Deep Mars: CNN Classification of Mars Imagery for the PDS Imaging Atlas

Kiri L. Wagstaff,[1] **You Lu,**[1] **Alice Stanboli,**[1]
Kevin Grimes,[1] **Thamme Gowda,**[1,2] **Jordan Padams**[1]
[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, {firstname.lastname}@jpl.nasa.gov
[2]Information Sciences Institute, University of Southern California, Marina Del Rey, CA 90292, tg@isi.edu

Abstract

NASA has acquired more than 22 million images from the planet Mars. To help users find images of interest, we developed a content-based search capability for Mars rover surface images and Mars orbital images. We started with the AlexNet convolutional neural network, which was trained on Earth images, and used transfer learning to adapt the network for use with Mars images. We report on our deployment of these classifiers within the PDS Imaging Atlas, a publicly accessible web interface, to enable the first content-based image search for NASA's Mars images.

Introduction

The ongoing exploration of Mars by orbiting spacecraft and surface rovers has generated an enormous amount of data that continues to grow on a daily basis. A large portion of the acquired data consists of images, which are stored and made available to the public by NASA's Planetary Data System (PDS). The PDS currently contains more than 31 million images, of which 22 million are from the planet Mars.

Connecting scientists and other users to images of interest is a major challenge. The PDS Imaging Atlas allows users to search by mission, instrument, target, date, and other parameters that filter the set of images. Previously, the searchable parameters were all values that were known *a priori*, i.e., before the image was collected. However, users are often interested in finding images based on content (e.g., "show me all images containing craters"), which is known only after acquisition and must be extracted through content analysis.

Content-based searches can help further both science and mission operations. In images collected by the Mars Science Laboratory (MSL) rover, engineers are interested in analyzing all images that contain the rover's wheels so that they can monitor wheel degradation over time. While some images are purposely aimed at the wheels, and therefore can be obtained due to knowledge about the intended target of the image, in other cases the wheel is captured serendipitously in an image that was targeted at a nearby rock or soil. Using content-based analysis to find all wheel images can greatly increase the amount of relevant images for analysis.

Convolutional neural networks have achieved high performance on a variety of image classification tasks (Razavian et al. 2014). However, they tend to require thousands to millions of labeled examples if trained from scratch. A common practical solution is to use transfer learning to adapt a previously trained network to a new problem. This "fine-tuning" approach enables practitioners to train networks for new problems while leveraging the representation already inferred by the original network. However, the limits of such fine-tuning are not known, and we wondered whether a network that was trained on Earth images could successfully be adapted to operate on images from another planet.

