Toward Autonomous Localization of Planetary Robotic Explorers by Relying on Semantic Mapping

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, kwc57@cornell.edu, dmitriy.kogan@jpl.nasa.gov, deegan.j.atha@jpl.nasa.gov, rschwa63@umd.edu, curtis.w.padgett@jpl.nasa.gov, hook@jpl.nasa.gov

***Abstract*—Highly accurate localization of planetary robotic explorers is crucial for robust, efficient, and safe path planning in unknown and extreme 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 in satellite imagery. While these methods have proven to be effective, they rely heavily on a human-in-the-loop. This paper is concerned with autonomous global localization of planetary robotic explorers in extreme and GPS-denied environments by relying on semantic segmentation of ground imagery. Using a trained convolutional neural network (CNN), saliency maps are obtained from semantic segmentation of ground imagery. These maps are then 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.**

TABLE OF CONTENTS

1. INTRODUCTION 1
2. RELATED WORK 2
3. METHODOLOGY 3
4. EXPERIMENTAL RESULTS 5
5. CONCLUSION AND FUTURE WORK 7
6. ACKNOWLEDGMENTS 7
REFERENCES 7
BIOGRAPHY 9

1. INTRODUCTION

In planetary missions, accurate landing site localization in a global reference system is crucial for achieving safe and efficient long-term autonomy. Since Global Positioning Systems are unavailable in planetary applications, localization is achieved by matching perceptual features in terrain images to an orbital map, analyzing X-band Doppler radio transmissions using satellites orbiting the planetary body, or by direct observation of the landing site from orbit. 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 the ground team achieved localization with a sufficient level of accuracy, it required significant human intervention. Though past Mars missions disqualified scientifically compelling landing sites due to hazardous terrain within the landing ellipse, NASA's most recent Mars exploration mission, the Perseverance rover, relied on Terrain Relative Navigation (TRN) in order to land more precisely in a challenging but geologically interesting area. 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. As shown in Figure 1, the TRN method shrunk the area of the landing ellipse of uncertainty for the Perseverance

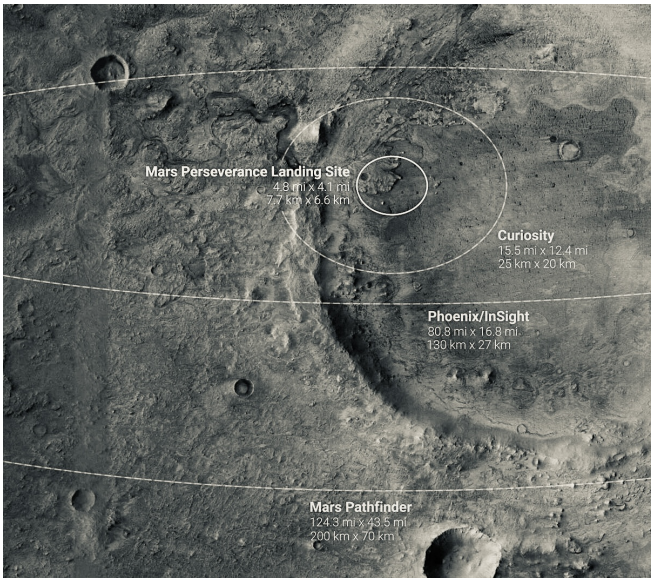

Figure 1: Visualization of landing ellipse size for Mars Pathfinder, Phoenix, InSight, and Curiosity, on the target landing area of NASA's Perseverance rover on Mars, Jezero Crater. Credit: NASA JPL-Caltech.

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. ©2022 California Institute of Technology.

2022 IEEE Aerospace Conference (AERO) | 978-1-6654-3760-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/AERO53065.2022.9843550

rover by a factor of ten, as compared to the Mars Curiosity mission, and 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].

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. A fast and accurate localization method to decrease the robot position estimation uncertainty can greatly benefit a planetary robotic mission and increase the mission's science return. For instance, the Perseverance rover can benefit from long-range traverses at higher speeds as it is tasked with gathering a diverse set of martian rock and regolith samples that could be returned to Earth by the Mars Sample Return campaign being planned by NASA and the European Space Agency. It is advantageous to enable fully on-board localization capability, not reliant on external satellite communications, to be used for landing site localization and periodic robot localization during the mission to reduce accumulated error 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], wheel odometry [10], or a combination of both are used to propagate the pose of the robot [11]. Localization of a planetary exploration rover by registering rover's terrain imagery to a known aerial map is studied in a Mars analogue environment in [12] where salient perceptual features from ground imagery are registered to an orbital map to achieve localization. While these methods can be effective in planetary applications, they are often terrain-dependent and could suffer from drift due to accumulation of odometry errors in estimated robot motions, particularly over long-range traverses in feature-less environments with smooth sand [8]. Off-board localization methods using satellite imagery have been used for high-precision localization of planetary robots [13], but these methods require reliable satellite communications and make the localization pipeline critically dependent upon remote systems.

