MARS ORBITAL IMAGE CLASSIFICATION (HiRISE)

Data

This dataset contains a total of 64,947 landmark images that were detected and extracted from HiRISE browse images, spanning 232 separate source images.

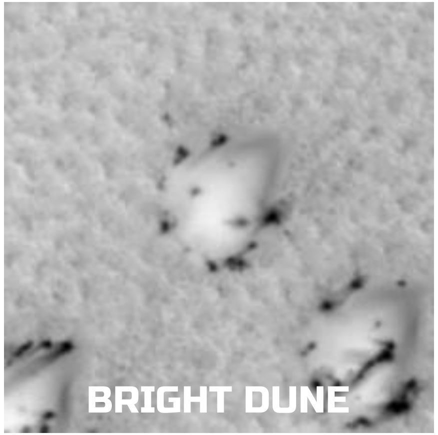
This set was formed from 10,815 original landmarks. Each original landmark was cropped to a square bounding box that included the full extent of the landmark plus a 30-pixel margin to the left, right, top, and bottom. Each landmark was then resized to 227x227 pixels. 9,022 of these images were then augmented to generate 6 additional landmarks using the following methods:

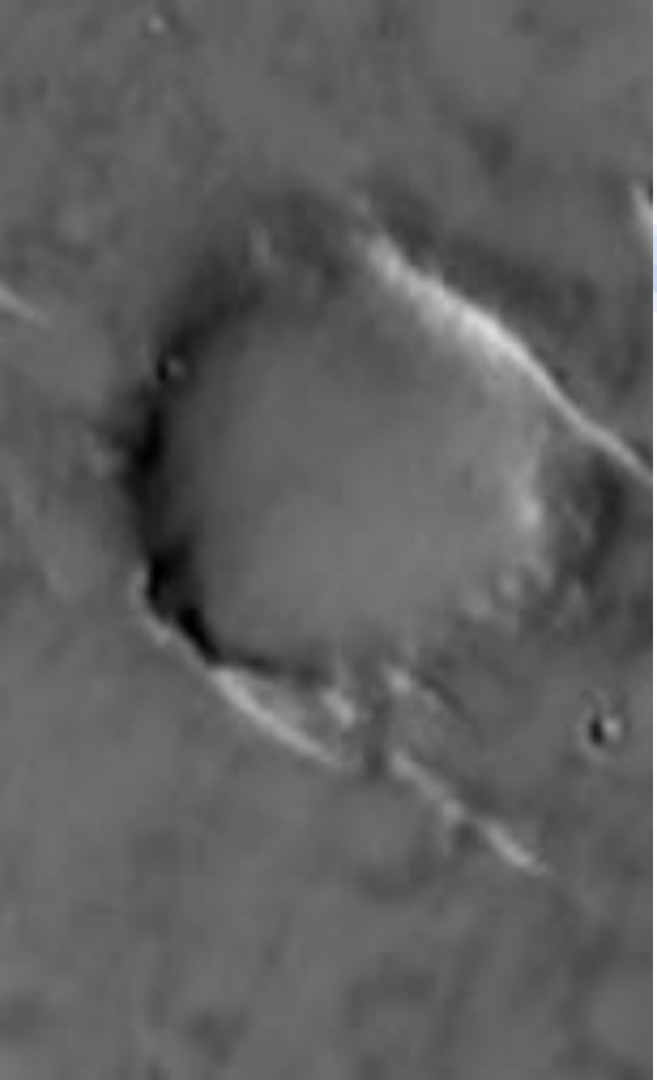
- 90 degrees clockwise rotation
- 180 degrees clockwise rotation
- 270 degrees clockwise rotation
- Horizontal flip
- Vertical flip
- Random brightness adjustment

The remaining 1,793 images were not augmented. Combining these with the 7*9,022 images, gives a total of 64,947 separate images.

Bright and Dark Dune

Bright dune and dark dune are two sand dune classes found on Mars. Dark dunes are completely defrosted, whereas bright dunes are not. Bright dunes are generally bright due to overlying frost and can exhibit black spots where parts of the dune are defrosting.



