- Open Source
- Go-to tooling for data scientists
- Over 3 million users
Extending Jupyter

3. Accessibility
Satellite Imagery is Big and Awkward

- A high res scene of satellite imagery = 40 gb
- Standard photo imagery is three band (red, green, blue = RGB)
- DigitalGlobe has eight bands
- Imagery needs to be atmospherically compensated
- Imagery needs to be color balance (dynamic range adjustment)
Absolute
Replaces the pixel values of an image by their absolute values.

Acomp
Apply ACOMP to a 1B image.

Add
Adds two images.

AddCollection
Adds a collection of images.

AddConst
Adds constants to an image.

ID: buildings-v2
Source: imagery
SegmentationType: pointOpBased
Mode: segmentation
SagemakerEndpoint: building-segmenter-v2
```python
from gdxtools.task import env
from gdxtools import CatalogImage
#matplotlib inline

catalog_id = env.inputs.get('catalog_id', '1040010041DB800')
bbox = env.inputs.get('bbox', '-122.4418330192566, 40.61330164978256, -122.42888230830432, 40.62104023560901')

image = CatalogImage(catalog_id, band_type="MS", blm=True, acom=True, pansharpen=True, bbox=map(float, bbox.split(',')))
image.preview()
```
Image Science Tutorial Notebooks

#1 Introduction to GBDX note...
By steve.roast@digitalglobe.com
This is first time beginner's guide to noteb...
Details Clone & Open

#2 Ordering and Working with...
By steve.roast@digitalglobe.com
Learn how to find satellite imagery and pul...
Details Clone & Open

#3 Beginner Band Math for Re...
By sean.gorman@timbr.co
This notebook is the third in a series of be...
Details Clone & Open

#4 Introduction to Imagery Ac...
By chris.helm@digitalglobe.com
This notebook is a detailed introduction to ...
Details Clone & Open

#5 Imagery and AOIs
By chris.helm@digitalglobe.com
This notebooks shows how to work with va...
Details Clone & Open

#6 Intro to DG Sensors
By chris.helm@digitalglobe.com
An introduction to various image classes s...
Details Clone & Open
Now that we have an interesting image to work with we'll construct the equation to calculate our NDVI index. Next we need to identify the appropriate bands to calculate our index. In Python we start with 0 and assign bands sequentially across the spectrum:

0** = **Coastal Blue** - Absorbed by chlorophyll in healthy plants and aids in conducting vegetative analysis. Least absorbed by water, and will be very useful in bathymetric studies. Substantially influenced by atmospheric scattering and has the potential to improve with atmospheric correction techniques.

1** = **Blue** - Readily absorbed by chlorophyll in plants. Provides good penetration of water. Less affected by atmospheric scattering and absorption compared to the Coastal Blue band.

2** = **Green** - Able to focus more precisely on the peak reflectance of healthy vegetation. Ideal for calculating plant vigor. Very helpful in discriminating between types of plant material when used in conjunction with the Yellow band.

3** = **Yellow** - Very important for feature classification. Detects the "yellowness" of particular vegetation, both on land and in the water.

4** = **Red** - Focused on the absorption of red light by chlorophyll in healthy plant materials. One of the most important bands for vegetation discrimination. Useful in classifying bare soils, roads, and geological features.

5** = **Red Edge** - Centered strategically at the onset of the high reflectivity portion of vegetation response. Very valuable in measuring plant health and aiding in the classification of vegetation.

6** = **NIR1** - Very effective for the estimation of moisture content and plant biomass. Effectively separates water bodies from vegetation, identifies types of vegetation and also discriminates between soil types.

7** = **NIR2** - Overlaps the NIR1 band but is less affected by atmospheric influence. Enables broader vegetation analysis and biomass studies.

To calculate NDVI we’ll take the red band ‘4’ and the NIR1 band ‘6’ to create our equation. The code for the NDVI equation is shown below.

```bash
In [11]: import numpy as np

red = image[4,:,:]  
nir = image[6,:,:]  
ndvi = np.clip((nir - red)/(nir + red), -1, 1)
```
4. UI/UX
Customizing Jupyter for Domain Specific Use

- Jupyter React as a framework
- Customizing within Jupyter Notebooks
- Customizing the organization of Jupyter Notebooks
Customizing Jupyter
Organizing Jupyter
Explore Notebooks

Published Notebooks

**Imagery and AOs**
by chris helm@digitalglobe.com
Published: 8/23/2018 10:01

This notebook shows how to work with va...

**Introduction to GBDX notebook**
by www.csaat@digitalglobe.com
Published: 8/23/2018 13:39

This is first time beginner’s guide to noteb...

