SPACE Conferences and Exposition
13 - 16 September 2016, Long Beach, California
AIAA SPACE 2016

SPOC: Deep Learning-based Terrain Classification for Mars Rover Missions

Brandon Rothrock*, Jeremie Papon*, Ryan Kennedy*, Masahiro Ono*, Matt Heverly†
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, 91109, USA

Chris Cunningham‡
Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, 15213, USA

This paper presents Soil Property and Object Classification (SPOC), a novel software capability that can visually identify terrain types (e.g., sand, bedrock) as well as terrain features (e.g., scarps, ridges) on a planetary surface. SPOC works on both orbital and ground-bases images. Built upon a deep convolutional neural network (CNN), SPOC employs a machine learning approach, where it learns from a small volume of examples provided by human experts, and applies the learned model to a significant volume of data very efficiently. SPOC is important since terrain type is essential information for evaluating the traversability for rovers, yet manual terrain classification is very labor intensive. This paper presents the technology behind SPOC, as well as two successful applications to Mars rover missions. The first is the landing site traversability analysis for the Mars 2020 Rover (M2020) mission. SPOC identifies 17 terrain classes on full-resolution (25 cm/pixel) HiRISE (High Resolution Imaging Science Experiment) images for all eight candidate landing sites, each of which spans over $\sim 100\text{km}^2$. The other application is slip prediction for the Mars Science Laboratory (MSL) mission. SPOC processed several thousand NAVCAM (Navigation camera) images taken by the Curiosity rover. Predicted terrain classes were then correlated with observed wheel slip and slope angles to build a slip prediction model. In addition, SPOC was integrated into the MSL downlink pipeline to automatically process all NAVCAM images. These tasks were impractical, if not impossible, to perform manually. SPOC opens the door for big data analysis in planetary exploration. It has a promising potential for a wider range of future applications, such as the automated discovery of scientifically important terrain features on existing Mars orbital imagery, as well as traversability analysis for future surface missions to small bodies and icy worlds.

I. Introduction

Identifying terrain type is critical for the safety of robotic operation on a planetary surface. Take, for example, the Mars rover Curiositys experience in the Hidden Valley. As shown in Figure 1, the narrow valley has a relatively steep slope on both sides and a floor that is constituted of rippled sand. Initially, the operation team commanded her to drive over the ripples. However, the deep sand was more hazardous than expected, causing high wheel slip and sinkage. As a result, the operations team backed up and chose an alternate path over a harder substrate to continue the traverse toward Mount Sharp.

In addition, Curiosity's wheels have also experienced an unexpectedly high rate of damage, particularly on Sol 450-515. The MSL Wheel Wear Tiger Team identified that the period of highest damage accrual occurred when the rover was driving over angular, embedded rocks. It also found that rocks on hard terrain, such as bed rock, are more likely to cause damage on the wheels. These examples highlight the importance of considering terrain type in planning safe and efficient rover traverses.

Terrain types on Mars can be identified on both orbital and ground images with well-trained eyes. For example, on orbital images taken by the most powerful telescopic imager in Mars orbit, the High Resolution

*Robotics Technologist, Mobility and Robotic Systems, 4800 Oak Grove Dr., Pasadena, CA.
†Robotics Mechanical Engineer, Project Systems Engineering and Formulation, 4800 Oak Grove Dr., Pasadena, CA.
‡PhD Candidate, Robotics Institute, Carnegie Mellon University, Pittsburgh PA.

Figure 1. The Hidden Valley, imaged by Curiosity's NAVCAM on Sol 717. Image Credit: NASA/JPL

Imaging Science Experiment (HiRISE) camera on the Mars Reconnaissance Orbiter (MRO), sandy terrain can be identified by finding dunes, which are typically separated from the adjacent dunes by 1-20 meters (shown in Figure 2). The wider separation between dunes indicate the better traversability. Also, we empirically know that polygonal ripples, which are often observed at the bottom of craters are significantly more challenging than linear dunes.

These important indicators of traversability can only be identified at the maximum zoom level on the 25cm-per-pixel-resolution HiRISE images, as shown in Figure 2. Manually performing such evaluation on all over the landing site, which typically spans over 10 km, is very laborious, if not impossible. Moreover, it requires the eyes of experienced geologists and rover drivers, who are not available for spending hundreds of hours on manual terrain classification.

To overcome the labor bottleneck of image labeling, we developed the Soil Property and Object Classifier (SPOC), an automated terrain classifier. Built upon a deep convolutional neural network (CNN), it learns from a small set of examples given by human experts, as shown in Figure 4. A training set consists of raw images (orbital or ground) and corresponding terrain labels. Once sufficiently trained, SPOC can automatically classify all pixels of input images. In this way, we can apply the judgment of experienced eyes on a significant volume of data with a marginal labor requirement.

We developed two versions of SPOC: SPOC-H for HiRISE orbital images, presented in Section III for Curiosity's Navigation Camera (NAVCAM) images, presented in Section IV. The two SPOCs have almost identical CNN architecture, but are trained by different data sets.

In this paper, we present the technical approach of SPOC, as well as two specific applications: landing site traversability analysis for the Mars 2020 Rover (M2020) mission, and slip prediction for the Mars Science Laboratory (MSL) mission. Section II presents the technical approach of SPOC, in particular the architecture of the convolutional neural network. Then, Section III presents the results of SPOC-H for the M2020 application, followed by Section IV-E presenting the result of SPOC-G for the MSL application.