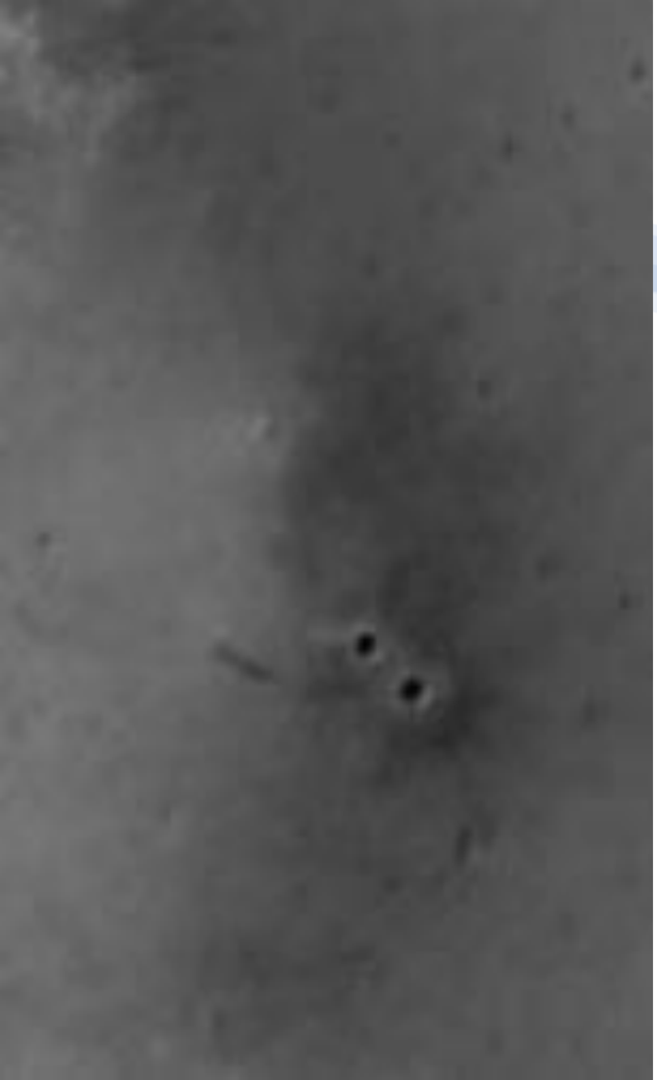


Crater

The crater class consists of crater images in which the diameter of the crater is greater than or equal to 1/5 the width of the image and the circular rim is visible for at least half the crater's circumference.

Slope Streak

The crater class consists of crater images in which the diameter of the crater is greater than or equal to 1/5 the width of the image and the circular rim is visible for at least half the crater's circumference.

