Semantic Mapping in Unstructured Environments: Toward Autonomous Localization of Planetary Robotic Explorers

Kamak Ebadi, Kyle Coble, Dima Kogan, Deegan Atha, Russell Schwartz, Curtis Padgett, Joshua Vander Hook

NASA Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Dr, Pasadena CA 91109

Email: kamak.ebadi@jpl.nasa.gov

***Abstract*—Highly accurate localization of planetary robotic explorers is crucial for robust, efficient, and safe exploration and path planning in unknown and unstructured planetary environments. In these environments, where satellite-based radio-navigation systems are unavailable, global localization can be achieved by relying on registration of ground imagery to an orbital map, X-band Doppler radio transmissions, or direct observation of the robotic explorers in satellite imagery. While these methods have proven to be effective, they rely heavily on a human-in-the-loop. This paper studies the feasibility of autonomous visual global localization of planetary robotic explorers in extreme and GPS-denied environments by using a trained convolutional neural network (CNN) to obtain saliency maps from semantic segmentation of ground imagery. The saliency maps are registered to projected views of the terrain elevation maps in the rover's general region of operation to find the optimal match that places tight constraints on the pose of the robot in a Mars body-fixed coordinate system. We provide details on the use of the DeepLab V3+ framework for semantic image segmentation of Martian landscape imagery, including fine-tune training of existing models on domain specific data. Furthermore, we provide performance analysis of the proposed method on a Martian landscape dataset obtained by NASA's Perseverance rover, and discuss the limitations of the proposed method and future research directions.**

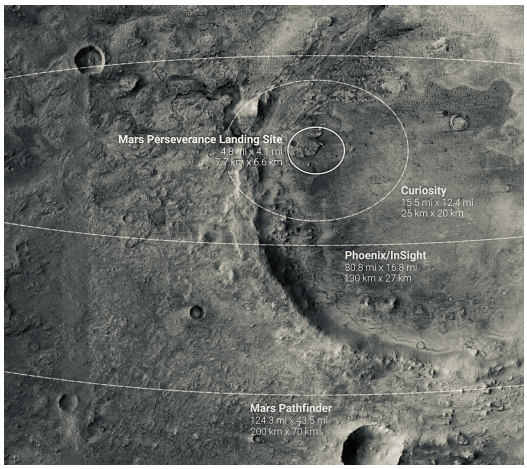

Figure 1: Visualization of landing ellipse size for Mars Pathfinder, Phoenix, InSight, and Curiosity, on the target landing area of NASA's Perseverance rover on Mars, Jezero Crater. Credit: NASA JPL-Caltech.

TABLE OF CONTENTS

1. INTRODUCTION

In planetary missions, topographic mapping and accurate landing site localization in a global reference system, and with respect to other salient features in the landing site, is crucial for determining the initial pose of the robot and enabling safe and efficient short- and long-range traverses to accomplish the science objectives of the mission. In early Mars exploration missions (e.g., NASA's Mars Exploration Rovers (MER)), the initial localization of the landing site was accomplished by the ground team within eight days of landing for both rovers [1]. This activity involved extensive data collection and tracking of the communications link in the inertial reference frame, reconstructing the entry, descent and landing (EDL) in returned descent images, and registration of features in the lander and orbital imagery. While localization was achieved with a sufficient level of accuracy, it required significant human intervention. NASA's most recent Mars exploration mission, the Perseverance rover, requires fast and highly accurate localization and navigation techniques, as the rover is tasked with gathering samples that could be returned to Earth by the Mars Sample Return campaign being planned by NASA and the European Space Agency. The rover is set to have long-range traverses at higher speeds than previous rover missions in order to gather a diverse set of martian rock and regolith samples throughout the life of the mission. A future Mars sample fetch rover would also rely on high-precision localization to retrieve the sealed samples from the Martian surface. The Perseverance rover relied on Terrain Relative Navigation (TRN) in order to land more precisely in a challenging but geologically interesting area. As shown in Figure 1, while past Mars missions disqualified scientifically compelling landing sites due to hazardous terrain within the

978-1-6654-3760-8/22/$31.00 ©2022 IEEE
The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

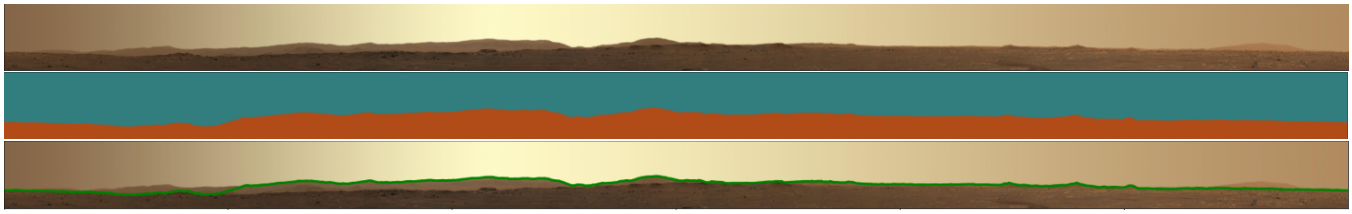

Figure 2: Semantic segmentation and skyline delineation demonstrated on a Martian panorama (top), with corresponding semantic segmentation prediction map (middle) and delineated skyline contour (green, bottom).

landing ellipse, the TRN method shrunk the area of the landing ellipse of uncertainty for the Perseverance rover by a factor of ten, as compared to the Mars Curiosity mission. This made it possible to target a landing ellipse with major and minor axes of 7.7 km and 6.6 km, providing access to previously inaccessible landing sites [2].