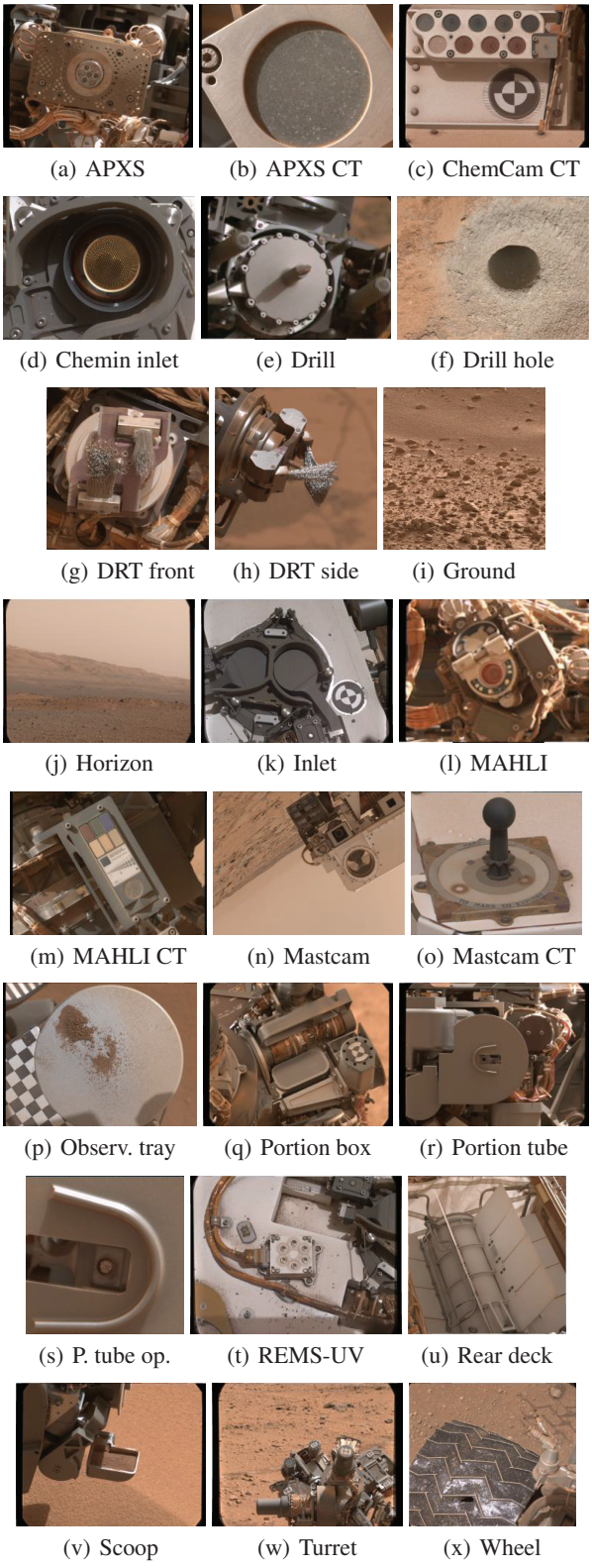
We tackled the problem of content-based image search for Mars images by adapting a typical Earth-image convolutional neural network to classify the content in images of Mars. One network, MSLNet, classifies images taken by the Mars Science Laboratory rover, and the other network, HiRISENet, classifies regions within large images collected from Mars orbit. Both networks achieved high performance in validation and testing, and their predictions have now been integrated into the PDS Imaging Atlas to enable the first public content-based image search capability for the Planetary Data System.

Related Work

Automated image analysis has been used for decades to automatically detect surface features such as craters (Urbach and Stepinski 2009) and dune fields (Bandeira et al. 2011) in orbital images. More recently, the strong performance of convolutional neural networks (CNNs) on image classification tasks (Razavian et al. 2014), as well as their ability to infer useful features rather than hand-coding them, has led to their adoption in planetary science as well. For example, CNNs recently were shown to out-perform a support vector machine classifier when detecting two Mars surface features of interest (volcanic rootless cones and transverse aeolian ridges) in orbital images (Palafox et al. 2017). They have also been applied to Mars rover images of the surface to classify terrain and inform navigation (Rothrock et al. 2016).

Deep Learning for Mars Images

We employed transfer learning to adapt the AlexNet image classifier (Krizhevsky, Sutskever, and Hinton 2012) to classify images from Mars. AlexNet was trained on 1.2 million (Earth) images from 1000 classes in the ImageNet data



(a) APXS (b) APXS CT (c) ChemCam CT

(d) Chemin inlet (e) Drill (f) Drill hole

(g) DRT front (h) DRT side (i) Ground

(j) Horizon (k) Inlet (l) MAHLI

(m) MAHLI CT (n) Mastcam (o) Mastcam CT

(p) Observ. tray (q) Portion box (r) Portion tube

(s) P. tube op. (t) REMS-UV (u) Rear deck

(v) Scoop (w) Turret (x) Wheel

Figure 1: MSL (surface) image classes. "CT" indicates a calibration target. "P. tube op." stands for "Portion tube opening."

set. We adapted this network by removing the final fully connected layer and re-defining the output classes, then retraining the network with Caffe (Jia et al. 2014). We started with Caffe's BVLC reference model, which is a replication of AlexNet that was trained for 310,000 iterations and provided by Jeff Donahue[1]. Following Caffe's recommendations for fine-tuning[2], we specified a small base learning rate and stepsize (iterations between reductions in learning rate) and a learning rate multiplier of 1 (how the rate is adjusted) for all layers except the final layer, which was set higher. Precise values are given below for each classifier.

We created two Mars data sets that contain images taken from different perspectives: Mars rover images of the surface and rover parts, and Mars orbital images of interesting surface features. Each data set enabled the training of a custom fine-tuned CNN that can provide classifications of new images from the same instruments as they are collected.