A. SPOC-H for M2020 Landing Site Traversability Analysis

The landing site traversability analysis for the Mars 2020 Rover mission involves the assessment of the likelihood that the M2020 rover could drive to the regions of interest (ROIs) within distance and duration constraints specified by an engineering requirement for all the candidate landing sites. This assessment is performed by Monte-Carlo simulations, where the probability distribution of a landing point is given by a point cloud, and an optimal route planner is used to obtain the distance and time to drive to the ROIs

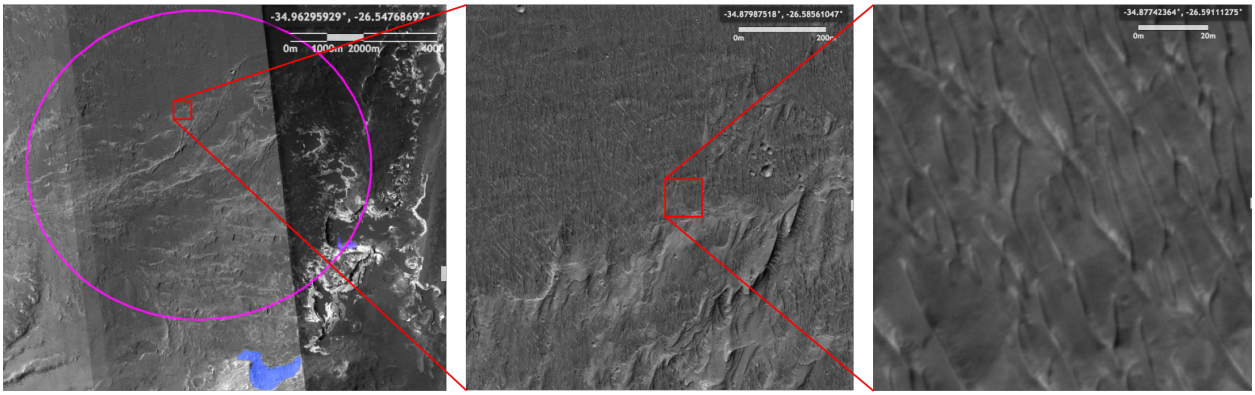

Figure 2. Holden Crater, one of the candidate landing site for the Mars 2020 Rover mission, imaged by the High Resolution Imaging Science Experiment (HiRISE) camera on the Mars Reconnaissance Orbiter (MRO). The magenta ellipse on the left figure is the landing ellipse, which spans $\sim 100\text{km}^2$. Sand ripples are visible on the right figure. As highlighted in this example, identifying terrain class requires inspecting orbital images from MRO's HiRISE camera at the maximum resolution (25 cm per pixel). Manually performing terrain classification for the entire landing site is impractical, if not impossible.

from all the landing points in the cloud.[Ono et al., 2016] Since the traversability and driving speed depends on terrain type, terrain classification is essential. However, for the reason explained above, manually classifying terrain for an single landing site is impractical, not to mention doing so for the eight candidate sites currently considered. SPOC-H made the impractical practical.

As explained in detail in Section III, SPOC-H takes HiRISE images as inputs, and classifies each pixels into one of 17 classes. The result is in turn used as the input to the cost function of the optimal route planner, which represents the inverse of the speed of the rover.

B. SPOC-G for Empirical Slip Prediction

Identifying terrain type is essential for predicting how much a rover will slip over its traverse. Intuitively, significantly greater slip is observed on sand than on cohesive soil or bed rocks. In hidden valley, for example, Curiosity experienced up to 81% slip at an average slope of merely 4 degrees.

However, obtaining a slip prediction model for Mars rovers is a challenging task. Roughly speaking, there are two approaches for slip prediction. One is a physics-based approach, where the mechanical forces and torques between the wheels and a soil are computed using terramechanics. However, this approach requires identifying soil parameters such as cohesion and friction angle.[Zhou et al., 2014] Unfortunately, accurately identifying these parameters is difficult without a soil sample and nearly impossible prior to contact.

The other is an empirical approach, where the relation between slip, slope angle, and terrain type is obtained through the regression on past driving data. However, this approach is not useful at the early phase of the surface mission due to the lack of sufficient driving data. Data from past rover missions, such as Mars Exploration Rovers (MER), is not useful due to the significant difference in the size and mass of the rovers as well as the variation in soil properties between landing sites.

As a result, slip prediction for the Curiosity rover has been primarily made by a model obtained from experiments on Earth.[Heverly et al., 2013] The model consists of three curves, corresponding to three distinct terrain types: sand, cohesive soil, and bed rock. For each terrain type, the curve gives the estimate of slip ratio as a function of the slope angle.

Four years of operation of Curiosity accumulated a significant body of data, which can be readily used for regression. The data includes ~8,000 instances of slip, detected by the difference between the wheel odometry and the visual odometry. However, still, there was one more issue: the terrain type can only be identified from images, and manually doing so for all the ~8,000 samples was not practical.

SPOC-G enabled automatic terrain class labeling from Curiosity's Navcam images, which subsequently enabled the generation an empirical slip prediction model from in-situ Curiosity mobility data. The results are reported in Section E.