Using imagery obtained during the descent stage, the TRN method relies on registration of salient perceptual features to an orbital map to refine the rover's position accuracy to 40 m of position error until 2000 m above the surface, where the rover is separated from the back-shell. Although TRN has proven to be highly effective in shrinking the landing ellipse of uncertainty, illumination variations and sensor measurement noise could introduce differences between aerial imagery and the onboard orbital map, which could lead to increased position estimation uncertainty. Before the rover can safely and efficiently start navigating the environment, a high precision initial localization is required to establish the rover's position in a Mars global reference frame. In planetary applications where a Global Positioning System (GPS) is unavailable, localization can be achieved by matching perceptual features in terrain images obtained by the rover's onboard camera to an orbital map, analyzing X-band Doppler radio transmissions using satellites orbiting the planetary body, or by direct observation of the rover's landing site in satellite imagery. It is advantageous to enable a fully onboard localization capability, not reliant on external satellite communications that can be used for landing site localization, and periodic robot localization during its mission, to reduce the accumulated drift in estimated robot trajectory.

This paper presents a vision-based localization method for planetary robotic explorers that relies on a trained convolutional neural network (CNN) to obtain saliency maps from semantic segmentation of terrain imagery. We study the problem of efficient and robust extraction of landscape contours and skyline delineation through semantically segmenting the terrain and sky regions of distant landscapes, as shown in Figure 2, so that unique and distinctive contours of peaks, boulders, and general topography of the local environment can be used as points of reference to establish localization on digital elevation maps (DEMs) obtained from the Mars Reconnaissance Orbiter's HiRISE camera. Global position estimates are obtained by finding the optimal match between contours of the delineated skyline and those of rendered skylines in the rover's general region of operation, based on projected views in the digital elevation models.

The rest of this paper is organized as follows: in Section 2 we discuss the related work with focus on semantic segmentation and planetary rover localization. Our CNN-based global localization method is presented in Section 3 and experimental results are presented in Section 4. Finally, Section 5 discusses the conclusion, and future research directions.

2. Related Work

Our work lies in the intersection of semantic image segmentation and localization of planetary rovers in large-scale and unstructured environments. Thus, we review the related literature in these domains. Accurate on-board localization of planetary rovers has been an active area of research [3], [4], [5], [6]. High-precision autonomous localization is an important capability for on-line path planning and autonomous navigation of mobile planetary robots (e.g., ground or aerial robots) to ensure their safety and maximize their science return. Most of the current frameworks primarily rely on passive vision-based localization and pose estimation methods, and multiple sensing modalities can be coupled in order to increase the position estimation accuracy of the robot. In most terrestrial robotic applications, a low-cost and efficient localization method is to fuse proprioceptive data, like that from an IMU or onboard vision system [7], with GPS data. In planetary applications, where GPS signal is unavailable, vision [8], [9], or wheel odometry [10] or combination of both are used to propagate the pose of the robot [11]. While the systems have been effective in planetary operations, they are often terrain-dependent and could suffer from drift due to accumulation of errors in estimated robot motions, particularly over long-range traverses in feature-less and sandy environments [8]. Off-board localization methods using satellite imagery can be used for high-precision localization of planetary robots [12], but these methods require reliable satellite communications and make the localization pipeline critically dependent upon remote systems. Localization of a planetary exploration rover by registering ground imagery from the rover to a known aerial map is studied in a Mars analogue environment in [13].

Recently, active localization [5] and perception-aware path planning methods have been proposed [6] in order to use the feature-rich terrain in the robot's local environment to reduce localization and pose estimation error. Moreover, some recent research investigates the use of simultaneous localization and mapping (SLAM) algorithms for autonomous planetary rovers to reduce both the relative and absolute localization errors [14], [15], [16]. SLAM is a commonly used method to enable robots to create a map of an unknown environment, while localizing themselves in the map at the same time [17], [18]. Geromichalos et al. [19] propose a SLAM algorithm that relies on matching high-resolution sensor scans to the local map created online to improve relative localization. The method relies on matching the current local map to the orbiter's global map at discrete times to avoid issues with drifting in absolute localization. An adaptive visual SLAM algorithm for performing traversability analysis and global localization is presented in [15]. A visual SLAM method for planetary UAVs that registers images with known DEM data is presented in [16]. To overcome the scale and appearance difference between on-board UAV images and a pre-installed digital terrain model, topographic features of UAV images

and DEM are correlated in the frequency domain via cross power spectrum.

In [20], a method of image-based planetary rover localization is presented by comparing detected skylines in images to DEM data. However, this localization method is reliant upon a wide field of view (FOV) panoramic image, and the skyline is delineated based on luminance in grayscale images. Localizing a robot by comparing observations to known terrain maps can be studied in the context of localizing an image taken in mountainous terrain [21], [22], [23]. Typically, these methods rely on a relatively-precise prior estimate of the GPS location where the image was taken. In [22], imagery of terrestrial terrain are aligned with topographic maps using edge detection, specifically of silhouette edges. In [23], terrestrial imagery are aligned with topographic maps using semantic segmentation of the query image. However, both of these methods require geotagged imagery, indicating a relatively small region of uncertainty where the photo was taken as an input and are unsuitable for global position estimation.