MSL Surface Data Set

We created a data set of 6691 images of the Mars surface environment that were collected by three instruments on the MSL (Curiosity) rover: Mastcam Right eye, Mastcam Left eye, and MAHLI (Mars Hand Lens Imager). These cameras differ in their focal length (100 mm, 34 mm, and 18-21 mm respectively) and resolution (150, 450, and 14 μm/pixel)), and field of view (5 degrees, 15 degrees, and 34-39 degrees). The data set is composed of fully calibrated RGB thumbnail (256x256 pixel) versions of each image. This is the image size used by Krizhevsky, Sutskever, and Hinton to train AlexNet and also sufficiently large to enable identification of the classes of interest. The labeled data set is available at http://doi.org/10.5281/zenodo.1049137.

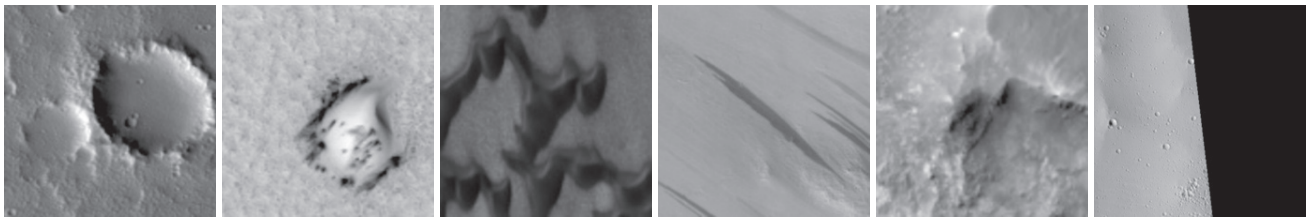
Twenty-four classes were identified by a Mars rover mission scientist. They include several instruments on the rover (e.g., APXS, drill, DRT, MAHLI) and associated calibration targets (marked "CT") as well as other rover parts (e.g., observation tray, inlet, scoop, wheel), plus two non-rover classes (ground and horizon). An example of each class is shown in Figure 1. The classes are in some cases very heterogeneous with objects imaged from different angles and magnifications, different backgrounds and illumination, multiple objects in the same image, objects that are out of focus, etc. Some images are RGB, while others are single-band (grayscale) or appear in other colors due to instrument filters.

HiRISE Orbital Data Set

Mars orbital images are collected in long strips as an instrument sweeps over the planet. For example, each image collected by the HiRISE (High Resolution Imaging Science Experiment) camera on the Mars Reconnaissance Orbiter covers several square kilometers at a resolution of 30 cm/pixel. Rather than assigning labels on a per-image basis, it is more useful to identify interesting features within

[1] https://github.com/BVLC/caffe/tree/master/models/bvlc_reference_caffenet

[2] http://caffe.berkeleyvision.org/gathered/examples/finetune_flickr_style.html



(a) Crater (b) Bright dune (c) Dark dune (d) Dark slope streak (e) Other (f) Edge

Figure 2: HiRISE (orbital) image classes.

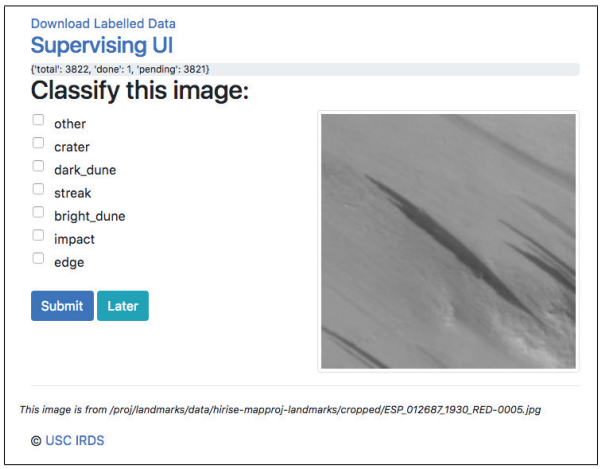


Figure 3: Supervising UI image labeling web tool.

each image and classify them. We employed dynamic landmarking (Wagstaff et al. 2012) to find visually salient "landmarks" within each HiRISE image and cropped out a square bounding box around each landmark, plus a 30-pixel border, to create a data set for image classification. Processing 168 map-projected HiRISE images yielded 3820 grayscale landmark images, each of which we resized to 227x227. The labeled data is available at http://doi.org/10.5281/zenodo.1048301.