Recently, active localization [5] and perception-aware path planning methods have been proposed [6] in order to exploit the feature-rich terrain in the robot's local environment to reduce localization and pose estimation error of onboard vision-based system. 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], [19]. Geromichalos et al. [20] 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 [21], a method of image-based planetary rover localization is presented by comparing

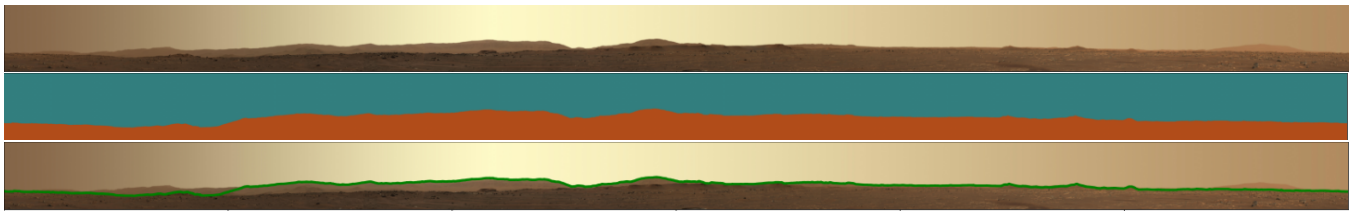

Figure 2: Semantic segmentation and skyline delineation demonstrated on a Martian panorama (top), with corresponding semantic segmentation prediction map (middle) and delineated skyline contour (green, bottom).

detected skylines in images to DEM data, where the 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 [22], [23], [24]. Typically, these methods rely on a relatively-precise prior estimate of the GPS location where the image was taken. In [23], imagery of terrestrial terrain are aligned with topographic maps using edge detection, specifically of silhouette edges. In [24], terrestrial imagery are aligned with topographic maps using semantic segmentation of the query image. While both methods show promising performance, they rely on 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 [22], 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 x-y grid of points at $\sim$ 100 m spacing to select top location candidates. Using the ICP algorithm [25], the entire skyline of the query image is then compared with 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 [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] to determine 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 x-y 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 the work presented by Maninis et al. in [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], [42]. However, DeepLab V3+ was selected due to the high performance, the open-source tensorflow implementation, and the well documented user instructions. In [43] and [44], a method is 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

In this section, we will show how saliency maps obtained from semantic segmentation of terrain imagery can be used for global localization of a rover in a Mars body-fixed coordinate system. Figure 3 presents our proposed semantic segmentation-based localization pipeline. In the rest of this section, we will introduce each component of the pipeline.

Semantic Segmentation

Semantic segmentation is performed using the most up-to-date open-source version of DeepLab V3+ [35], an existing technology. We take common architectures, namely MobileNet-v2 [45] and Xception65 [46], pretrained on the ADE20k dataset [47] and perform a fine-tune training 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 [48]. 24 images were selected and annotated using the labelme tool [49], 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,