In [21], a method for global localization of monocular camera images by obtaining a rough position estimate, on the range of 100 m, of an image taken anywhere in a large DEM map is introduced. The method involves using color and gradient likelihoods to detect the skyline in a query image and representing this skyline as a collection of small, normalized, overlapping sections dubbed "contourlets". These are compared to DEM-generated contourlets found by rendering a 360-degree FOV projected skyline from an xy grid of points at ~ 100 m spacing to select top location candidates. The entire skyline of the query image is then compared with the ICP algorithm [24] to corresponding FOVs in the top render candidates using a sliding window and the top ICP match is selected as the most-likely location and orientation of the image. In [25] and [26], a CNN-based approach to finding skylines trained on labels generated through Canny edge detection [27] and Hough Voting [28] is presented. The method adapts the MOSSE correlation filter [29] for determining a position estimate of the query frame with GPS-level accuracy by rendering a view based on DEM data from a known camera heading and FOV at each point in an xy grid of points surrounding the true location of the vessel. The MOSSE filter correlation score between the query image and each rendered view is computed, with the final position estimate based on a second-order polynomial fit of the maximum MOSSE correlation scores in the position search grid.

Semantic segmentation is a means of understanding an image at the pixel level. That is, to predict a class label representing each pixel in an image and define connected components of pixels with the same label [30], [31]. DeepLab uses Deep Convolutional Neural Networks for performing semantic segmentation [32], [33], [34], [35]. The current version of DeepLab utilized in this paper, DeepLab V3+, incorporates Atrous Convolution, Fully Connected Conditional Random Fields, Atrous Spatial Pyramid Pooling, and encoder-decoder modules. Alternatives to semantic segmentation for finding connected components in images include modern grab-cut style segmentation implementations based on Graph Cut [36], [37], such as [38]. These methods, however, typically rely on some user input to perform the object or foreground-background segmentation. Grab-cut based object segmentation could be coupled with CNN-based object detection methods to remove the need for human input in the segmentation pipeline, but this would not be an efficient alternative to Deep Learning-based tools developed for semantic segmentation (e.g., DeepLab) for performing object segmentation with pixel-wise labels. There are modern alternative network architectures that compete with and occasionally outperform DeepLab V3+ in semantic segmentation [39], [40], [41]. However, DeepLab V3+ was selected due to the high performance, the open-source tensorflow implementation, and the well documented user instructions. In [42] and [43], a method and dataset are presented for semantic segmentation of Martian terrain into seventeen terrain categories. Notably, sky is not included as a class in these works, as their purpose is for traversability analysis, rather than localization, based on the segmented terrain.

3. METHODOLOGY

Semantic Segmentation

Semantic segmentation is performed using the most up-to-date open-source version of DeepLab V3+ [35]. We perform fine-tune training of existing models, namely MobileNet-v2 [44] and Xception65 [45] pretrained on the ADE20k dataset [46], on domain specific data composed of 3-channel monocular camera images of Martian landscapes taken by the Curiosity Rover, selected from NASA JPL's publicly available Planetary Data System (PDS) Image Atlas [47]. 24 images were selected and annotated using the labelme tool [48], with 20 used for training and 4 for validation. The model was trained for 750 iterations, with a batch size of 2 images. This is a short fine-tune training for a semantic segmentation model, as the model quickly learns to distinguish between the two classes it is trained to identify at the pixel level, "terrain" and "sky". Inference is performed with the trained model to generate semantic segmentation prediction maps of query images, such that each pixel is identified and labeled as one of the two aforementioned classes.

Test Data Selection

The test data is composed of 39 monocular camera images of Martian skylines taken by the Perseverance Rover Mastcam-Z, selected from the publicly available PDS Image Atlas [47]. The semantic segmentation model was trained on images taken by the Curiosity rover, while all experimental results presented in this paper use images taken by the Perseverance rover to ensure sufficient distinction between the training and test sets and to demonstrate the generalization capabilities of our method. This test set is the newest publicly available data, released on August 20, 2021, and is composed of images and corresponding labels, including ground truth location and orientation, for the first 90 sols $(0 - 89)$ of the Perseverance mission on Mars. A hand-selected set of images containing Martian terrain and skylines is used rather than a random sampling, as the majority of images in the set do not contain skylines, which are necessary for our localization method.

Skyline Delineation

The "skyline" contour in query images are delineated by finding the highest pixel in each column of the "terrain" cluster in the semantic segmentation prediction map. If any columns are segmented entirely as "sky", the contour is assigned a value of "NaN" in that spot such that a lack of detected "terrain" will not impact position estimation. However, our set of images was selected such that a visible skyline spanning the width of the image is present in all samples. Currently, the semantic prediction map is used rather than the semantic probability map for skyline delineation, though a probabilistic representation of the skyline based on the joint probability distributions of "terrain" and "sky" classes is under consideration for future investigation. Optionally,

we can cluster all pixels in a frame labeled as "terrain" with the Density-based spatial clustering of applications with noise (DBSCAN) algorithm [49] to remove noisy terrain pixel detections from the skyline contours. However, we did not find this step necessary with our Martian landscape model and have omitted it to decrease run time of localization calculations.