The classes of interest that we identified in these Mars orbital images are craters, bright sand dunes, dark sand dunes, and dark slope streaks (see Figure 2). Landmarks that did not contain one of these features were labeled "other." Because the landmarks were detected within map-projected HiRISE images, they also included many spurious detections of the triangular black image border, which is very statistically salient. Therefore, we included an "edge" class in our labeled data to capture (and filter out) such images.

We developed a web-based image labeling tool[3] to facilitate the labeling of the thousands of images in this data set (see Figure 3). This Supervising UI server takes in a set of images and a list of class names and enables multiple users to simultaneously contribute labels to a given data set, thereby distributing the workload. Unlabeled images are randomly presented until all images in the data set acquire labels, which are stored in an SQLite database and can be downloaded as a .csv file. Users who are unsure about the classification of a given image can click "Later" to skip it so that others can label the image (i.e., users are not forced to make a possibly unreliable guess).

[3]https://github.com/USCDataScience/supervising-ui

Experimental Results

We trained two classifiers to operate on different kinds of Mars images: (1) MSLNet: Mars surface images collected by the MSL rover and (2) HiRISENet: Mars orbital images collected by the Mars Reconnaissance Orbiter.

MSLNet: Mars Rover Surface Images

We divided the MSL images into train, validation, and test data sets according to their sol (Martian day) of acquisition. This strategy is more appropriate than a random split since it models how the system will be used operationally with an image archive that grows over time. The images were collected from sols 3 to 1060 (August 2012 to July 2015); Table 1 shows the division into train, validation, and test sets by sol range. Because images are not collected with equal frequency by every instrument on every sol, the number of examples in each of the data set varies.

We fine-tuned the MSLNet classifier for 3000 iterations with a base learning rate of 0.0001, stepsize 500, and final layer learning rate multiplier 25. We compared the classification accuracy of the trained model to that of random selection as well as a simple baseline that always predicts the most common class observed in the training set, which was "ground" (see Table 2). MSLNet strongly outperformed both baselines. Interestingly, performance was lower on the test set than the validation set. The imbalanced class representation as well as evolving mission practices in terms of which items were imaged and with what instrument settings

Table 1: MSL image data sets, by instrument and sol range.

	Train 3–181	Val 182–564	Test 565–1060
Mastcam Left (ML)	1491	189	202
Mastcam Right (MR)	1935	94	373
MAHLI (MH)	320	1357	730
Total num. images	3746	1640	1305

Table 2: Classification accuracy on MSL (rover) surface images. The best performance on each data set is in bold.

Classifier	Train	Val	Test
Random	4.2%	4.2%	4.2%
Most common	62.5%	5.3%	19.5%
MSLNet	**98.7%**	72.8%	**66.7%**
MSLNet-inst	**98.7%**	**83.7%**	50.2%

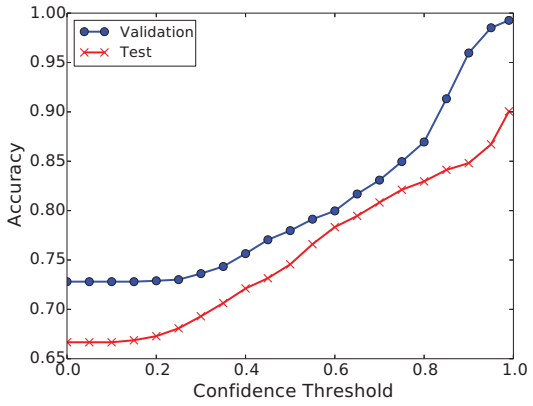


Figure 4: MSL image classification accuracy as a function of confidence threshold.

are likely contributing factors. This result is evidence that it will be important to periodically update MSLNet by providing a labeled sample of newly acquired images.

