Mars Image Classification

Joy, Rovina, Sahil, Wang

Motivation and About the Project

This project is based on the work done by Lu et al in 2018 and 2019 which grew out of a **need to enable content analysis** to better find images of interest from the publicly accessible Planetary Data System (PDS) Imaging Node which hosts millions of images obtained from the planet Mars.

The class distribution of the datasets is severely imbalanced and will change with influx of data and changing classes or labels with better understanding of the surface of Mars.

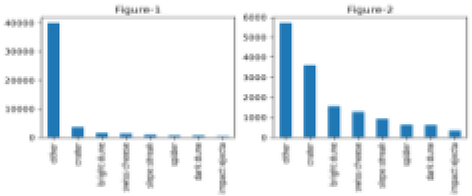
Therefore handling the latest dataset (09.2020), different from the authors (01.2019), presented us with the opportunity to utilize the paper as a template while looking out for unexpected issues.

The dataset consists of images from Mars that SOTAs are not trained with; therefore we think it will be interesting to see how the popular models such as VGG, Inception, etc perform on this unique dataset.

We extend the project to implement image completion using Autoencoders, where we look to understand the effects of architecture on robustness of image reconstruction.

Data and Labels

HiRISE
64,947 landmark images extracted from 232 source images
48,979 training, 14,175 validation and 1793 test images
Consists of 10,815 original landmarks
9,022 of them were then augmented. 8 classes as seen below
MSL:
6,820 images => 5,920 training (including augmented); 300 validation; 600 test images. Theses set were randomly sampled from sol (Martian day) range in temporal order. 19 classes of interest.



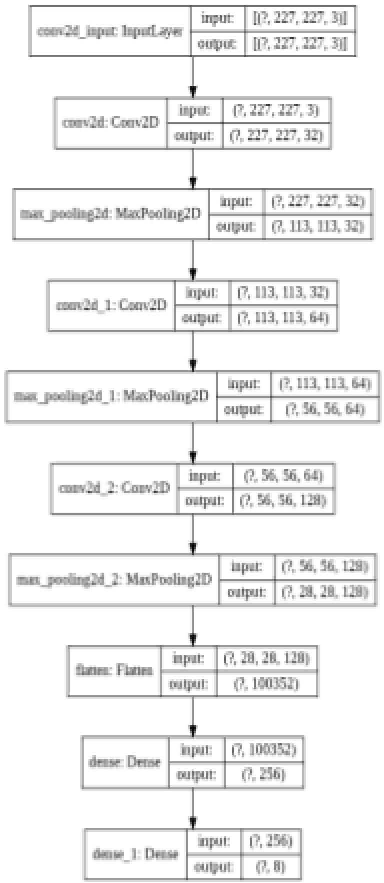
References

[1] Gary Doran, Emily Dunkel,Steven Lu,and Kiri Wagstaff. Mars orbital image (HiRISE) labeled dataset version 3.2. https://doi.org/10.5281/zenodo.4002935.

[2] Steven Lu and Kiri L. Wagstaff. MSL Curiosity Rover Images with Sci-ence and Engineering Classes. https://doi.org/10.5281/zenodo.4033453.

[3] Kiri L. Wagstaff, You Lu, Alice Stanboli, Kevin Grimes, Thamme Gowda,and Jordan Padams. Deep mars:Cnn classification of mars imagery forthe pds imaging atlas. InAAAI, pages 7867–7872, 2018.

Models



Model Part-A
Custom CNN models employing:
Class weights
Crop Preprocessing
RESNET50 + logistic Regression

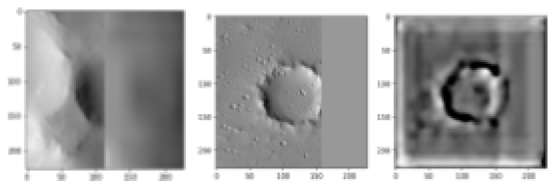
Model Part-B
HiRISE
MobileNet with additional dense and dropout layers
Inception with same added dense and dropout layers

Model Part-B
MSL
VGG16 with additional dense and dropout layers
InceptionV3 with same added dense and dropout layers

Model Part-C
Autoencoder
Convolutional autoencoder which takes half (227,114,1) image as input and predicts the other half (227,113,1)
Convolutional autoencoder trained on full (227,227,1) images and tested on fractional images.

Results

Part A : Test accuracy of 80% but the data imbalance affects the performance
Part A : 84% accuracy with deeper CNN and improved the accuracy to 85.72% using crop-preprocessor.
Part B : Mobilenet gives 90% and Inception gives 91.75% accuracy on HiRISE
Part B: VGG gives 74.33% and Inception gives 79.17% accuracy on MSL
Part C: Autoencoder approach 1 identifies general crater region but does not give good results visually and approach 2 gives good results as it reconstructs edges of crater.



Conclusion

	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			0.88	1793
macro avg	0.39	0.50	0.41	1793
weighted avg	0.83	0.88	0.81	1793

Test accuracies are not everything. Performance metrics are important with multi-classification problems.

Custom deeper layered CNN but with less learnable parameters along with custom preprocessor (PatchPreprocessor) as a train-time augmentation to help reduce overfitting and crop-preprocessor as a test-time augmentation to improve the accuracy were required to get better performance.

ResNet50 as a feature extractor and trained Logistic Regression Classifier on the extracted features resulted in an accuracy boost of 88%.

We initially wanted to apply transfer learning via the weights saved from our custom models onto MSL data given that the models were trained on Mars images and would possibly perform better with MSL data as compared to SOTA. But since we weren't able to get satisfactory results from the custom model on HiRISE data itself, we decided to drop the plan as it would be resources spent unwisely.

We constructed an autoencoder with the aim of reconstructing a fraction of an image. The performance was not as expected, but we believe that the model needs more training time and fine tuning to achieve the results we're looking for.