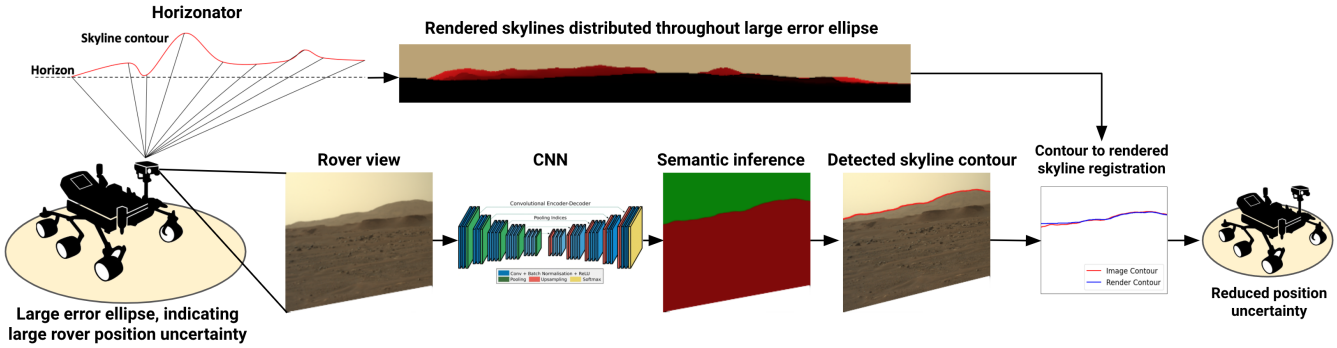

Figure 3: An overview of the proposed semantic segmentation-based localization pipeline.

terrain and *sky*, it is trained to identify at the pixel level. 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 [48]. The semantic segmentation model was trained on images taken by the Curiosity rover, while all experimental results presented use images taken by the Perseverance rover to ensure sufficient distinction between the training and test sets. This test set, from the first 90 sols $(0 - 89)$ of the Perseverance mission, was the newest publicly available data at the time of experimentation, released on August 20, 2021, and is composed of images and corresponding labels including ground truth location and orientation. A hand-selected set of images containing Martian skylines is used rather than a random sampling, as the majority of images in the set do not contain the skylines necessary for our localization method.

Skyline Delineation

The first contribution of this paper is the use of deep learning semantic segmentation predictions for the extraction of Martian skylines. The *skyline* contour in query images are detected by finding the highest pixel in each column of the *terrain* cluster that falls directly below a pixel of the *sky* cluster in the semantic segmentation prediction map. Therefore, no value is assigned in any columns segmented entirely as *terrain* or *sky*, so that any out-of-frame regions of skyline will not impact position estimation. Extracting the contour in this way, in contrast to existing methods using either the highest terrain pixel or the lowest sky pixel in each column, increases robustness and allows for the omission of occlusions of the natural skyline by objects (e.g., other planetary exploration robots) in the camera view if an additional semantic segmentation prediction class is added. Optionally, all pixels in a frame labeled as *terrain* can be clustered with the Density-based Spatial Clustering of Applications with Noise (DBSCAN) algorithm [50] to remove noisy terrain pixel detections from the skyline contours. The accuracy of semantic predictions from our Martian landscape model enable the omission of the DBSCAN step to decrease runtime of localization calculations.

Contour Confidence

The second contribution of this paper is a point-wise contour confidence vector, based on local intensity gradients of detected Martian skylines. Each point along the detected skyline is assigned a probabilistic confidence score that evaluates the likelihood that the point on the detected skyline represents the true natural skyline, by evaluating local intensity gradients. The confidence measure is based on the difference between the maximum and minimum intensity value in a 0.25-degree vertical window centered on each skyline point. This leads to high confidence skyline points where a distinct skyline (i.e. sharp intensity change from terrain to sky) is visible and low confidence skyline points in regions of obfuscated (e.g., hazy) skylines, dual skylines from tiered mountain ranges, or incorrectly detected skylines. This confidence vector can be included in the position estimation calculations to reduce the negative impact any erroneously delineated segments along a skyline contour can have on the skyline-based position estimation method.

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* [51] 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

The third contribution of this paper is a method that relies on the detected Martian skylines to achieve global position estimation. A rough, global position estimate is obtained using a method loosely inspired by [22]. 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 runtime. The prior processing steps include:

1. Defining 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. Obtaining the corresponding DEM data covering the span of points, with an extension in all directions equal to or greater than the view horizon (e.g., 5 km)
3. Rendering a 360-degree FOV projected view of the rover from each grid point
4. Extracting the skyline contours from the rendered views

To estimate the position of each query image, the delineated skyline (obtained by processing the semantic segmentation prediction) is compared to a sliding window of the 360-degree FOV rendered skylines. The corresponding FOV (i.e., window size), camera pitch, and camera roll are found in the accompanying image label and are considered known. 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 the contourlet matching step and ICP comparison used in [22]. RMSE comparison is used as it generates correlated comparisons to alternative signal processing methods (e.g, ICP [25], Dynamic Time Warping (DTW) [52], [53], and FastDTW [54]) at a significantly lower runtime complexity. The skyline confidence vector can be included in the position estimation calculations by multiplying the squared error between detected and rendered skyline by the confidence score before taking the square root and mean, as in

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{width} conf(n) * (y_{det}(n) - y_{rend}(n))^2}.$$

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 $P_{x,y} = (RMSE_{min}/RMSE_{x,y})^2$. When the confidence vector is included in the RMSE calculation, the position

score for each candidate point is divided by the average of the confidence vector to penalize skylines of low overall confidence so they appropriately return a lower likelihood of the position estimate. 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 the highest position score density based on approximating the grid of position scores as a second-order polynomial surface, as in [26].