The CNN provides a posterior probability for each prediction that we employed to further increase the classifier's operational accuracy. We specified a confidence threshold such that predictions with lower probabilities were omitted from the results. Figure 4 shows accuracy as a function of confidence threshold for the validation and test sets. Employing a confidence threshold of 0.9 elevates validation accuracy to 94.3% and test accuracy to 84.0%. The fraction of abstentions was 41% and 52% respectively. For this application, users would prefer to have fewer results of higher quality than more results with lower quality.

We also experimented with training instrument-specific classifiers to see if they would perform better than a single generic MSL classifier. MSLNet-ML was trained using only images captured by the Mastcam Left eye camera, and the same strategy was employed to train MSLNet-MR and MSLNet-MH on their respective image data subsets. We evaluated classification accuracy of an ensemble model in which new images were classified by the appropriate instrument's model. The results appear in Table 2 under "MSL-inst". We found that training accuracy was identical but that validation performance increased significantly. However, performance on the test set instead went down. Again, the imbalanced class distributions and evolving image characteristics likely influenced the results.

Table 3: Classification accuracy on HiRISE (Mars orbital) images. The best performance on each data set is in bold.

Classifier	Train ($n = 2734$)	Val ($n = 546$)	Test ($n = 540$)
Random	16.7%	16.7%	16.7%
Most common	51.8%	56.2%	54.4%
HiRISENet	**99.1%**	**88.1%**	**90.6%**

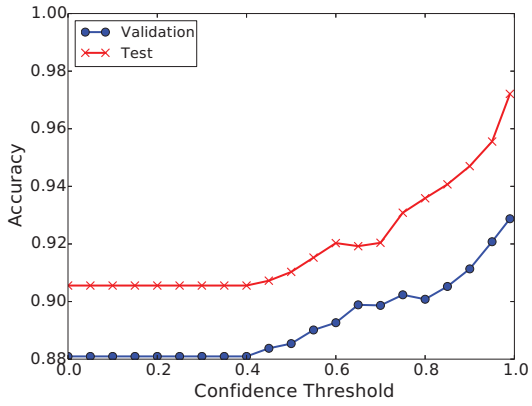


Figure 5: HiRISE image classification performance as a function of confidence threshold.

HiRISENet: Mars Orbital Images

We split the HiRISE data set into train, validation, and test sets by the HiRISE source image identifier, to ensure that landmark images from the same source image did not appear in more than one data set. We used 80% of the source images for training, 15% for validation, and 15% for testing.

We fine-tuned the HiRISENet classifier for 5300 iterations with a base learning rate of 0.0001, stepsize 20000, and final layer learning rate multiplier 10. The results are shown in Table 3, along with the performance expected for random predictions and the most-common baseline. In this domain, the "other" class was the most commonly observed class in the training set. The HiRISENet classifier strongly outperformed both baselines.

In comparison to the MSLNet results, performance on the HiRISE data set was generally higher. The data sets were more balanced and representative, and the total number of classes (5 versus 24) was lower, so the task may be inherently easier. Figure 5 shows classification accuracy as a function of confidence threshold (note change in y-axis scale compared to Figure 4). Applying a confidence threshold of 0.9 yields a validation set accuracy of 90.3% (abstaining on only 11% of the images) and a test set accuracy of 94.5% (abstaining on 13%).

The most common errors in the test set were 19 images from the "other" class that HiRISENet classified as "edge", and 14 "bright dune" images that HiRISENet classified as "other." These classes have significant conceptual overlap (e.g., how much of the black image edge must be present

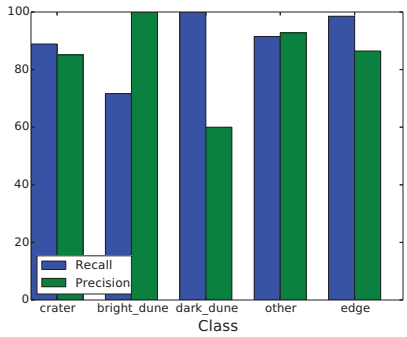


Figure 6: HiRISE test set precision and recall for each class.

to label an otherwise featureless terrain as "edge" instead of "other"?). Figure 6 shows the per-class precision and recall achieved on the test set. The "streak" class is not shown as HiRISENet did not classify any test images into this rare class (so precision is undefined). There are only two streak images present in the test set; both were missed. Recall was highest (100%) for the "dark dune" class, while precision was highest (100%) for the "bright dune" class. The apparently low precision for "dark dune" occurred because HiRISENet incorrectly classified two "other" images as "dark dune"; both predictions had low posterior probabilities (0.48 and 0.70).

