MARS Images Classification

Joy Parikh, Moh Sahil, Wang Dorjee Sherpa and Rovina Pinto

December 13, 2020

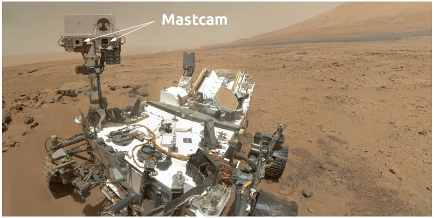
HiRISE and MLS dataset

HiRISE

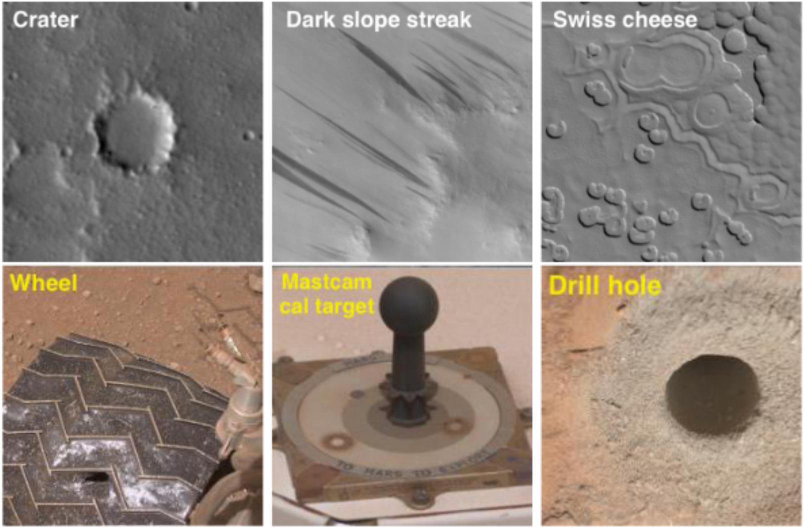
Dataset contains a total of 64,947 landmark images spanning 232 separate source images for Mars Reconnaissance Orbiter. Has 8 classes.

MSL

The data set consists of 6,820 images that were collected by the MSL Curiosity Rover. Has 19 classes.



Dataset Classes: Swiss Cheese, Wheels and More



Setting the Stage: Part A

Objective:

Classify images using a custom CNN architecture

Our Approach:

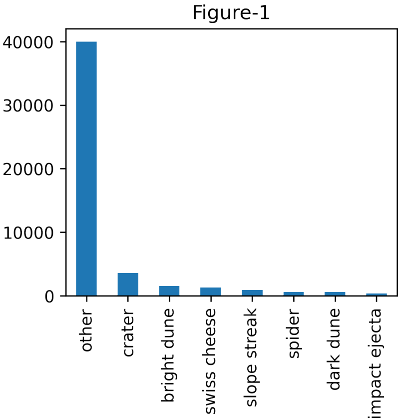
- Inspect data
- Make a simple CNN and have a look at results
- Improve results by suiting our model to the data, as well as employing some data processing

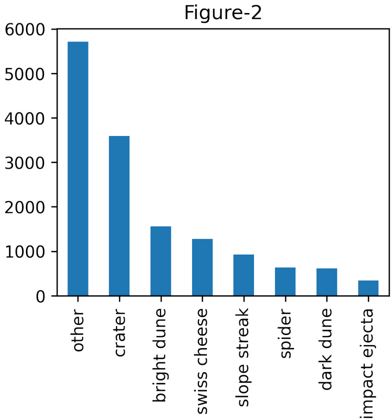
Algorithms employed:

- CNNs
- Downsampling
- Class Weights

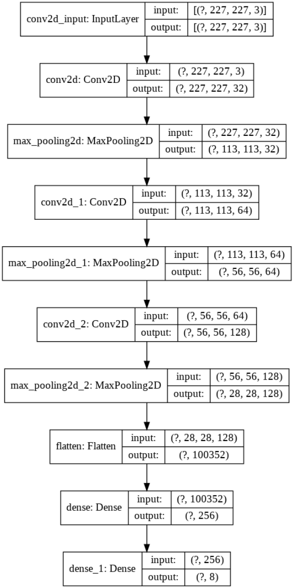
Trouble: Class Imbalance

Heavy Imbalance in the data is clearly visible in the ’other’ class. Fig. 1 is with data augmentation and Fig. 2 is after removing augmentation from ’other’ class