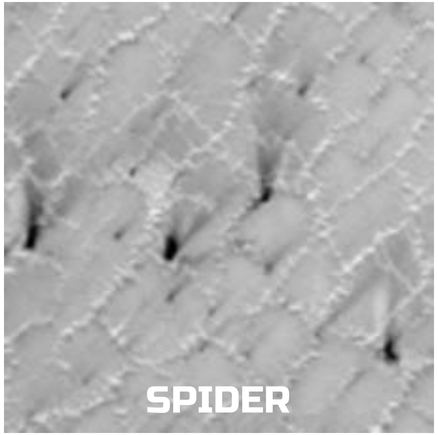


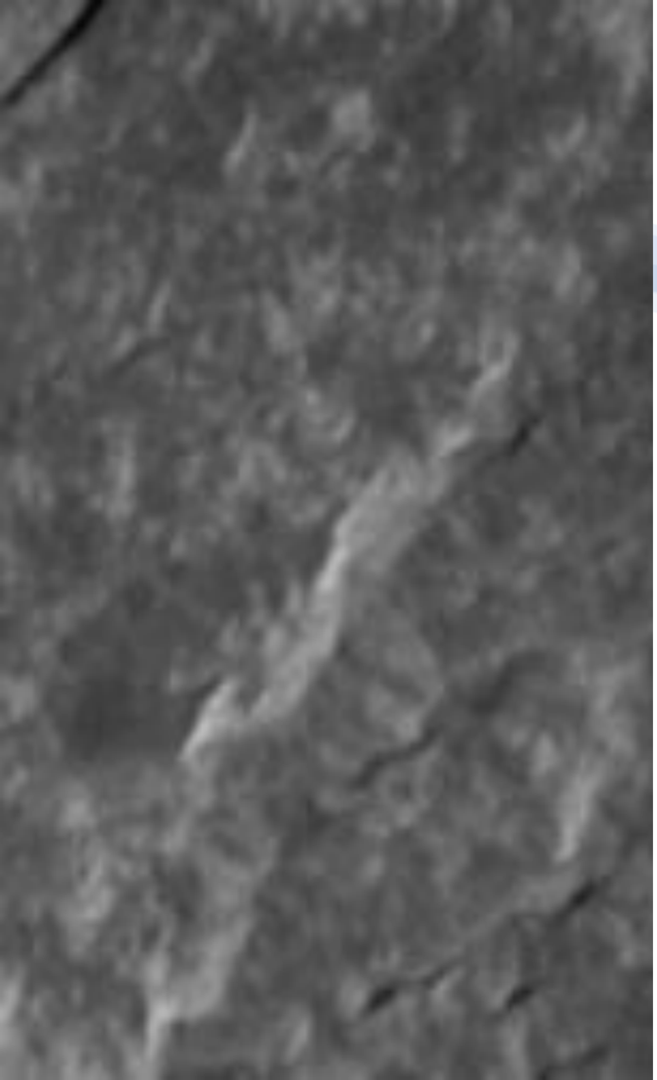
Impact Ejecta

Impact ejecta refers to material that is blasted out from the impact of a meteorite or the eruption of a volcano. We also include cases in which the impact cleared away overlying dust, exposing the underlying surface. In some cases, the associated crater may be too small to see. Impact ejecta can also include lava that spilled out from the impact (blobby ("lobate") instead of blast-like), more like an eruption (triggered by the impact). Impact ejecta can be isolated, or they can form in clusters when the impactor breaks up into multiple fragments.

Spider and Swiss Cheese

Spiders and Swiss cheese are phenomena that occur in the south polar region of Mars. Spiders have a central pit with radial troughs, and they are believed to form as a result of sublimation of carbon dioxide ice. This process can produce mineral deposits on top, which look like dark or light dust that highlights cracks in the CO2 ice. Spiders can resemble impact ejecta due to their radial troughs, but impact ejecta tends to have straight radial jets that fade as they get farther from the center. The spider class also includes fan-like features that form when a geyser erupts through the CO2 layer and the material is blown by the wind away from the cracks. Fans are typically unidirectional (following the wind direction), whereas impact ejecta often extends in multiple directions. Swiss cheese is a terrain type that consists of pits that are formed when the sun heats the ice making it sublimate (change solid to gas).

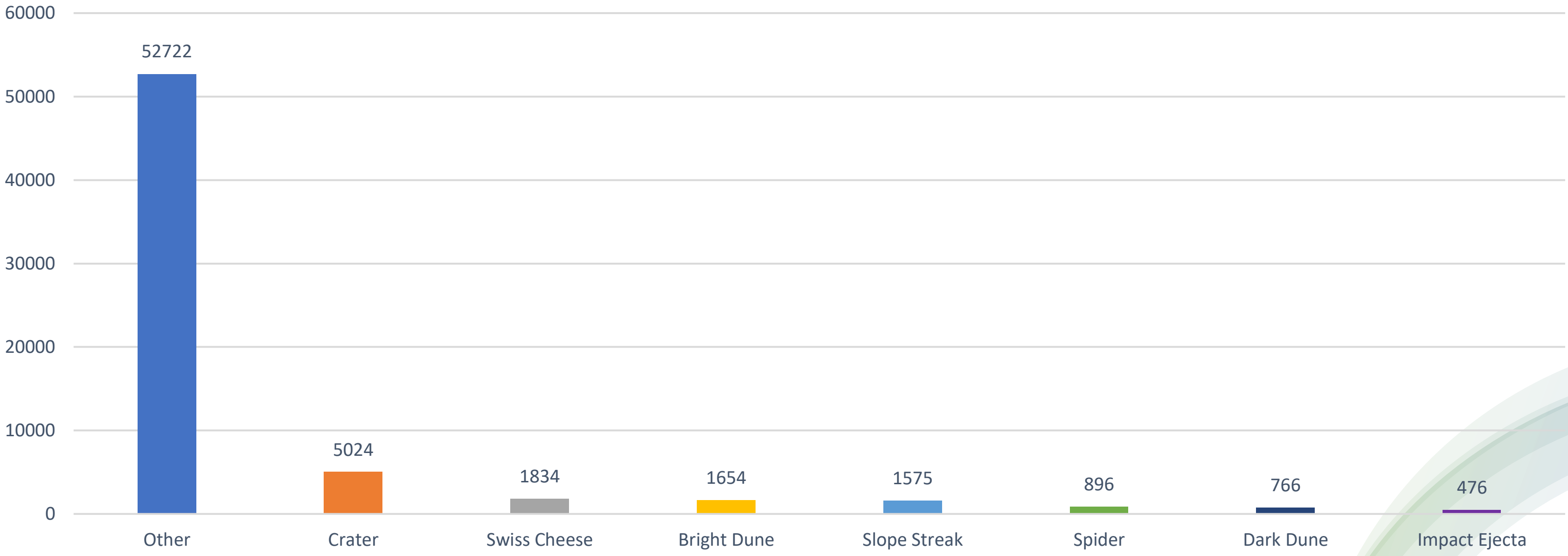




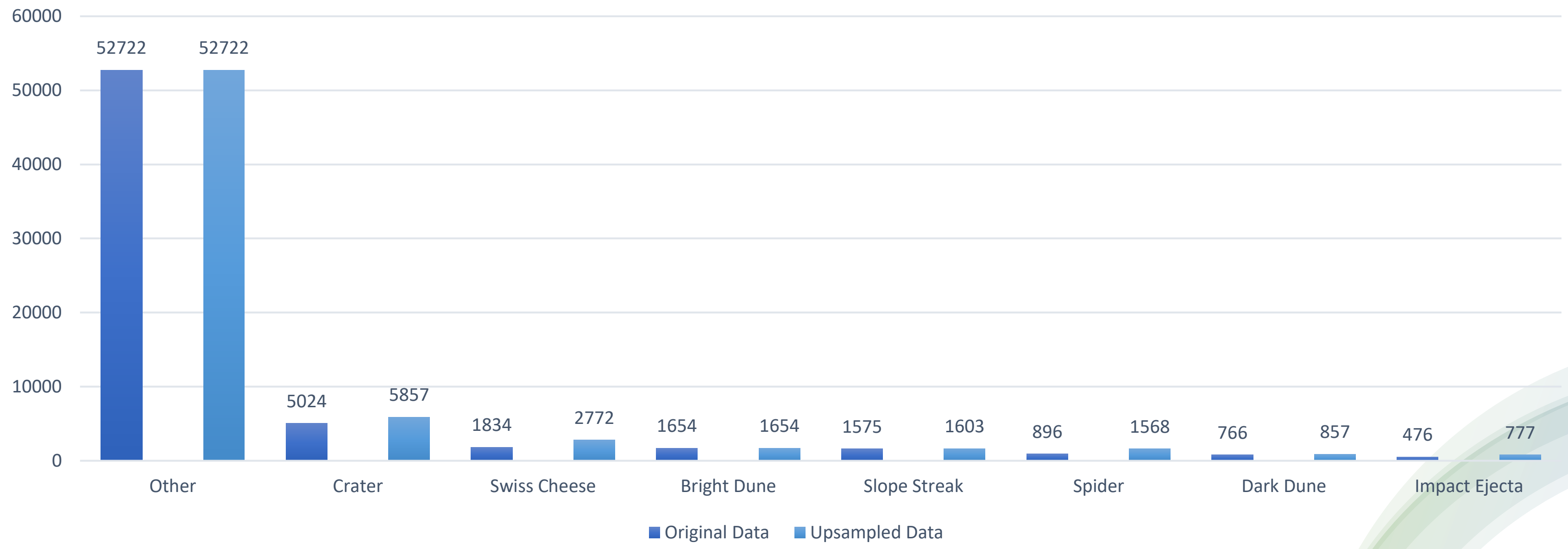
Other

Other is a catch-all class that contains images that fit none of the defined classes of interest. This class makes up most of our data set.

The classes in this dataset are highly imbalanced (shown in the training dataset distribution count plot graph). Most of the images classified as "Other" including data augmentation has roughly 53 thousand images, while other images from different classes have at most approximately 6 thousand.



The project is provided with the following two files: “**labels-map-proj-v3_2.txt**”, which includes two columns separated by a space: filename, class_id. The second file “**labels-map-proj_v3_2_train_val_test.txt**” includes train/validation/test labels and upsampling used for the trained model. The file includes three columns separated by a space: filename, class_id, and set. The following chart represents the difference in the number of images for each label in the dataset.