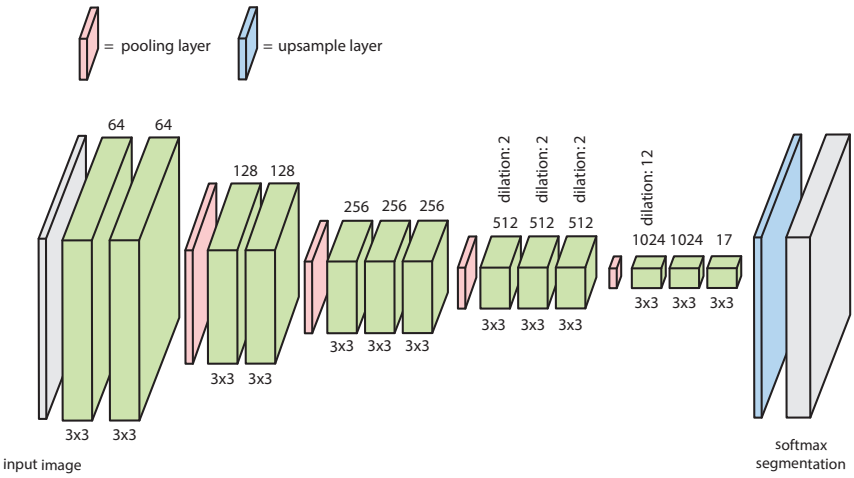

Figure 3. Fully-convolutional network architecture for terrain classification utilizes multiple stages of filtering and downsampling before upsampling the penultimate layer to the input resolution. The entire network is fully differentiable and trained end-to-end.

II. Technical Approach

The model used for both orbital and ground-based terrain classification utilizes essentially the same architecture, although trained on different data. Terrain classification can be viewed as a semantic segmentation problem, where a class label is assigned to every pixel in an input image. Recent developments in deep learning have produced dramatic improvements to semantic segmentation, and the current state-of-art is dominated by the use of fully-convolutional neural networks (FCNNs). Similar to the conventional neural networks used for image classification,[Krizhevsky et al.,] FCNNs classify a pixel by viewing the region around the pixel, often called the receptive field, and passing it through a deep network to compute a class prediction. The use of deep learning provides the capability to learn very complex and effective feature representations from data, without the use of hand-designed features used in essentially all previous approaches.

Individually classifying a window around each pixel is extremely inefficient, however, and the main innovation of FCNNs over conventional neural networks is the ability to infer the full semantic segmentation of the input image in a single forward pass. This is achieved by replacing the final fully-connected layers with a functionally equivalent bank of 1×1 filters. This produces an identical result, but amortizes the computational cost of the filter convolutions over overlapping receptive field windows. This also allows the network to be trained end-to-end from input image to output segmentation without any post-processing using a spatial cross-entropy loss. Unlabeled pixels in the training data are ignored in the loss function.

Our terrain classification model is based on the "DeepLab"[Chen et al., 2015] FCNN implementation. The network front end replicates the VGG architecture[Simonyan and Zisserman, 2014] modified to use "atrous" convolutions, also referred to as dilated convolutions.[Yu and Koltun, 2015] These specialized convolutions effectively increase the receptive field of the filters without increasing the filter size. An illustration of the network is shown in Figure 3. The network is trained with standard backpropagation and stochastic gradient descent, and takes on the order of 6 hours to train on a modern desktop GPU. Ground-based images can be processed entirely within the GPU in under $200ms$. For orbital images, memory limitation on the GPU require that the site is processed in tiles that are overlapped by the size of the receptive field. Classification of a typical HiRISE mosaic for a landing site takes approximately 12 hours. For training in both cases, tiles are extracted from the training data at random shifts and rotations at each training iteration.

III. Deployment for M2020

A. Class definitions

Table 1 shows the definitions of the 17 terrain types classified by SPOC-H in our M2020 application. The classes are sorted in the ascending order of the difficulty in traverse. These terrain classes then map to

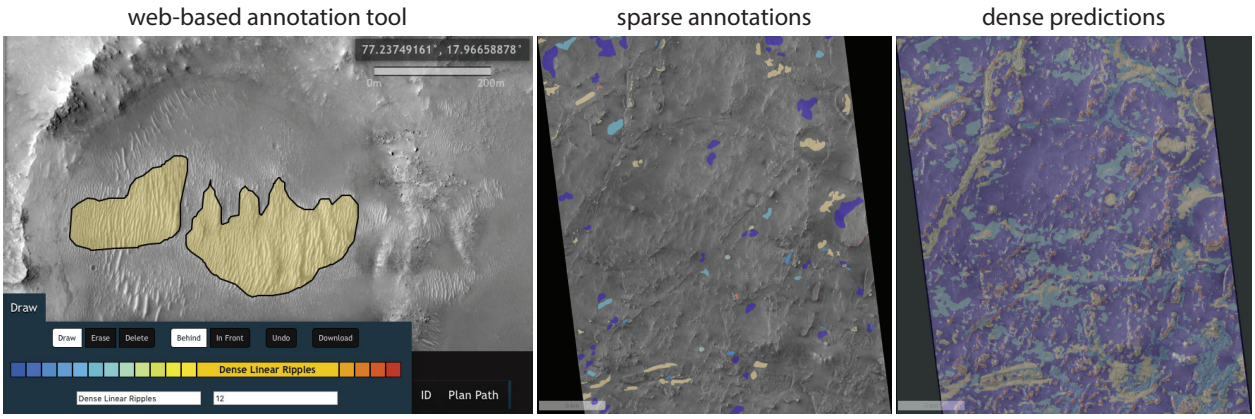

Figure 4. Web-based annotation tool for orbital imagery (left) is used to sparsely annotate terrain on large maps (center). After training a classifier, these maps are densely reclassified (right).

a traversability class, which is ultimately used to estimate the limitations and traverse rates of the rover. These traversability classes are shown in Table 2.

B. Data set

Due to the unique terrain morphologies that are often present at a given site, training data sets are curated individually for each site. Training labels are produced using a web-based tool developed for this task that allows annotators to navigate and zoom into the site and paint in regions with their appropriate terrain label. Typically only regions of high confidence are annotated, and due to the large size of the site the annotations often only cover about 1% of the total area.

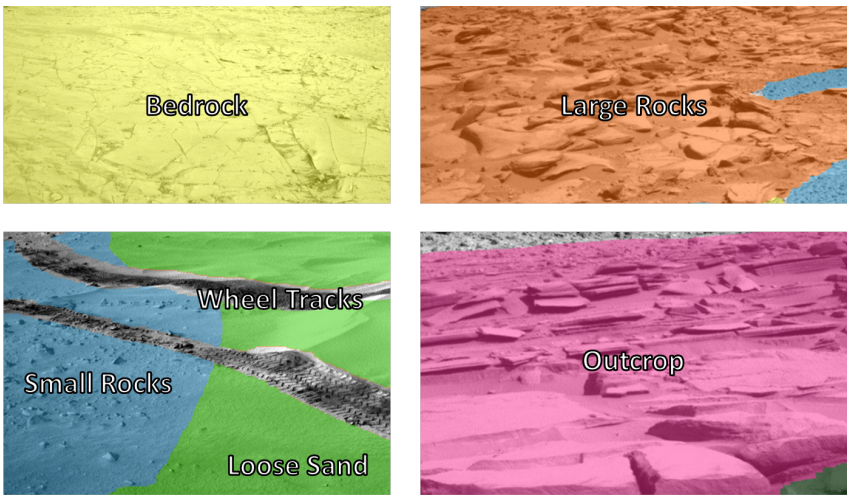

Figure 5. Examples of manually labeled terrain classes in Curiosity Navcam imagery.

C. Performance evaluation

For reporting the terrain classification performance of SPOC-H we use Columbia Hills, which is a candidate landing site for M2020 as well as the destination of the MER Spirit rover. The HiRISE mosaic for this site is approximately $40km \times 40km$, which corresponds to approximately 25 gigapixels. Performance is reported as a confusion matrix using an 80%/20% randomized split of the annotated labels for training and testing respectively. Average precision across all classes is 90.2%. Terrain classes not present at this site were omitted.

Table 1. Terrain class definitions for the M2020 landing site traversability analysis application.

Class	Definition
Smooth Regolith	Terrain that is not solid bedrock, but that is a firm surface without obstacles or significant terrain features. It is expected that the grousers will penetrate the regolith, but that the skin of the wheel will not see any substantial sinkage. The regolith has enough shear strength that the rover slip will be minimal.
Smooth Outcrop	Outcrop that is solid such that the wheel grousers will not penetrate in to the terrain. The outcrop is smooth with no obstacles that need to be avoided.
Fractured Outcrop	Outcrop that is solid such that the grousers will not penetrate in to the terrain. The outcrop has polygonal fractures that are often filled with sand or regolith, but the fractures do not constitute a mobility hazard and can be driven over without hesitation. There are no obstacles such as loose rocks on the outcrop that need to be avoided.
Sparse Ripples Firm Substrate	Ripples that are spaced 5 to 25 meters apart over the top of a firm substrate that is easily traversable. The distance between the base of one ripple and the base of the adjacent ripple is wide enough that the rover can traverse through the ripple field without needing to drive on the sand.
Moderate Ripples Firm Substrate	Ripples that are spaced 1 to 10 meters apart over the top of a firm substrate that is easily traversable. The distance between the base of one ripple and the base of the adjacent ripple is not always wide enough to enable passage of the full rover width without the rover driving on the ripple, but the rover can always have either the left or right wheels on the firm substrate.
Rough Regolith	Terrain that is not solid bedrock, but that is a firm surface where it is expected that the grousers will penetrate the regolith, but that the skin of the wheel will not see any substantial sinkage. The terrain has significant undulations or small rocks less than 50 cm in height on the surface such that a non-straight path may be necessary and such that terrain occlusions will prevent long, unobstructed blind driving.
Rough Outcrop	Outcrop that is solid such that the grousers will not penetrate the terrain and is rough such that the rover will have to take a circuitous path to drive a significant distance. The outcrop shows loose rocks on top of the outcrop or the outcrop is not smooth which generates small features greater than 25 cm in height that will need to be avoided.
Sparse Ripples Sandy Substrate	Ripples that are spaced 5 to 25 meters apart over the top of a sandy substrate. The sand may be lose or indurated, but its ease of traversability is not certain. The distance between the base of one ripple and the base of the adjacent ripple is wide enough that the rover can traverse through the ripple field without the need to climb over any ripples.
Moderate Ripples Sandy Substrate	Ripples that are spaced 1 to 10 meters apart over the top of a sandy substrate. The sand may be lose or indurated, but its ease of traversability is not certain. The distance between the base of one ripple and the base of the adjacent ripple is not always wide enough to enable passage of the full rover width without the rover driving on the ripple, but the rover is not required to be fully on the ripple to traverse the terrain.
Dense Ridges	Tight spacing (< 10 meters) of small (< 50 cm tall) sized outcrop scarps with collections of sand or loose regolith between the ridges
Rock Field	A concentration or rocks that will be difficult to negotiate the rover through. This is sometimes seen on the rim of a crater or at the base of a scarp.
Solitary Ripple	A single ripple with at least 25 meters to the next ripple. The length and height scale of the ripple are intentionally not mentioned as this can apply to small and large ripples. While the ripple may be traversable with caution, we will treat this terrain type as untraversable in order have the route planner identify a route around these easily avoidable features.
Dense Linear Ripples	A ripple field where the base of one ripple is coincident with the base of the adjacent ripple such that there is no flat region for the rover to drive between ripples.
Sand Dune	A large scale dune where the length of the slope is at least 3 meters such that the rover will be fully on the face of the dune and at a constant slope to traverse the dune.
Deep Sand	A sand accumulation up against a feature like a crater or scarp
Polygonal Ripples	A ripple field where the crests of the ripples are not all parallel but instead form a polygonal pattern with the ripple crests.
Scarp	A cliff or step in the terrain larger than 50 cm in height and longer than 3 meters (a rover) in length.

Table 2. Confusion matrix of classified terrain on Columbia Hills

Ground Truth	Prediction: SR	SO	SRF	RR	RO	SRS	DLR	PR	% of data
Smooth regolith	81.2%	1.0%	0.0%	13.3%	0.0%	0.2%	0.0%	0.0%	19.8%
Smooth outcrop	0.0%	93.5%	0.0%	0.0%	0.4%	0.0%	0.0%	0.0%	5.4%
Sparse ripples firm	0.6%	0.0%	99.6%	0.0%	0.0%	0.6%	0.0%	0.0%	2.5%
Rough regolith	12.1%	4.7%	0.0%	83.6%	6.2%	1.1%	0.0%	0.2%	11.8%
Rough outcrop	0.0%	0.0%	0.0%	0.6%	71.6%	0.3%	0.0%	0.0%	9.6%
Sparse ripples sandy	5.7%	0.9%	0.2%	1.5%	21.8%	97.6%	0.0%	0.0%	20.8%
Dense linear ripples	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	97.2%	1.9%	6.0%
Polygonal ripples	0.4%	0.0%	0.1%	1.0%	0.0%	0.3%	2.8%	97.8%	24.1%

IV. Deployment for MSL

A. Class definitions

The classifier was trained on approximately 700 manually labeled images. Example images are shown in Figure 5. The terrain classes are: (1) *Sand*, (2) *Small Rocks*, (3) *Large Rocks*, (4) *Bedrock*, (5) *Outcrop*, and (6) *Wheel Tracks*. *Sand* corresponds to loose sand without any visible rocks. *Small Rocks* is consolidated soil with small rocks. *Large Rocks* corresponds to regions with larger rocks on the surface. *Bedrock* is relatively flat, exposed bedrock. *Outcrop* refers to areas generally in the sides of hills with large rock formations that are generally not traversable. *Wheel Tracks* refers to visible wheel tracks from the rover. *Wheel Tracks* were not originally included, but were necessary to prevent misclassifications.

B. Preprocessing of MSL Navcam Images

Classification of terrain from rover-based images imposes unique challenges not generally observed in semantic classification tasks. First of all, rover-based images are often taken at different times of the day with significant variation in the level of lighting. We therefore normalize each image before it is processed by brightening/darkening it so that the median pixel value is 0.5.

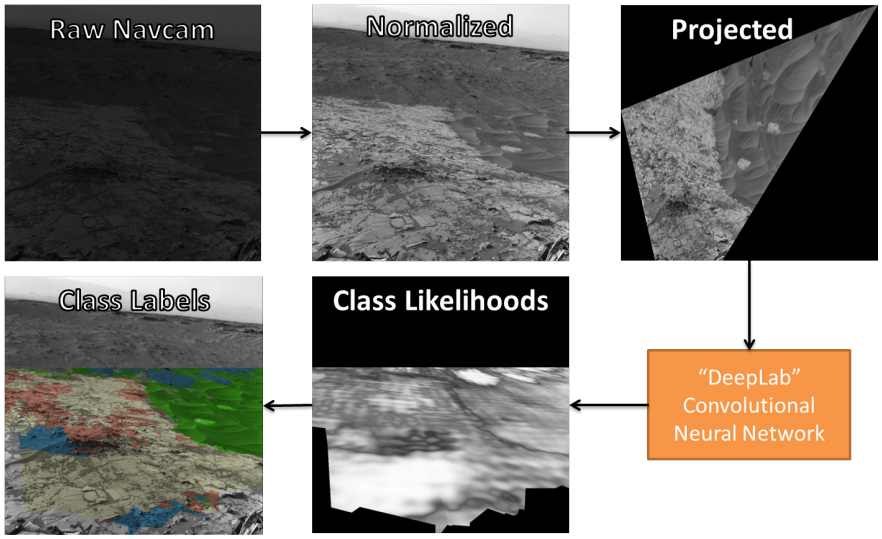

Figure 6. MSL Navcam preprocessing pipeline. The raw images are first radiance normalized, then zero centered. A ground plane is then fit to the local geometry, and the current image is orthoprojected to a top-down view. This scale-normalized image is then passed to a deep convolutional neural network for terrain classification.

A somewhat more challenging aspect of the problem is imposed by the goal of separating large rocks

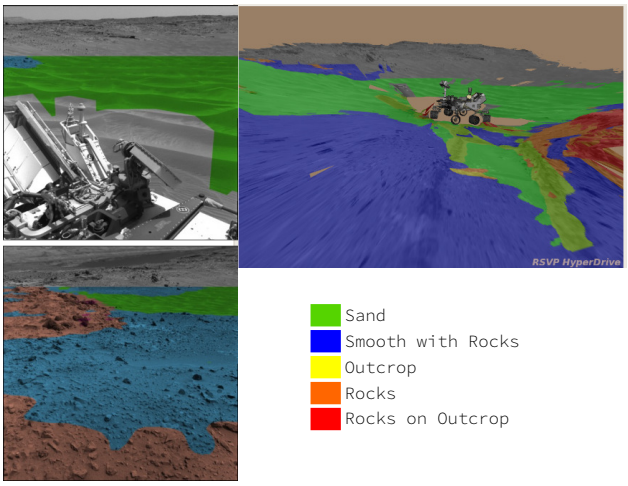

Figure 7. Example of classifier output on terrain images from MSL Navcam. On the left are some of the individual frames which are used for classification. On the right side the full classified panoramic mosaic is shown.

and small rocks into different classes. This means that, somewhat counter-intuitvely, we wish to avoid scale invariance within our classifier - as a small rock in the near-field will often be identical in appearance to a large rock in the far-field. While in principle a classifier could learn to understand projective geometry through training, we found that in practice this added too much complexity to the problem and prevented convergence. This was especially true given the small size (500 images) of our training set. To account for this, we developed a pre-processing scheme for scale-normalizing our input images using the local 3D geometry found using dense stereo reconstruction. First, we fit a plane to the local ground geometry using RANSAC, and then find the homography between this and the image plane. The image is then warped to a birds-eye view, cropped to a maximum distance of 30 meters from the rover, and all rover pixels are masked out. We chose to limit range to 30 meters as we found that beyond this point the sparsity of pixels made classifier results unreliable. An example of this pipeline is shown in Figure 6.

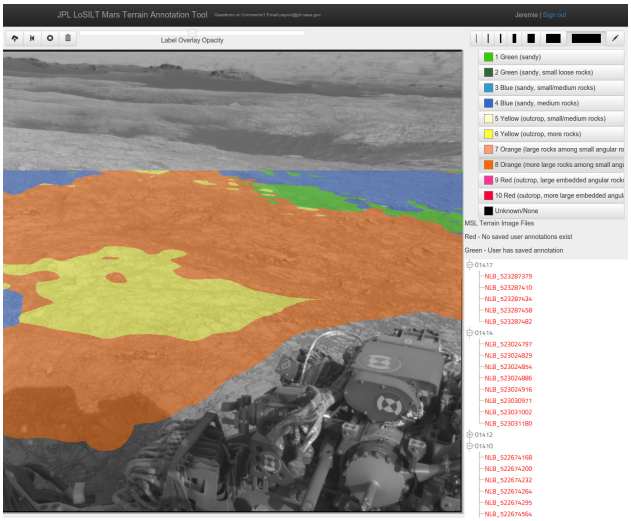

Figure 8. LoSILT Annotation Tool. This browser-based utility allows scientists and rover planners to review and correct the output of the classifier quickly from any location. This allows for iterative improvement of the classifier as more data of various terrain types is collected and added to the training set.

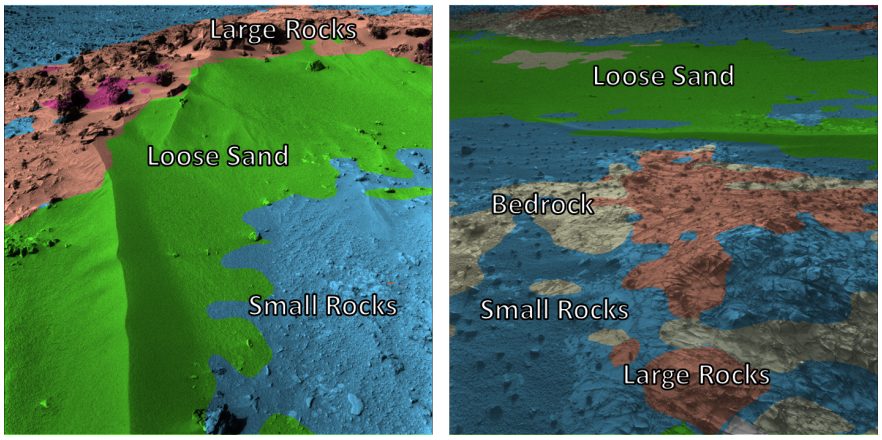

Figure 9. Examples of terrain classifier predicted classes in two Curiosity Navcam images.

C. Collection of Annotated Data

In order to collect annotations of terrain classes, a browser based annotation tool was developed which allows for scientists and operational rover planners to view MSL navcam imagery with classifier output overlayed. A screenshot of the tool can be seen in Figure 8. Using this tool, operators and scientists can review the classified output from previous Sols, and correct it when necessary. As more observations of terrain are collected, and the SPS team corrects it, our dataset will expand, allowing retraining and improvement of the classifier. Additionally, as the rover moves into new areas which contain novel terrain types, new classes can be added and annotated in the imagery. These can then be added to the training set and used to tune the classifier.

D. Results

Example classification results are shown in Figure 9. In both cases, the classifier does a reasonable job at predicting the correct label. However, some limitations become clear. In some cases, even a human might struggle to determine classes. For example on the right image, determining the boundary between *Sand* and *Small Rocks* is difficult. In the case where there are large rocks on top of bedrock, the correct class is also unclear. A confusion matrix on the holdout set for the classifier is shown in Table 3. These results show that in most cases the classifier performs well. Most misclassifications occur between classes that are similar and for which the boundary is unclear. Problems from unclear class boundaries are compounded by inconsistencies in human labeling. The other main source of misclassifications is due to *Small Rocks* making up more than half of the training data. Other classes are often misclassified as *Small Rocks* due to the frequency with which it appears in the training set. Such class imbalances are often problematic for classifiers, and we are investigating ways (e.g. data augmentation) to reduce the effect of these imbalanced priors.

Table 3. Confusion matrix of terrain classification on MSL Navcam imagery.

		Prediction						
		Sand	S. Rocks	Bedrock	L. Rocks	Outcrop	Tracks	% of data
Ground Truth	Sand	50.5%	44.4%	0.8%	1.7%	0.2%	2.4%	10.9%
	S. Rocks	2.0%	90.2%	0.6%	7.0%	0.2%	0.02%	54.6%
	Bedrock	9.9%	16.8%	44.5%	28.6%	0.3%	0.0%	7.7%
	L. Rocks	1.1%	57.5%	1.81%	38.3%	1.3%	0.02%	23.5%
	Outcrop	1.6%	34.6%	0.0%	34.0%	28.6%	1.1%	2.2%
	Tracks	0.7%	39.0%	0.0%	0.9%	0.8%	58.6%	1.1%

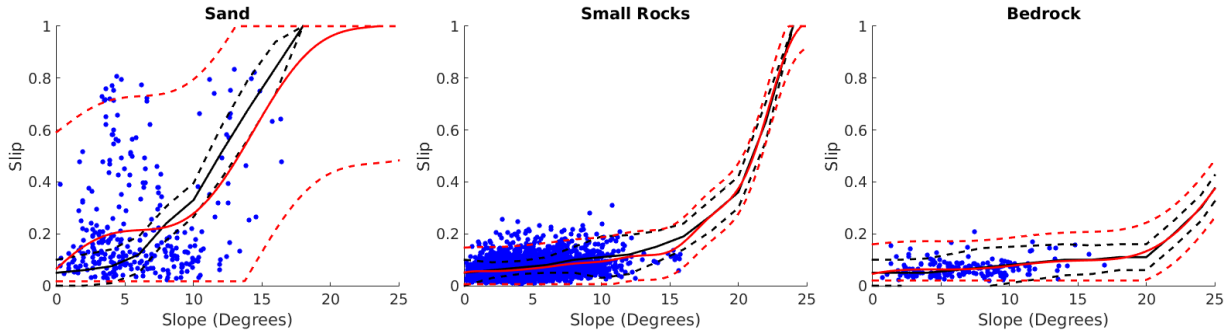

Figure 10. Slip vs. Slope curves for three terrain classes. Blue points indicate data extracted from Curiosity data. The black lines show the mean and error bars for the Earth-calibrated model. The red lines show the mean and 2.5 standard deviations around a Gaussian Process regression fit.

E. Rover Wheel Slip Prediction

Accurately estimating wheel slip is essential for estimating rover position through dead reckoning of wheel encoders. Slip prediction for MSL daily, tactical planning is currently done manually on Earth. Rover operators first analyze the type of terrain, then estimate slip by checking slope vs. slip curves for a given terrain class. These curves were calculated through extensive Earth-based testing in the JPL Mars Yard for loose sand, consolidated sand, and bedrock classes.[Heverly et al., 2013] Evaluating the accuracy of these curves on Mars is difficult due to the need to determine the terrain class at each position along a path for thousands of rover drives.

This section presents work that uses SPOC-G labeled Curiosity Navcam images in order to evaluate the currently used terrestrially calibrated slip models and the ability of regression based methods to improve upon these. The analysis uses data from Curiosity Sols 0-938, which contained 5501 slip measurements from visual odometry. First, the "best" image of the ground under the robot for each slip measurement is found using a heuristic-based search. Next, the class of that terrain is estimated using SPOC-G. Finally, the results for each terrain class are analyzed.

The first step to evaluating slip prediction methods was aggregating the slip and slope data for each terrain class. Terrain slope was estimated from the average rover tilt over a drive. Slip values were taken from on-board VO measurements. For each slip measurement, the rover's position on the ground was estimated from VO-corrected telemetry data. The region of interest on the ground for each measurement is the union of the area under each wheel.

The "best" image was found of each region of interest on the ground by searching through all Navcam images. An image was only considered if the image was taken within 10 m of the center of the region of interest. Images were then ranked using a manually-tuned heuristic. The heuristic was a weighted sum of the amount of the region of interest in the image, the amount of the image taken up by the region of interest, and the difference between the total distance traveled the rover had traveled at the time the image and slip measurements were taken. Heavy preference was given for images taken before driving over the terrain. In most cases, this heuristic-based method performed well. However, localization error was still a minor problem. These errors primarily occurred in loose sand regions with high slip and intermittent VO measurements. The terrain class was for a given slip measurement was estimated as the mode of all classes in the region of interest of the "best" image.

After obtaining terrain class labels for all ground truth slip measurements, both the Earth-calibrated models and Gaussian process regression-based models were evaluated using Curiosity data from Mars. Figure 10 shows slip vs. slope curves for three terrain classes. A Gaussian process regression model was fit to the data for comparison using the GPstuff toolbox.[Vanhatalo et al., 2013] Table 4 shows a comparison of the errors for both models over four of the terrain classes. *Outcrop* is not shown due to the lack of sufficient data.

In all four terrain classes, the Gaussian process performs at least as well as the Earth-calibrated model at predicting slip when compared based on root mean squared error. The results show that in *Small Rocks*, *Bedrock*, and *Large Rocks*, both the Gaussian process model and the Earth-calibrated model perform well. However, neither model is able to adequately predict slip in *Sand*. This is due to several factors that are difficult to identify in imagery including: terrain geometry, sand depth, and granular material properties.[Arvidson et al., 2016] Most of the samples with low slope but high slip occurred in very deep, loose,

	Root Mean Squared Slip Error		Percent Above Max Bound	
	Earth-Calibrated	Gaussian Process	Earth-Calibrated	Gaussian Process
Sand	0.23	0.19	37.46	4.83
Small Rocks	0.04	0.03	11.49	1.97
Bedrock	0.03	0.03	5.79	0.53
Large Rocks	0.05	0.05	20.94	1.11

Table 4. Root mean squared error (RMSE) between predictions and ground truth slip data and the percentage of data points that lie above the max bound for both the Earth-Calibrated model and the Gaussian process regression model. Max bounds for Gaussian processes were 2.5 standard deviations from the mean.

rippled sand.

This work illuminates one shortfall of the current terrestrially calibrated models, the inability to understand the uncertainty bounds of a slip estimate. For all terrain classes, the Earth calibrated model performs significantly worse than the Gaussian process in capturing all samples within its bounds. In fact, in sand more than a third of the samples lie outside of the bounds. Using data from Mars in the training set for slip prediction helps to make the uncertainty bounds more accurate for future slip predictions.

V. Conclusion

In this paper, we presented a methodology for autonomously classifying Martian terrain from both MRO HiRISE orbital imagery as well as MSL Navcam surface imagery. In both cases, we trained a state-of-the-art deep neural network implementation on hand-annotated labels, and then demonstrated their performance on a hold-out set. We then described ongoing work which uses the classifier outputs for landing site traversability analysis (in the case of orbital imagery), and wheel slip prediction (in the case of surface imagery). In the case of landing site traversability, we showed how classifier labels can be used to replace costly and time-consuming manual labeling by geologists. For surface imagery, we used the classifier results to improve our wheel slip models, and described how future rover planning will incorporate local terrain class to improve wheel slip prediction. While there remains much to be done in improving segmentation results, the work we have presented shows the potential of computer vision to reduce operator work-load and enable novel science mission capabilities in robotic space exploration.

Acknowledgments

We extend our thanks to Fred Calef and Tariq Soliman for annotation tool support, Richard Otero, Hallie Gengl, Anthony Campbell, and Hiroka Inoue for data preparation and labeling, Jeng Yen, Amanda Steffy, Oleg Pariser, Douglass Alexander, Adrian Tinio, and Nick Toole for supporting the development of SPOC-G. We also thank Prof. Ray Arvidson for providing guidance on terrain class definitions. This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. A portion of this work was supported by a NASA Space Technology Research Fellowship.

References

Arvidsonet al.,2016Arvidson, R. E., Haverly, M. C., Iagnemma, K. D., Bellutta, P., Maimone, M., Rubin, D., Fraeman, A. A., Stein, N. T., Zhou, F., Grotzinger, J. P., and Vasavada, A. R. (2016). Mars Science Laboratory Curiosity Rover Megaripple Crossings up to Sol 710 in Gale Crater. *Journal of Field Robotics*.

Chenet al.,2015Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2015). Semantic image segmentation with deep convolutional nets and fully connected crfs. In *International Conference on Learning Representations (ICLR)*.

Heverlyet al.,2013Heverly, M., Matthews, J., Lin, J., Fuller, D., Maimone, M., Biesiadecki, J., and Leichty, J. (2013). Traverse performance characterization for the mars science laboratory rover. *Journal of Field Robotics*, 30(6):835–846.

Krizhevskyet al.,Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, page 2012.

Onoet al.,2016Ono, M., Rothrock, B., Almeida, E., Ansar, A., Otero, R., Huertas, A., and Heverly, M. (2016). Data-Driven Surface Traversability Analysis for Mars 2020 Landing Site Selection. In *Proceedings of the IEEE Aeospace Conference*.

SimonyanandZisserman,2014Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Vanhataloet al.,2013Vanhatalo, J., Riihimaki, J., Hartikainen, J., Jylanki, P., Tolvanen, V., and Vehtari, A. (2013). GPstuff: Bayesian modeling with Gaussian processes. *Journal of Machine Learning Research*, 14:1175–1179.

YuandKoltun,2015Yu, F. and Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv preprint arXiv:1511.07122*.

Zhouet al.,2014Zhou, F., Arvidson, R. E., Bennett, K., Trease, B., Lindemann, R., Bellutta, P., Iagnemma, K., and Senatore, C. (2014). Simulations of Mars Rover Traverses. *Journal of Field Robotics*, 31(1):141–160.