* In the sdoml\_dataset notebook, continue debugging from where marked

—

* Get 9K images for 1 year as files in googledrive
  + We know how to get the images in a dask-array (sub\_image)
  + We will get the images as rows in a pandas df
    - Columns should be: “datetime”, “image\_pixels\_json”

“[[0.2, 0.7, 0.76, …], [.45, .22, .99, …], … [256th]]”

# convert 'sub\_image' dask-array into pd dataframe:images\_df

import pandas as pd

images\_df = pd.DataFrame(columns=["datetime", "image\_pixels\_np\_str"])

for image\_num in range(len(sub\_image)):

images\_df.loc[len(images\_df)] = {

"datetime": "", # this will be someting like df\_time["Time"][time\_index].values[image\_num]

"image\_pixels\_np\_str": "", # this will be something like np.array2string(np.asarray(sub\_image[image\_num, :, :]))

}

images\_df.to\_csv("images\_df.csv", index=False)

—

* First in google collab load/read the SDO data, make sure I can get the datetime information of the SDO images and then match the datetime from GOES to SDO and find out if we have 127 images to work with.’
  + (D) Create my own google collab and read an image from one day in 2015 and plot it
  + (C)12/13 - Get an image for every hour in each day which is ~9000 images and put it on a folder in gd archive
  + And enrich the dataset with all images taken during flare events.
    - some things to consider, cadence and resolution of SDO images
  + tag it using the clean GOES events data
  + do for all SDO images in 2015

— for later —

* + Consider using transfer learning.