A Simple CNN



	precision	recall	f1-score	support
bright dune	0.19	0.81	0.30	16
crater	0.55	0.47	0.51	89
dark dune	0.48	0.79	0.59	66
impact ejecta	0.22	0.29	0.25	7
other	0.92	0.88	0.90	1482
slope streak	0.25	0.35	0.29	49
spider	0.00	0.00	0.00	42
swiss cheese	0.52	0.38	0.44	42
accuracy			0.80	1793
macro avg	0.39	0.50	0.41	1793
weighted avg	0.83	0.80	0.81	1793

A good test accuracy doesn’t have to mean anything. Need better suitable performance metric to access the results!

Class weights = cost sensitive learning

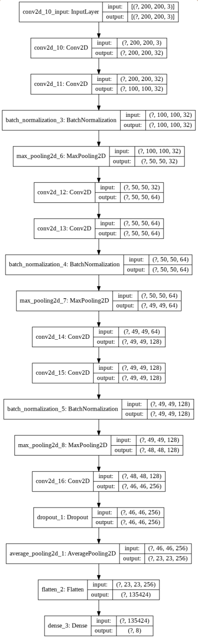
A not-so-simple CNN

*********** CLASSIFICATION REPORT (TEST SET)***********

	precision	recall	f1-score	support
bright dune	0.73	1.00	0.84	16
crater	0.51	0.67	0.58	89
dark dune	0.91	0.89	0.90	66
impact ejecta	0.16	0.57	0.25	7
other	0.95	0.91	0.93	1482
slope streak	0.42	0.71	0.53	49
spider	0.29	0.05	0.08	42
swiss cheese	0.95	1.00	0.98	42
accuracy			0.88	1793
macro avg	0.61	0.73	0.64	1793
weighted avg	0.89	0.88	0.88	1793

Patch Preprocessor for Train-Time Augmentation and Crop-Preprocessor for Test-Time Augmentation used

	Test Accuracy(%)
Normal Prediction	84.27
Using Crop Preprocessing	85.72
ResNet50+Logistic Regression	87.67



PartB: Transfer learning on HiRISE

Models used:

- MobileNetV2
- InceptionV3

Approach:

- Try to freeze as many layers as possible
- Try to reduce learnable parameters
- Maintain similar added layers to compare both model’s performance

Results:

- Mobilenet gave 90% test accuracy with 2.6mil parameters and 90 layers freezed
- Inception gave 91.75% test accuracy with 7mil parameters and 380 layers freezed

Part B: Transfer learning on MSL

Models used:

- VGG16
- InceptionV3

Approach:

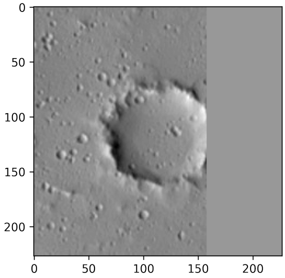
- Try to freeze as many layers as possible
- Try to reduce learnable parameters
- Maintain similar added layers to compare both model's performance

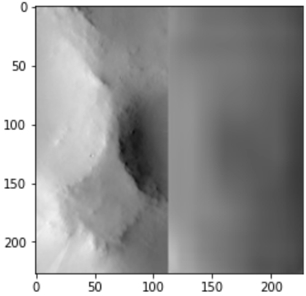
Results:

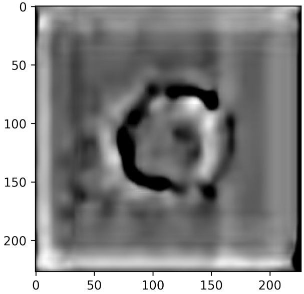
- VGG16 gave 74.33% test accuracy with 0.6mil trainable parameters and 19 layers freezed
- Inception gave 79.17% test accuracy with 0.8mil trainable parameters and 230 layers freezed

Part C: AutoEncoder

We tried two approaches for reconstruction, and the second one worked the best. It even predicted shadows in reconstructed part!







Learning

Handling domain shift problem is a non trivial task and cannot be addressed with class balancing methods. Therefore, sometimes it is important to accept the data as is while finding better ways to interpret the results and performance.

Working with real world data is a lot different than working with well structured data, it requires a lot of processing before training the model to get some good results.

Be very very VERY patient to get good results.

With ML and AI, it's important to remember:
"Garbage In, Garbage Out"