DEM Renders

To compare the detected skyline contour to a known map based on DEM data, the DEM data is converted to the same format as the detected skyline contour. An open-source tool named "Horizonator" [50] is used to render terrain data to simulate the view that would be captured by the rover's camera from a given location, orientation, and FOV. This tool uses an equi-rectangular projection and assumes the planet is flat locally, which is not problematic in our application with relatively short view horizons. Parameters for this tool can be tuned, and for our application we generate a 3-channel image displaying a 360-deg FOV projected view with a 5 km view horizon from ground level at each point in an xy grid at 100 m spacing. The skyline contours from the DEM renders are delineated similarly to those of the semantic segmentation predictions. However, in lieu of performing semantic segmentation, a binary mask segmenting all rendered terrain from all background pixels is used.

Location Estimation

In our implementation, a rough, global position estimate is obtained using a method loosely inspired by [21]. Before localizing any query images, a prior is established by pre-processing DEM data in the rover's region of operation to reduce position estimation run-time. The prior processing steps include:

1. Selecting a grid of candidate locations for localizing the rover, with span and resolution based on application (e.g., 4 km^2 at 100 m resolution)
2. Downloading the DEM data covering the span of points, with an extension in all directions equal to the view horizon (e.g., 5 km)
3. Rendering a 360-degree FOV projected view from each grid point
4. Extracting the skyline contours from the rendered views

To estimate the position of each query image, the delineated skyline is compared to a sliding window of the 360-degree FOV rendered skylines. The corresponding FOV (i.e., window size) and camera pitch are found in the accompanying image label and are considered known, while the camera roll is assumed to be 0 deg. The comparison at each step of the sliding window (e.g., 1-deg steps) is calculated using root-mean-square error (RMSE) on the equal length skyline contours to find the best view angle at each point, a deviation from contourlet matching and ICP used in [21]. Using the best view angle from each candidate point, the position scores, $P_{x,y}$, are calculated through inversion, normalization with the best candidate, and exponentiation such that all scores are ≤ 1.0 and the distribution of scores better distinguishes the top candidates, as given by $P_{x,y} = (\frac{RMSE_{min}}{RMSE_{x,y}})^2$. Currently, the predicted heading and location are selected at the heading and position where the overall minimum RMSE was observed. Given the discrete representation of candidate locations, an improvement would be extracting the local maximum of the region of highest position score density based on approximating the grid of position scores as a second-order polynomial surface, as in [25].

4. Experimental Results

The method is tested on 39 3-channel monocular camera images of Martian skylines taken by the Perseverance Rover Mastcam-Z, selected from the publicly available PDS Image Atlas [47] as described in Section 3. The resolution and FOV of these images may vary, with typical resolutions and FOVs of 1648 x 1200 pixels and 20.4 x 14.8 deg. Details on image capture and pre-processing steps, the Perseverance Rover Mastcam-Z, and the image labels are available at [47].

Semantic Segmentation

Semantic segmentation between martian terrain and sky proves to be a fairly easy to solve semantic segmentation problem, as seen in Figure 3 - examples 1 and 2. The martian terrain in our landscapes is feature and texture-rich, while the sky is devoid of any features or texture, suggesting that the model quickly learned to segment based on the amount of texture in a region. Because of this, our semantic segmentation model may incorrectly segment distant, hazy mountain ranges (see Figure 3 - example 3) or texture-less faces of large, smooth boulders (see Figure 3 - example 4). Due to the small training set of 20 images that did not include these environments, our model did not learn to segment these regions as "terrain". Unlike traditional methods for skyline delineation based on, e.g., Canny Edge Detection, this can easily be overcome by fine-tuning the model with a more diverse train set including a few examples of these types of environments. Neither of these errors proves problematic for the skyline delineation and position estimation results presented later in this section, as segmenting the patches of the boulders as sky does not affect the delineated skyline and the distant, hazy mountain range is past the 5 km view horizon

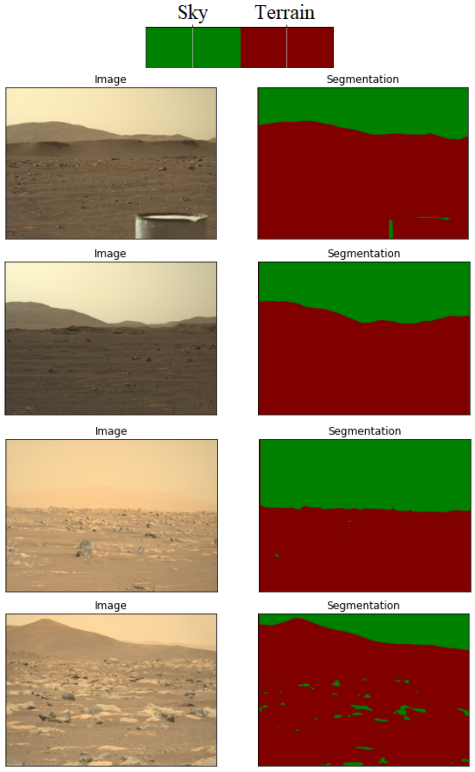

Figure 3: Semantic segmentation inference results demonstrated on various images collected by the Mars Perseverance Rover, with corresponding color map.

used in the rendered views.

Skyline Delineations

To demonstrate the utility of this method, we compare the delineated skyline of the Martian terrain image, obtained from the semantic segmentation prediction, to the corresponding delineated skyline of the rendered view with matching ground truth location, camera orientation, and FOVs. In general the skyline delineation method performs very well, with accurately delineated skylines in the images taken by the Perseverance rover. Additionally, the corresponding rendered views and skylines closely match, with full skyline matching RMSE errors of $\sim$ 10 pixels, corresponding to $\sim$ 0.3 degrees, detected for the majority of test images. The notable exception is the result seen in Figure 4 - example 4 in which a portion of the rendered skyline is missing due to a rendering error. This error can occur if the landscape is either outside of the region covered by the DEM used to generate the renders or is outside the view horizon (i.e., maximum distance to render), a parameter selected during the rendering step. For skylines outside the view horizon, this may be permissible as the semantic segmentation prediction will likely match (see Figure 3 - example 3). In this example, however, the region omitted is correctly segmented as terrain, leading to a full skyline matching RMSE error of $\sim$ 2.0 degrees. This will cause the ground truth location and orientation to not be selected in the position estimation results, but can be resolved by including these regions in the rendered views with full DEM coverage and well-selected view horizons.

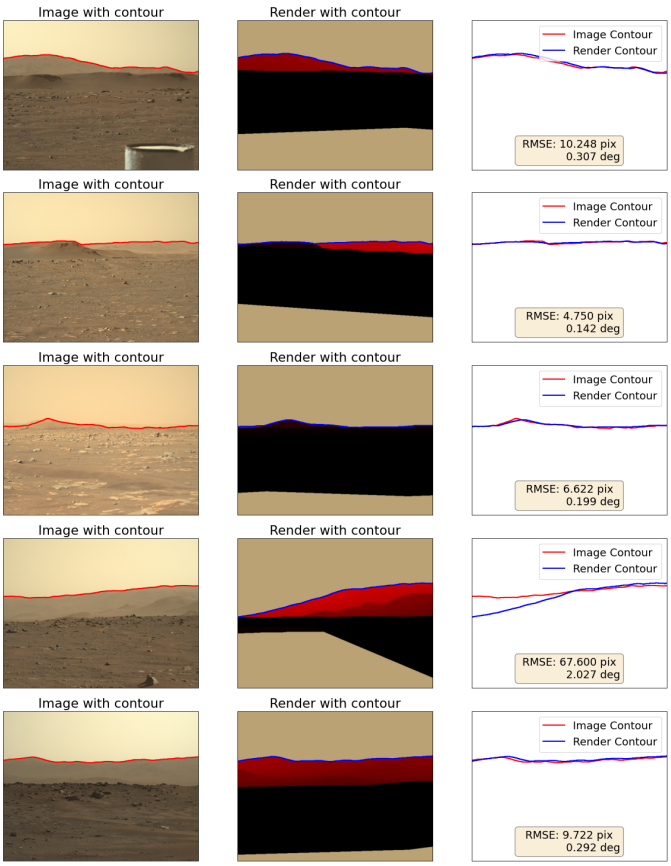

Figure 4: Image and corresponding rendered view, generated with the Horizonator tool using ground truth information from the image label, with delineated skylines.

Location Estimation

In Figure 5, we show that our skyline delineation method using semantic segmentation predictions can be used to obtain an accurate global position estimate. These results show an xy grid of candidate location points, with position scores based on how accurately the delineated skyline matches with a corresponding window of the rendered skyline, as detailed in Section 3. The delineated skyline in this example contains unique contours that closely match with those in the corresponding rendered views from the ground truth location of the rover, and does not closely match with those rendered from other position candidates. Additionally, the estimated heading is correctly identified within ± 1 degree, which corresponds to the step size used for RMSE comparison.

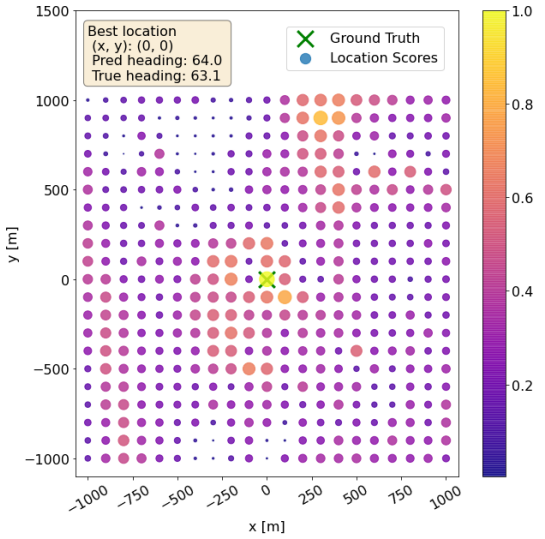

Figure 5: Successful location and orientation estimate, with respect to Perseverance rover landing site. 1.0 is the maximum position score.