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 [48] 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 in [48].

Semantic Segmentation

Segmentation between martian terrain and sky proves to be a fairly easy problem to solve for a semantic segmentation network, as seen in Figure 4 - 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 4 - example 3) or texture-less faces of large, smooth boulders (see Figure 4 - 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 training set including a few examples of these types of environments. Neither of these errors tend to prove problematic for the skyline delineation and position estimation results presented later in this section, as the skyline delineation method is robust to small, erroneous regions of sky in the terrain cluster and the distant, hazy mountain range is past the 5 km view horizon used in the rendered views.

Skyline Delineation

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 Figure 5. In general, the skyline delineation method performs robustly, with accurately delineated skylines in the images taken by the Perseverance rover. The corresponding rendered views and skylines closely match, with full skyline matching RMSE errors of $\sim 5-15$ pixels, corresponding to $\sim 0.06-0.18$ degrees, for the majority of test images. The precision of these matches at the ground truth position and orientation is paramount to the success of the global position estimation method. The notable exception is the result seen in Figure

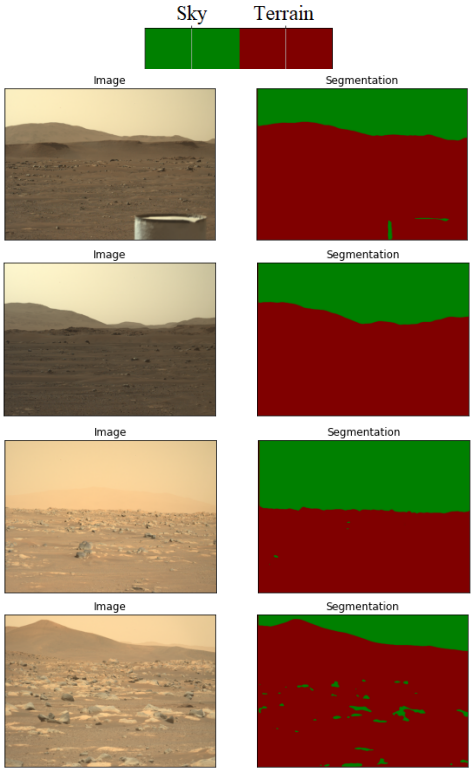

Figure 4: Semantic segmentation inference results demonstrated on various images collected by the Mars Perseverance Rover, with corresponding color map.

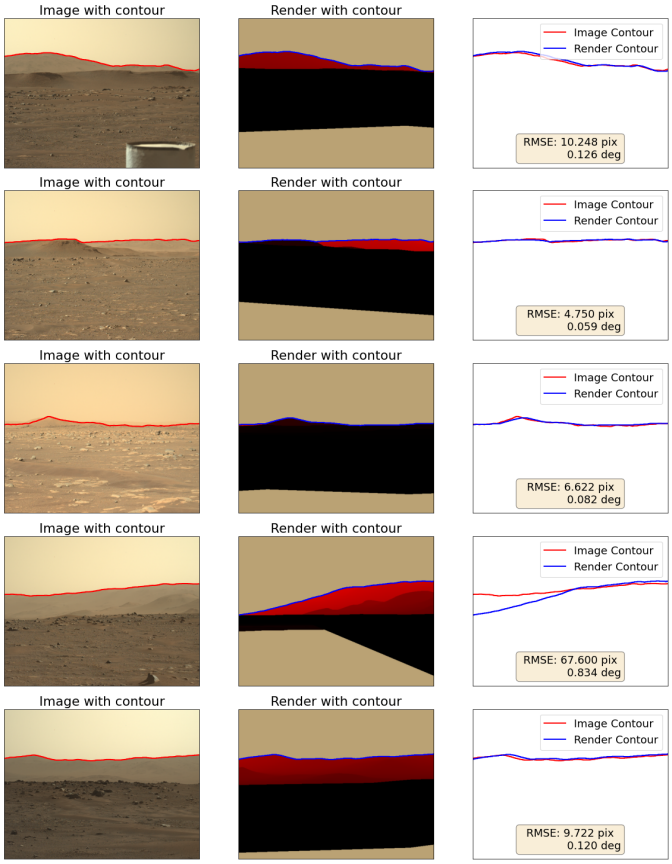

Figure 5: Image and corresponding rendered view, generated with the Horizonator tool using ground truth information from the image label, with delineated skylines.