**Beginner Band Math for Remo**
by sean.german@nsmbra
Published: 8/23/2018 10:17

This notebook is the third in a series of be...
5. Scaling
```
In [2]: import numpy as np

def ndvi(chunk):
    return np.expand_dims(np.clip((chunk[6,::] - chunk[4,::]) / (chunk[6,::] + chunk[4,::]), -1, 1), axis=0)

In [17]: processed = image.map_blocks(ndvi, dtype=image.dtype, chunks=(1,512,512))

In [19]: import matplotlib.pyplot as plt

plt.figure(figsize=(60,20))

for i, window in enumerate(processed.iterwindows(10, (512,512))):
    aoi = window.read(quiet=True)
    plt.subplot(2,5,i+1)
    plt.axis('off')
    plt.imshow(aoi[9], cmap='jet')

plt.show()
```
**Inputs**

Inputs are defined directly from your code, and are used as configurable input values to your deployed task on GBDX. These values are what can be passed into the task you create.

<table>
<thead>
<tr>
<th>Variable Name</th>
<th>Type</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>approximate_only</td>
<td>String</td>
<td>required</td>
</tr>
<tr>
<td>catalog_id</td>
<td>String</td>
<td>required</td>
</tr>
</tbody>
</table>

**Outputs**

Outputs are string values from your code that you want to make accessible to other tasks within GBDX workflow chains. Note: Everything you write to disk during execution of the notebook will be saved as standard output directory called `task_output`.

- `out_geojson`
  - GeoJSON

**Task Information**

- **Name**: coastline_extraction_[my initials]
- **Description**: Extract coastlines from an image
- **Version**: 0.0.1
- **Instance Type**: default

**Outputs**

Outputs are the results from your tasks. You can explore outputs below and inject data into notebooks for use in an analysis.

<table>
<thead>
<tr>
<th>Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>out_geojson</code></td>
<td>coastline_1040010057E7D00.geojson</td>
</tr>
</tbody>
</table>
5. Community
Building Community with Notebooks

- Notebooks publishing and role based access control
- Notebook cloning
- Community tier and open data
Carr Fire Burned Building Detection

Published to: https://notebooks.geobigdata.io/hub/notebooks/5b63a1e42486966ea89b7e9c

Metadata

**Title**
Carr Fire Burned Building Detection

**Description**
This notebook uses a modified building footprint segmentation model to detect buildings that have been burned by the Carr fire by doing footprint segmentations before and after the fire.

**Tags**
carr fire, footprints, wildfire, segmentation

Permissions

Controlling who can use your published notebook is quick and easy. Set each notebook to public or private. ‘Public’ allows any registered user to view and reuse your notebook. ‘Private’ will keep this notebook private to you and anyone you choose.

- Private
- Public

Your notebook will be available for any user to view, and use. Please note this also means your notebook will be shared under an open source MIT license as stated in the terms of service.
Published to: https://notebooks.geobigdata.io/hub/notebooks/5b63a1e42486966ea89b7e9c

Apply updates to your published version of this notebook.

Update your published notebook

Updating your published notebook will reflect any adjustments made to metadata and permissions as well as any edits or additions made to your notebook cells.

Metadata

Title
Carr Fire Burned Building Detection

Description
This notebook uses a modified building footprint segmentation model to detect buildings that have been burned by the Carr fire by doing footprint segmentations before and after the fire.

Tags
carr fire, footprints, wildfire, segmentation

Permissions

Controlling who can use your published notebook is quick and easy. Set each notebook to public or private. ‘Public’ allows any registered user to view and reuse your notebook. ‘Private’ will keep this notebook private to you and anyone you choose.

- Private
- Public

‘Private’ notebooks will only be available to the organizations and teams that you choose. They will not be available to the other members. We’ve imported your Github teams to make sharing even easier.

Expecting some GitHub groups that aren’t listed here? Try refreshing.

Organizations
- DigitalGlobe
- timbr-io

0 Teams

Cancel Update Published
Creating Training/Test Data with Image Labeling

Publisher: michael.gleson@digitalglobe.com

In this Notebook, we demonstrate some basic functionality and workflows that can be used for small scale labeling of satellite imagery. These methods can be used to build up small sets of training or test data to help in algorithm development.

#tomnod  #labeling  #roads  #feature-extraction  #dgn-new-methods-and-models

Sean Gorman

Can the data labeling widget push data to Veda to allow it to be reused as training data?

August 24th, 2018 7:08 AM

Reply  Edit  Delete

Comment

Share
Ikonos, Sentinel2, Landsat, and Event Data
Thank you.

DigitalGlobe
Sean Gorman PhD.
@seangorman