APOD Image classification

Problem Statement Brief:

The Data science challenge deals with the astronomy picture of the day images where the task is to classify between the real and fake images using CNN.

The Data:

We are provided with a training dataset which contains images of 2 classes real (5000 images) and fake(2000 images). There is also a validation dataset which contains around ~1500 real images and ~700 fake images.

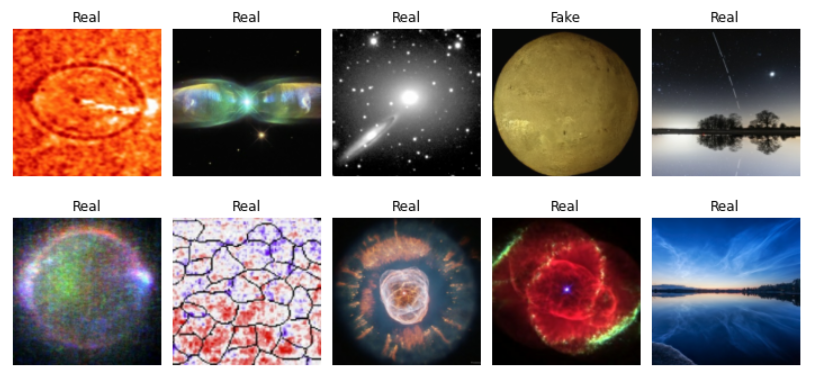
Data Loading:

The data is loaded from the directory using tf.keras generator functions, where the data is loaded in batches of specified batch\_size which in our case is 32. The image size is set at (112,112).

Samples of Training Data:

The RGB images are plotted after performing mix max scaling to bring the range to 0,1.



We can observe that the real image samples are a wide range of deep astro images both from space as well as terrestrial. The significant difference we can immediately notice wrt fake images is the vignette, contrast irregularities and improper shapes.Let us take a look at few random samples from the validation dataset 

Pre processing and augmentation: