CRISMIS GSOC Project

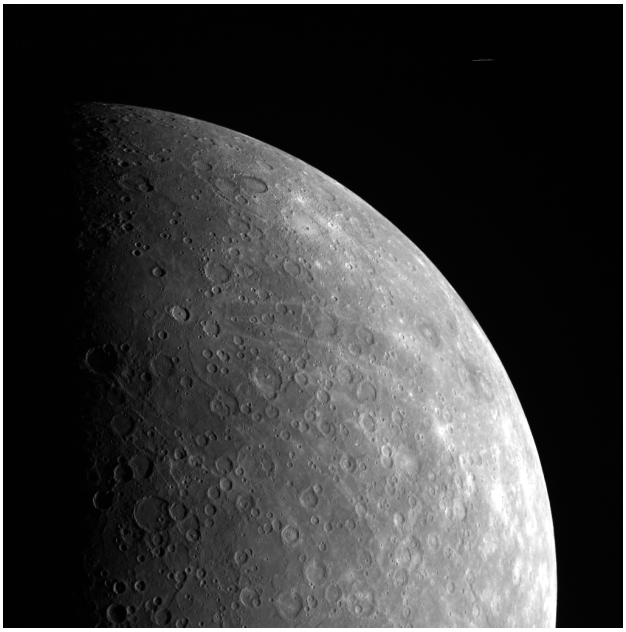
Step 1 – Retrieving and Plotting Image Files

The CRISMIS GSOC project requires retrieving image files from NASA public data repositories, reading/displaying the images, and performing machine-learning techniques to recognize cosmic-ray artifacts in the image. As an initial task, we'd like you to demonstrate that you can write a Python code, using Jupyter Notebook, that retrieves and displays image files acquired by NASA spacecraft.

The image files we will start with are from NASA's MESSENGER mission, which orbited the planet Mercury from March 2011 to April 2015. MESSENGER image data is stored in NASA's Planetary Data System (PDS), located at:
https://pds-imaging.jpl.nasa.gov/portal/messenger_mission.html

As a trial, retrieve and plot the data collected on 26 July 2011, which was day-of-year (DOY) 207 of year 2011. The data from this date can be found at:
https://pdsimage2.wr.usgs.gov/archive/mess-e_v_h-mdis-2-edr-rawdata-v1.0/MSGRMDS_1001/DATA/2011_207/
Note that your code should be written allow for data from any date to be easily retrieved by changing the DOY and year values to any desired value. During the project, this step will be automated.

The images take the form of "IMG" files. IMG files are binary data with a header, followed by the image data. In this data section, every pixel is listed, in order, from left to right and top to bottom, using however many bytes are necessary for the bit depth of an image. So a 512-by-256-pixel image with 16-bit pixels will contain (512 x 256 x 2 =) 262,144 bytes of image data, with each pixel represented by the 2 bytes necessary to record the 16-bit number corresponding to the digital number for that pixel.



One of the images in the 2011 207 directory is shown to the left. Its file name is EW0220137668G.IMG. Please display it in the Jupyter notebook you provide at the end of this step. Next, search the 2014 215 directory for file EN1049375684M.IMG and also display it on your notebook. Cosmic rays that hit the camera CCD appear as streaks of saturated pixels in the images. Can you spot the cosmic ray artifacts in both images?

Step 2 – Dataset preparation

Images with cosmic ray artifacts:

Year	Day of Year	File Name
2011	207	EW0220137668G.IMG
2014	215	EN1049375684M.IMG
2011	155	EW0215651174C

In fact, a large number of images from the 4 June to 6 June time frame are contaminated by cosmic ray artifacts, due to a strong solar event that was bombarding the MESSENGER spacecraft with energetic protons. For example, see:

2011	157	EW0215844301G
2011	156	EN0215728629M
2011	156	EW0215726557D

Some of the images in this time period are completely saturated with cosmic ray events! Learn to identify cosmic ray artifacts in the images, and prepare a large dataset with images that do and do not have cosmic ray artifacts to train a machine learning algorithm on.

Image EW0215726557D, which is saturated with multiple cosmic-ray events. Turning up the contrast will reveal Mercury's surface, which is otherwise below the contrast as shown above.

Step 3 – Machine Learning Algorithm

Apply a machine learning technique of your choice (we recommend to start with a computer-vision inspired algorithm) to the classification of cosmic ray artifacts. You can start with a basic convolution neural network (CNN) and then try something more advanced (feel free to impress us at this stage). Make sure that all your work is clearly shown in Jupyter Notebook. Start thinking about a pipeline that connects everything together (automatic data retrieval, dataset creation, classification).