Model Training

Model Training

For this project we will be using a pre-trained model ResNet50V2 with "ImageNet" weights. The model includes 20,488 trainable parameters and 23,568,896 non-trainable (frozen) parameters.

We will replace the top layers with our own in order to classify landmark images instead of the original ImageNet classes.

Parameters for model compilation:

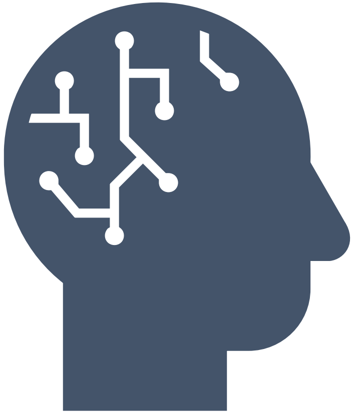
- **Loss** – CategoricalCrossEntropy
- **Optimizer** – Adam with learning rate of 0.0001
- **Metrics** – Accuracy (precision / recall will be used separately)

Parameters for model fitting:

- **Iterations** – 30 epochs
- **Callback** – EarlyStopping with validation loss monitor and patience of 4. 4 is the number of epochs with no improvement after which training will be stopped
- **Batch Size** - 32

Transfer Learning

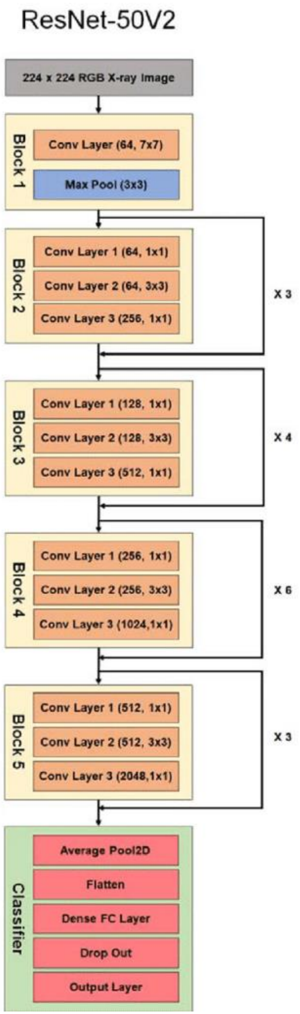
Transfer learning is a machine learning technique where a model trained on one task is used as the starting point for a model on a second, related task. This can save time and resources as the model can use the learned features from the first task as a starting point, rather than having to learn them from scratch. It can also improve the performance of the second model, as the features learned on the first task may be more general and transferable than if the model had to learn them from scratch.



ResNet50v2

ResNet50v2 is a deep convolutional neural network (CNN) designed for image classification tasks. It is a variant of ResNet (Residual Network) which is a type of deep learning model that uses skip connections to make the optimization of deep networks easier.

ResNet50v2 is known for its strong performance on image classification tasks, and it has been trained on a large dataset called ImageNet, which contains over 14 million images and more than 22,000 classes. ResNet50v2 is considered a powerful model with around 25 million parameters, which allows it to learn complex features from images. This model is widely used in computer vision, medical imaging, and other applications that require high accuracy in image classification.



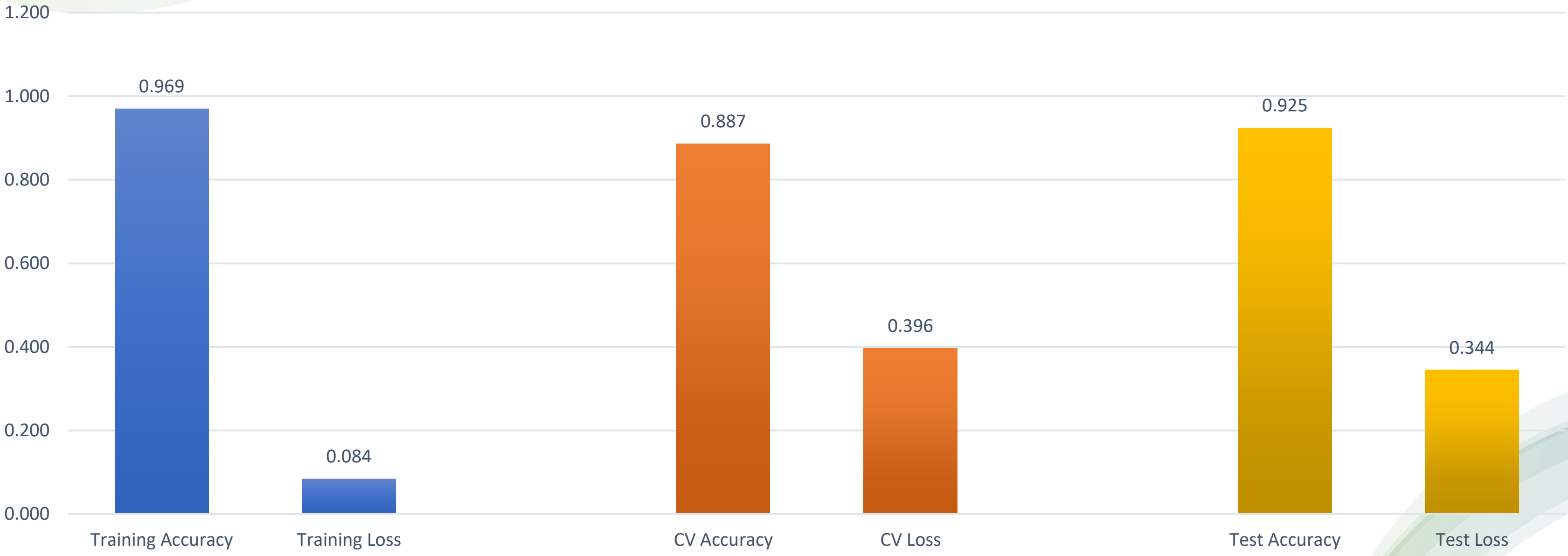
Rebuilding the Top Layers

The top layers include the following:

- **GlobalAveragePooling2D** - performs downsampling by computing the mean of the height and width dimensions of the input.
- **BatchNormalization** - a technique used to improve the performance and stability of neural networks. It is a normalization method that is applied to the activations of hidden layers in a neural network.
- **Dropout** - a regularization technique for reducing overfitting in neural networks (in this case the probability rate is set to 0.2).

Model Evaluation

The following chart represents the accuracy and loss of each segment – training, cross-validation and test. Although, in the next slide I will talk about a different set of metrics – precision / recall which are more accurate metrics when it comes to imbalanced dataset.



Precision / Recall and F1 Score

Accuracy is not a good metric to use when it comes to imbalanced dataset. Therefore, we will use precision and recall scores.

Precision is the fraction of relevant instances among the retrieved instances, while recall is the fraction of relevant instances that were retrieved. Both precision and recall are therefore based on relevance.

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive} \qquad Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

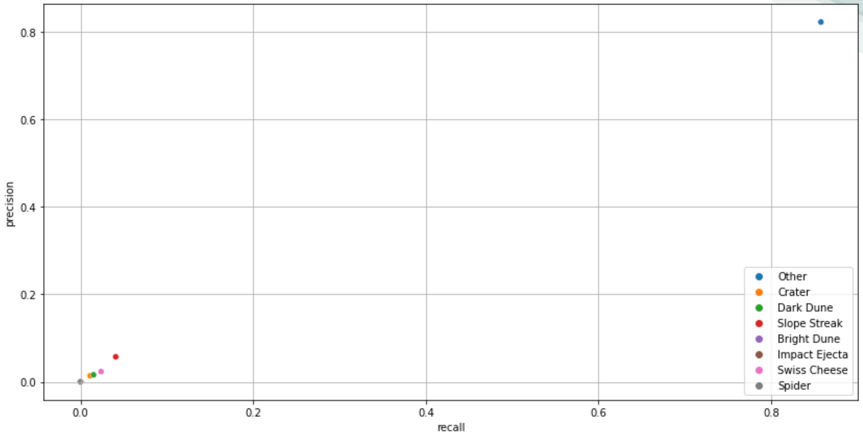
The f1-score takes the “harmonic mean” of precision and recall and combines the two into one metric.