Deployment for the PDS Imaging Atlas

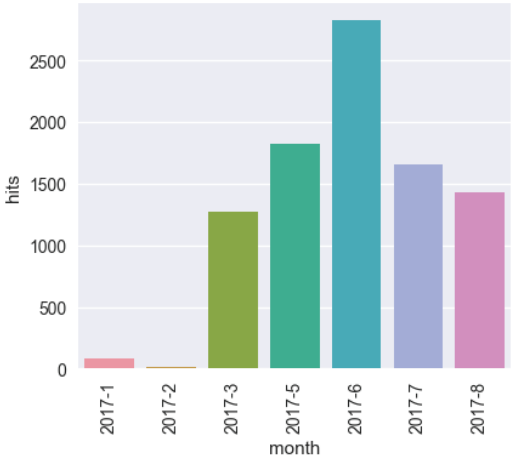
The goal of training Mars image surface and orbital classifiers is to benefit scientists and members of the general public who are interested in finding images that contain certain types of features. Therefore, we integrated the classifier's predictions into the PDS Imaging Atlas[4], a publicly accessible search interface to NASA planetary images.

The PDS Imaging Node Atlas III provides faceted navigation, an interactive style of browsing datasets that allows users to progressively filter a set of items to those of most interest. Faceted navigation has been utilized in the retail industry since the early 2000s (e.g., amazon.com and Google Shopping). A facet is a distinct feature or aspect of a set of objects (e.g., cost, size, style) or a way in which a resource can be classified (e.g., content classification).

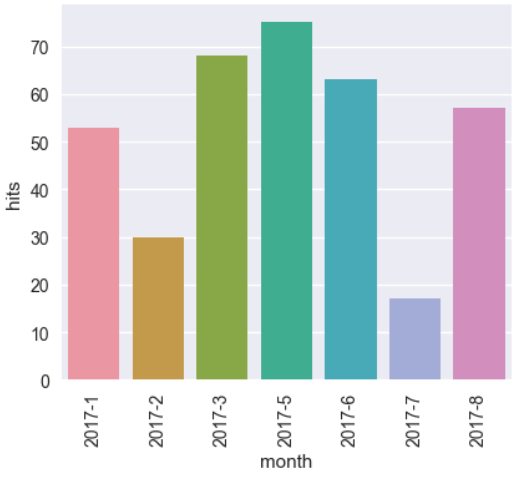
In the Atlas III, facets are defined by the most commonly used search criteria for imaging datasets, including mission name, instrument name, target, location meta-data (latitude, longitude), time constraints, etc. We have now added a new facet for Image Class (see Figure 7). When the user specifies a facet value, such as "Image Class" = "wheel", the results are filtered to contain only matching images, and the counts associated with all remaining facet values are updated accordingly. Older search systems required the user to guess which constraint they should apply next to narrow down the results; faceted search reduces the need for this kind of prior knowledge, which was a common user complaint.

[4]https://pds-imaging.jpl.nasa.gov/search/

Figure 7: PDS Imaging Atlas faceted search. Users can click image classes (left panel) to obtain images that contain content of interest, or enter class names via free text search.



(a) MSL image class queries



(b) HiRISE image class queries

Figure 8: PDS Imaging Atlas queries for image classes between January and August 2017. Note: metrics were unavailable for April 2017.

The Atlas also now provides free text search. A user can type "crater", and the resulting crater images are displayed. The search for crater is applied as both a facet-based query and a free text search of the ingested label meta-data.

We populated the Atlas database with classifications for the entire archive of MSL Mastcam, Navcam, and MAHLI images as well as all MRO HiRISE (calibrated) images. For MSL, we stored classifications for 3.7 million images (omitting thumbnail images) and provided the user with a confidence threshold slider to filter as desired. For HiRISE, we faced a different situation since the image classifications were of individual surface features within a larger image, not the entire image itself. In this setting, the number of detected landmarks of each class within each image is meaningful. We generated a catalog of all HiRISE landmarks with a posterior probability of at least 0.9 and provided the user with a slider to specify the minimum number of landmarks within each image. The high-confidence landmarks spanned a total of 16,344 HiRISE images.