5 - 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 may match, depending on visibility conditions. In this example, however, the region omitted is correctly segmented as terrain, leading to a full skyline matching RMSE error of ~ 0.8 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.

The contour confidence metric displayed in Figure 6 is used to automatically detect ambiguous (bottom left) or erroneous (bottom right) regions of detected skyline and reduce their impact on the position estimation calculations, as explained in Section 3. The method automatically detects such incorrectly segmented regions of skyline and reduces the point-wise confidence appropriately, while well detected skylines (top) yield perfect, or near perfect, contour confidence vectors.

Location Estimation

In Figure 7, 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 x-y 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 is a unique

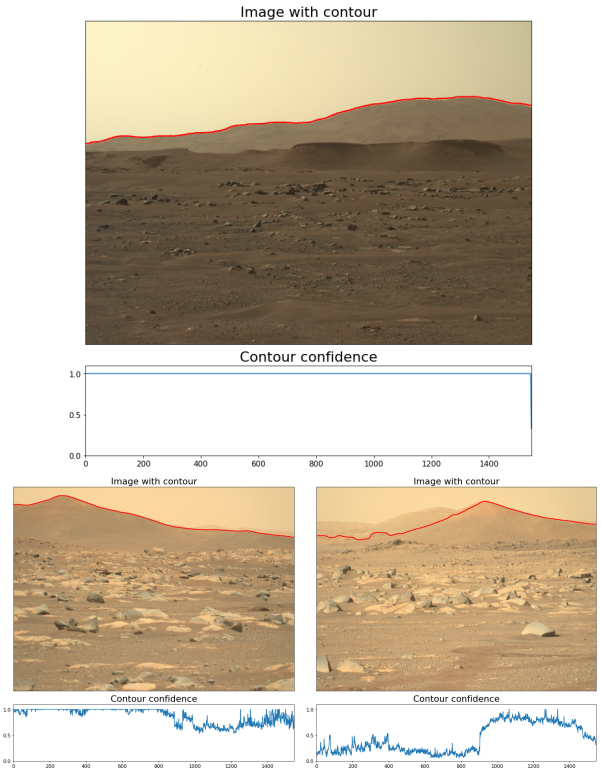

Figure 6: Skyline contour confidence with fully accurate detection (top), regions with dual skylines (left), and erroneous detection (right).

contour that closely matches the skyline in the corresponding rendered view from the ground truth location of the rover. The image localized in Figure 7 corresponds to the contour in Figure 5 - example 3, the second best match of the test set, with an RMSE of 6.622 pixels or 0.082 degrees in the image FOV between detected and rendered contour. The correct position of the rover is detected and the camera orientation is correctly identified within ± 1 degree, which corresponds to the step size used for RMSE comparison. Additionally, we find that the sliding-window RMSE contour comparison method can be used to reliably estimate camera orientation for well segmented skylines, even in the presence of ~ 1 km of positional uncertainty.

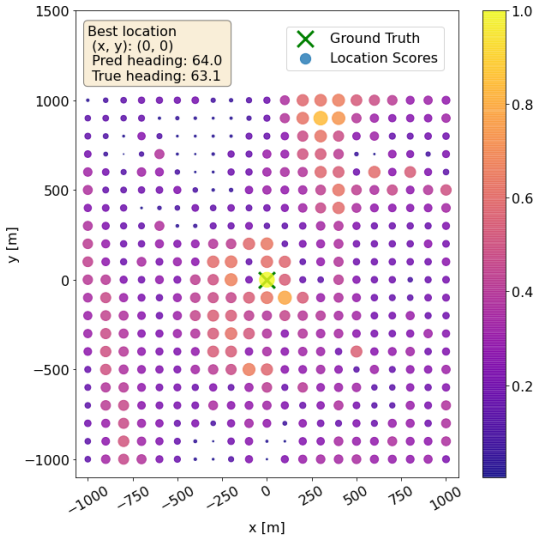

Figure 7: Successful location and orientation estimate, with respect to Perseverance rover landing site. 1.0 is the maximum position score.