However in the absence of salient features in the delineated skyline, our method will not detect a precise and accurate position estimate. In Figure 6 we show that our position estimation method assigns a relatively high position score to a large portion of xy grid points in such situations. The delineated skyline from the image in this example contained flat, uninteresting contours that loosely and relatively evenly matched some window from many of the location candidate points. To autonomously detect and reject such position estimates from state estimation filters, one could calculate the standard deviation on the distribution of high scoring position candidates and reject those with standard deviations above a threshold.

5. Conclusion and Future Work

In this paper we consider a method for onboard, automated global position estimation of planetary robotic explorers by performing semantic segmentation for skyline delineation of martian terrain images. We demonstrate that the proposed method of using semantic segmentation for skyline delineation is a well-founded method with accurate performance by comparing delineated skylines to those of rendered views based on DEM data from a rover's ground truth position and FOV. We also show that our global localization method can

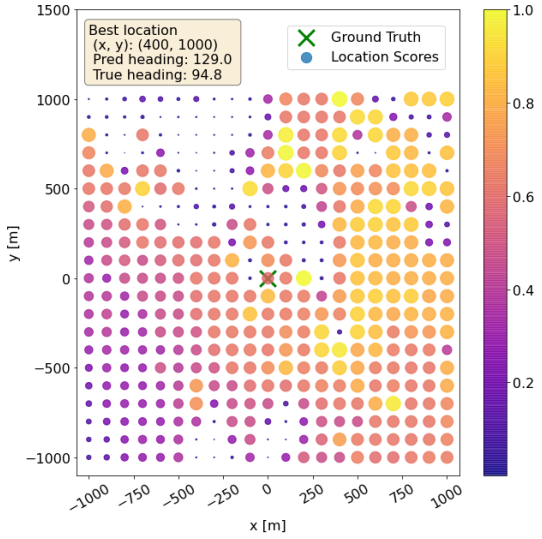

Figure 6: Imprecise location and orientation estimate, with respect to Perseverance rover landing site. 1.0 is the maximum position score.

correctly localize images in a Mars body-fixed coordinate system based on these delineated skylines by comparing them to delineated skylines from rendered views in a rover's general region of operation (e.g. a 4 km^2 region surrounding the rover landing site). In this method, unique and distinctive contours of all landscape features in a rover's local environment, including peaks, boulders, and general topography, are efficiently and robustly extracted and used for global position estimation. However, our method is dependent upon the presence of salient features in the delineated skyline, as the absence of such features can lead to perceptual aliasing and position ambiguity.

For this reason, our future work will strive to increase robustness of our localization method by investigating a method to extract peaks from the semantically segmented skyline and correlate these peaks to known landmarks. In this way, only the salient features will be used for localization and the regions of flat, nondescript skyline will be omitted. Additionally, our future work will be towards perception aware localization, with global path planning and camera orientations influenced by the amount of useful information (i.e., density and quality of salient features) expected in a region based on analysis of the DEM. These future works will be in conjunction with the future work of [51], which autonomously detects highly-visible Martian peaks and landscape features in DEM data, and calculates an upper bound on predicted localization accuracy from bearing measurements made to these landmarks through use of a QGIS plugin for viewshed analysis.

6. Acknowledgments

The research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. ©2021 California Institute of Technology. Special thanks to the Mars Perseverance entry, descent, and landing team for furnishing Mars DEMs used in this work. Particular thanks to Andrew Johnson, NASA JPL's lead for Mars2020 Terrain Relative Navigation.

References

[1] S. W. S. R. E. A. B. A. A. J. B. Y. C. L. C. e. a. Li, Rongxing, "Initial results of rover localization and topographic mapping for the 2003 mars exploration rover mission." *Photogrammetric Engineering Remote Sensing*, vol. 71, no. 10, pp. 1129–1142, 2005.

[2] K. H. W. K. M. S. R. B. A. C. M. d. l. T. K. H. e. a. Farley, Kenneth A., "Mars 2020 mission overview." *Space Science Reviews*, vol. 216, no. 8, pp. 1–41, 2020.

[3] L. Matthies and S. Shafer, "Error modeling in stereo navigation." *IEEE Journal on Robotics and Automation*, vol. 3, no. 3, pp. 239–248, 1987.

[4] M. L. H. S. M. Olson, C. F. and M. W. Maimone, "Rover navigation using stereo ego-motion." *Robotics and Autonomous Systems*, vol. 43, no. 4, pp. 215–229, 2003.

[5] O. M. T. S. Inoue, H. and S. Adachi, "Active localization for planetary rovers," *IEEE Aerospace Conference*, pp. 1–7, 2016.

[6] K. O. Strader, Jared and A. Agha-mohammadi, "Perception-aware autonomous mast motion planning for planetary exploration rovers." *Journal of Field Robotic*, vol. 3, no. 5, pp. 812–829, 2020.

[7] Y. Zhao, X. Wang, Q. Li, D. Wang, and Y. Cai, "A high-accuracy autonomous navigation scheme for the mars rover," *Acta Astronautica*, vol. 154, pp. 18–32, 2019.

[8] C. Y. Maimone, M. and L. Matthies, "Two years of visual odometry on the mars exploration rovers." *Journal of Field Robotics*, vol. 24, no. 2, pp. 169–186, 2007.

[9] G. S. C. Y. Johnson, A. and L. Matthies, "Robust and efficient stereo feature tracking for visual odometry." *IEEE International Conference on Robotics and Automation*, pp. 39–46, 2008.

[10] J. C. A. R. V. R. C. A. C. J. B. R. B. D. F. B. e. a. Grotzinger, John P., "Mars science laboratory mission and science investigation." *Space science reviews*, vol. 170, no. 1, pp. 5–56, 2012.

[11] D. S. C. L. H. M. S. R. e. a. D. M. Helmick, Y. Cheng, "Path following using visual odometry for a mars rover in high-slip environ-ments." *IEEE Aerospace Conference*, vol. 2, pp. 772–789, 2004.

[12] R. Li, S. He, Y. Chen, M. Tang, P. Tang, K. Di, L. Matthies, R. E. Arvidson, S. W. Squyres, L. S. Crumpler, T. Parker, and M. Sims, "Mer spirit rover localization: Comparison of ground image– and orbital image–based methods and science applications," *Journal of Geophysical Research: Planets*, vol. 116, no. E7.

[13] K. Ebadi and A.-A. Agha-Mohammadi, "Rover localization in mars helicopter aerial maps: Experimental results in a mars-analogue environment," in *Proceedings of the 2018 International Symposium on Experimental Robotics*, J. Xiao, T. Kröger, and O. Khatib, Eds. Cham: Springer International Publishing, 2020, pp. 72–84.

[14] P. P. Hidalgo-Carrió, Javier and F. Kirchner, "Adaptive localization and mapping with application to planetary rovers." *Journal of Field Robotics*, vol. 35, no. 6, pp. 961–987, 2018.

[15] M. Azkarate, L. Gerdes, L. Joudrier, and C. J. Pérez-del-Pulgar, "A GNC architecture for planetary rovers with autonomous navigation capabilities,"

CoRR, vol. abs/1911.09975, 2019. [Online]. Available: http://arxiv.org/abs/1911.09975

[16] X. Wan, Y. Shao, and S. Li, "Planetary UAV localization based on multi-modal registration with pre-existing digital terrain model," *CoRR*, vol. abs/2106.12738, 2021. [Online]. Available: https://arxiv.org/abs/2106.12738

[17] K. Ebadi, Y. Change, M. Palieri, A. Stephens, A. H. Hatteland, E. Heiden, A. Thakur, B. Morrell, S. Wood, L. Carlone, and A. akbar Agha-mohammadi, "LAMP: Large-scale autonomous mapping and positioning for exploration of perceptually-degraded subterranean environments," *IEEE International Conference on Robotics and Automation*, 2020.

[18] K. Ebadi, M. Palieri, S. Wood, C. Padgett, and A. akbar Agha-mohammadi, "DARE-SLAM: Degeneracy-Aware and Resilient Loop Closing in Perceptually-Degraded Environments," *IEEE International Conference on Robotics and Automation*, 2020.

[19] M. A. E. T. L. G. L. P. Geromichalos, Dimitrios and C. P. D. Pulgar, "Slam for autonomous planetary rovers with global localization." *Journal of Field Robotics*, vol. 37, no. 5, pp. 830–847, 2020.

[20] S. Chiodini, M. Pertile, S. Debei, L. Bramante, E. Ferrentino, A. G. Villa, I. Musso, and M. Barrera, "Mars rovers localization by matching local horizon to surface digital elevation models," in *2017 IEEE International Workshop on Metrology for AeroSpace (MetroAeroSpace)*, 2017, pp. 374–379.

[21] G. Baatz, O. Saurer, K. Köser, and M. Pollefeys, "Large scale visual geo-localization of images in mountainous terrain," in *Computer Vision – ECCV 2012*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 517–530.

[22] L. Baboud, M. Čadík, E. Eisemann, and H.-P. Seidel, "Automatic photo-to-terrain alignment for the annotation of mountain pictures," in *CVPR 2011*, 2011, pp. 41–48.

[23] G. Baatz, O. Saurer, K. Köser, and M. Pollefeys, "Leveraging topographic maps for image to terrain alignment," in *2012 Second International Conference on 3D Imaging, Modeling, Processing, Visualization Transmission*, 2012, pp. 487–492.

[24] P. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[25] B. Grelsson, A. Robinson, M. Felsberg, and F. S. Khan, "HorizonNet for visual terrain navigation." *IEEE International Conference on Image Processing, Applications and Systems (IPAS)*, vol. 12, no. 1, pp. 149–155, 2018.

[26] ——, "Gps-level accurate camera localization with horizonnet," *Journal of Field Robotics*, vol. 37, no. 6, pp. 951–971.

[27] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-8, no. 6, pp. 679–698, 1986.

[28] P. V. Hough, "Method and means for recognizing complex patterns."

[29] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2544–2550.

[30] B. Li, Y. Shi, Z. Qi, and Z. Chen, "A survey on semantic segmentation," in *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2018, pp. 1233–1240.

[31] M. Thoma, "A survey of semantic segmentation," *CoRR*, vol. abs/1602.06541, 2016. [Online]. Available: http://arxiv.org/abs/1602.06541

[32] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Semantic image segmentation with deep convolutional nets and fully connected crfs," *arXiv preprint arXiv:1412.7062*, 2014.

[33] ——, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," 2017.

[34] L. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *CoRR*, vol. abs/1706.05587, 2017. [Online]. Available: http://arxiv.org/abs/1706.05587

[35] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *ECCV*, 2018.

[36] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary region segmentation of objects in n-d images," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 1, 2001, pp. 105–112 vol.1.

[37] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut -interactive foreground extraction using iterated graph cuts," *ACM Transactions on Graphics (SIGGRAPH)*, August 2004. [Online]. Available: https://www.microsoft.com/en-us/research/publication/grabcut-interactive-foreground-extraction-using-iterated-graph-cuts/

[38] K.-K. Maninis, S. Caelles, J. Pont-Tuset, and L. Van Gool, "Deep extreme cut: From extreme points to object segmentation," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 616–625.

[39] T. Takikawa, D. Acuna, V. Jampani, and S. Fidler, "Gated-scnn: Gated shape cnns for semantic segmentation," *CoRR*, vol. abs/1907.05740, 2019. [Online]. Available: http://arxiv.org/abs/1907.05740

[40] Y. Zhu, K. Sapra, F. A. Reda, K. J. Shih, S. D. Newsam, A. Tao, and B. Catanzaro, "Improving semantic segmentation via video propagation and label relaxation," *CoRR*, vol. abs/1812.01593, 2018. [Online]. Available: http://arxiv.org/abs/1812.01593

[41] H. Wu, J. Zhang, K. Huang, K. Liang, and Y. Yu, "Fastfcn: Rethinking dilated convolution in the backbone for semantic segmentation," *CoRR*, vol. abs/1903.11816, 2019. [Online]. Available: http://arxiv.org/abs/1903.11816

[42] B. Rothrock, R. Kennedy, C. Cunningham, J. Papon, M. Heverly, and M. Ono, *SPOC: Deep Learning-based Terrain Classification for Mars Rover Missions*. [Online]. Available: https://arc.aiaa.org/doi/abs/10.2514/6.2016-5539

[43] R. M. Swan, D. Atha, H. A. Leopold, M. Gildner, S. Oij, C. Chiu, and M. Ono, "Ai4mars: A dataset for terrain-aware autonomous driving on mars," in *Proceedings*

of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, June 2021, pp. 1982–1991.

[44] M. Sandler, A. G. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "Inverted residuals and linear bottlenecks: Mobile networks for classification, detection and segmentation," *CoRR*, vol. abs/1801.04381, 2018. [Online]. Available: http://arxiv.org/abs/1801.04381

[45] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *CoRR*, vol. abs/1610.02357, 2016. [Online]. Available: http://arxiv.org/abs/1610.02357

[46] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Scene parsing through ade20k dataset," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5122–5130.

[47] L. Huber, "PDS Image Atlas," https://pds-geosciences.wustl.edu/missions/mars2020/, 2014.

[48] K. Wada, "labelme: Image Polygonal Annotation with Python," https://github.com/wkentaro/labelme, 2016.

[49] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise." AAAI Press, 1996, pp. 226–231.

[50] D. Kogan, "Horizonator!" 2021, last accessed 23 April 2021. [Online]. Available: https://github.com/dkogan/horizonator

[51] J. Vander Hook, R. Schwartz, K. Ebadi, K. Coble, and C. Padgett, "Topographical landmarks for ground-level terrain relative navigation on mars," in *2022 IEEE Aerospace Conference*, 2022.

BIOGRAPHY

Dr. Kamak Ebadi *is currently a Robotics Technologist at NASA JPL, Pasadena, CA, USA. He received his Ph.D degree in Computer and Electrical Engineering from Santa Clara University in 2020, and was a Postdoctoral Fellow with NASA JPL-California Institute of Technology from 2020 to 2021. His research interests include computer vision, multi-robot perception and autonomy in perceptually degraded and extreme environments.*

Kyle Coble *received his B.S. degree in Mechanical Engineering in 2016 from Cornell University. He is scheduled to complete an M.S. degree in Systems, Control, and Robotics from the KTH Royal Institute of Technology in late 2021. His interests include computer vision and state estimation for autonomous robots in extreme environments.*

Dima Kogan *has a Master's degree in Control and Dynamical Systems from California Institute of Technology. He has many years of industry experience working on tracking, calibration, path planning, mapping, data analysis, visualization, software infrastructure, embedded development and board design. He has 20+ years of professional software development and is an active contributor to many Free Software projects.*

Deegan Atha *is a robotics technologist in the Perception Systems group. He joined after completing his Bachelors in Electrical Engineering from Purdue University. Prior to JPL, he was at NASA Langley, where he researched perception algorithms for UAVs. While at Purdue, he conducted research on robotic visual inspection of structures.*

Russell Schwartz *is an undergraduate at the University of Maryland in his senior year, scheduled to graduate with a dual B.S. in Mathematics and Computer Science in Spring 2022. He is also an intern at JPL and plans on pursing a Master's degree in Robotics beginning the following Fall.*

Dr. Curtis Padgett *is the Supervisor for the Maritime and Aerial Perception Systems Group. He completed his doctoral work in pattern recognition while working at JPL. His interests include computer vision, structure from motion, automated calibration, pattern classification, star identification and machine learning.*

Dr. Joshua Vander Hook *received a PhD in Computer Science at the University of Minnesota in 2015. His research involved designing autonomous robots that can assist in surveying and data-gathering in remote areas. He held a Doctoral Dissertation Fellowship at the University of Minnesota.*