$$F1\ Score = 2 * \frac{Precision\ * Recall}{Precision + Recall}$$

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

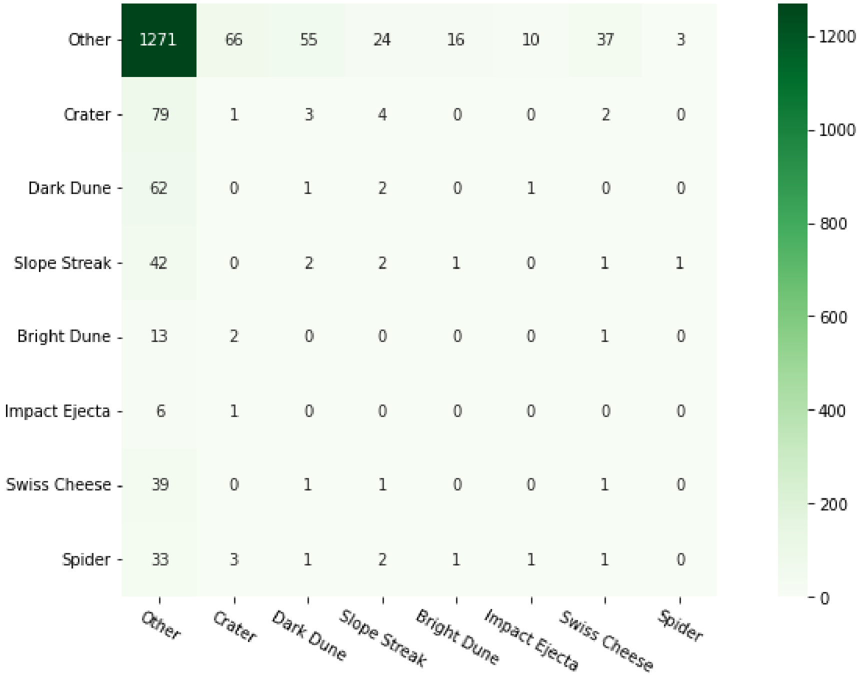
The following table and chart represent the precision/recall and f1-score metrics of each landmark:

Landmarks	Precision	Recall	F1-Score
Other	0.823	0.858	0.840
Crater	0.014	0.011	0.012
Dark Dune	0.016	0.015	0.016
Slope Streak	0.057	0.041	0.048
Bright Dune	0.000	0.000	0.000
Impact Ejecta	0.000	0.000	0.000
Swiss Cheese	0.023	0.024	0.024
Spider	0.000	0.000	0.000



Because the label Other has the most images it receives the highest scores out of the 8 landmarks with precision of 0.823, recall 0.858 and f1-score 0.84.

Confusion Matrix



Conclusion

The classes in this dataset are highly imbalanced. Most of the images classified as "Other" including data augmentation has roughly 53 thousand images, while

other images from different classes have at most approximately 6 thousand. We applied transfer learning technique using a pre-trained model ResNet50V2 with "ImageNet" weights. The model includes

20,488 trainable parameters and 23,568,896 non-trainable (frozen) parameters. The model gives a fairly good performance on the class "Other" with an f1-score of 0.84.

However, on other categories such as Crater, Dark Dune, Slope Streak and Swiss Cheese the model gives a very poor performance: f1-score of at most 0.05 on class Slope Streak.

Classes such as "Bright Dune", "Impact Ejecta" and "Spider" were classified incorrectly 100% of the time (shown in the confusion matrix).

The best way to improve performance on the less common classes is to increase their representation in the entire dataset.

THANK YOU

Do you have any questions?

aryeh.rotberg@gmail.com

www.linkedin.com/in/aryeh-rotberg

BIBLIOGRAPHY

The data used in this project comes from the DOI:

10.5281/zenodo.2538135

Gary Doran, Emily Dunkel, Steven Lu, & Kiri Wagstaff. (2020). Mars orbital image (HiRISE) labeled data set version 3.2 (3.2.0) [Data set]. Zenodo. https://doi.org/10.5281/zenodo.4002935

Idea for this project and the data originates from the following paper:

Kiri Wagstaff, Steven Lu, Emily Dunkel, Kevin Grimes, Brandon Zhao, Jesse Cai,

Shoshanna B. Cole, Gary Doran, Raymond Francis, Jake Lee, and Lukas Mandrake "Mars Image Content Classification: Three Years of NASA Deployment and Recent Advances" The Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI-21)