We have compiled usage statistics by tracking Atlas queries that contain search terms related to the MSLNet or HiRISENet classes. As shown in Figure 8, we have seen increasing usage since the beginning of 2017. The raw number of queries is much higher for MSLNet classes than for HiRISENet classes, which might be due to the richer set of available classes, higher interest in rover part images in the user community, and/or our presentations to the MSL science team that directly advertised this capability.

Conclusions and Next Steps

In this work, we found that a convolutional neural network that was trained on Earth images can successfully be fine-tuned to adapt to classify Mars images. The transfer learning process was successful despite significant differences in image properties, imaging conditions, and classes of interest. We deployed networks that were trained on surface and orbital images of Mars to the PDS Imaging Atlas to enable the first content-based search of Mars images. Users have adopted this capability with enthusiasm as shown in the increasing number of content-based queries received by the Atlas.

An important lesson that came from our MSLNet experiments was that, for a mission that moves into new environments and changes its imaging practices over time, periodic labeling of new data and updating of the trained classifier will be needed. HiRISE is in a fixed orbit and surveys the entire planet, so the distribution of images over time does not change the way MSL's image distribution does.

We are now in the process of integrating the classifiers into the data ingestion pipeline for the Atlas. Each mission delivers a batch of new images to the PDS on a regular schedule (e.g., every three months). We have classified the entire archive of current images and will automate image classification as new images arrive.

We are currently developing an extension to the Atlas web interface that will provide overlays for each HiRISE image to show the location of each classified landmark. Rather than seeing that a given image contains five craters, users will see those craters outlined with their bounding boxes. We expect this visualization to increase the utility and interpretability of the classifier's predictions.

Acknowledgments

We thank Joy A. Crisp (JPL) for her assistance in identifying classes of interest in the MSL images and providing labeled examples of each class. We thank Norbert Schörghofer (Planetary Science Institute / University of Hawaii) for selecting relevant HiRISE images. We also thank the Planetary Data System (PDS) and the Multimission Ground System and Services (MGSS) program for funding and enthusiastically supporting this work. This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. Government sponsorship acknowledged.

References

Bandeira, L.; Marques, J. S.; Sarav, J.; and Pina, P. 2011. Automated detection of martian dune fields. *IEEE Geoscience and Remote Sensing Letters* 8(4):626–630.

Jia, Y.; Shelhamer, E.; Donahue, J.; Karayev, S.; Long, J.; Girshick, R.; Guadarrama, S.; and Darrell, T. 2014. Caffe: Convolutional architecture for fast feature embedding. *arXiv preprint arXiv:1408.5093*.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E. 2012. Imagenet classification with deep convolutional neural networks. In Pereira, F.; Burges, C. J. C.; Bottou, L.; and Weinberger, K. Q., eds., *Advances in Neural Information Processing Systems 25*. Curran Associates, Inc. 1097–1105.

Palafox, L. F.; Hamilton, C. W.; Scheidt, S. P.; and Alvarez, A. M. 2017. Automated detection of geological landforms on Mars using convolutional neural networks. *Computers & Geosciences* 101:48–56.

Razavian, A. S.; Azizpour, H.; Sullivan, J.; and Carlsson, S. 2014. CNN features off-the-shelf: An astounding baseline for recognition. In *Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 512–519.

Rothrock, B.; Kennedy, R.; Cunningham, C.; Papon, J.; Heverly, M.; and Ono, M. 2016. Spoc: Deep learning-based terrain classification for mars rover missions. In *Proceedings of the AIAA SPACE Forum*.

Urbach, E. R., and Stepinski, T. 2009. Automatic detection of sub-km craters in high resolution planetary images. *Planetary and Space Science* 57:880–887.

Wagstaff, K. L.; Panetta, J.; Ansar, A.; Greeley, R.; Hoffer, M. P.; Bunte, M.; and Schorghofer, N. 2012. Dynamic landmarking for surface feature identification and change detection. *ACM Transactions on Intelligent Systems and Technology* 3(3). Article number 49.