However in the absence of salient features in the delineated skyline, our method will not detect a precise and accurate position estimate. In Figure 8 we show that our position estimation method assigns a relatively high position score to a large portion of x-y 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.

Runtime Considerations

Semantic segmentation model training time will not have an impact on image localization runtime, as the model can be trained offline using suitable hardware on Earth and uploaded to the robot. Likewise, the prior processing related to building the database of render contours for comparison can be performed prior to any image localization and will not impact the time to localize a given image. Evaluation of true runtime performance for localization of each image in deployment will need to be tested and verified on the computing hardware selected for future planetary exploration robots. However, it is still useful to approximate the order on which the algorithm runtime will be. In [55] it is demonstrated that semantic segmentation inference with ICNet on images of 384×576

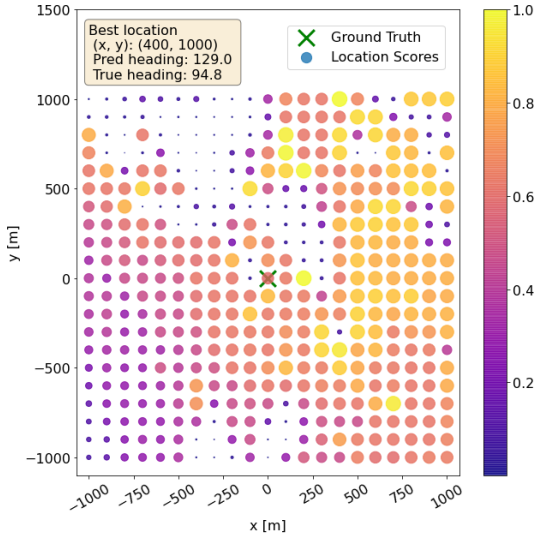

Figure 8: Imprecise location and orientation estimate, with respect to Perseverance rover landing site. 1.0 is the maximum position score.

resolution runs at 1299 ms on a Snapdragon 800, 930 ms on a Snapdragon 810, 313 ms on a Snapdragon 820, and 365 ms on a Snapdragon 821. The Mars Ingenuity Helicopter runs on a Snapdragon 801 which performs similarly to, if not slightly better than, the Snapdragon 800. Considering the proposed method does not require a specific semantic segmentation framework (DeepLab V3+ was chosen for convenience and proven accuracy), it is reasonable to expect an inference time of $\sim$ 1300 ms assuming comparable image resolution, the use of ICNet for semantic segmentation, and the same hardware used on the Ingenuity Helicopter.

Runtime of the position estimation calculations scales linearly with the number of position candidate points. In our experiments, we present localization in a 4 km^2 area at 100 m spacing, leading to 441 candidate points. On a single Intel(R) Core(TM) i5-8250U CPU @ 1.60GHz, position estimation of a frame in reference to 441 candidate points runs in 387.4 seconds with non-optimized Python code. A conservative runtime reduction estimate of 50$\times$ by optimization and translation to C++ would lead to $\sim$ 8 sec to localize a frame with 441 candidate points. With a landing site ellipse with axes of 6.6 km and 7.7 km at 100 m spacing, $\sim$ 5000 candidate points would be required to consider the entire ellipse. In this case, total runtime for inference and position estimation could take $\sim$ 90 sec, assuming comparable processor performance. As this localization method is designed for position correction, it is not meant for real-time operations, thus the current runtime would be acceptable. Furthermore, as position correction ellipses of uncertainty will be significantly smaller, the runtime will decrease correspondingly.

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 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. This would enable the use of only salient landmarks for localization and the regions of flat, nondescript skyline will be omitted. Additionally, our future work will be focused on 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 data. These future works will be in conjunction with the future work discussed in [56], 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. ©2022 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] L. Rongxing, S. W. Squyres, R. E. Arvidson, B. A. Archinal, J. Bell, Y. Cheng, and L. C. et al., "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. A. Farley, K. H. Williford, K. M. Stack, R. Bhartia, A. Chen, M. de la Torre, and K. H. et al., "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] C. F. Olson, L. H. Matthies, M. Schoppers, and M. W. Maimone, "Rover navigation using stereo ego-motion." *Robotics and Autonomous Systems*, vol. 43, no. 4, pp. 215–229, 2003.

[5] H. Inoue, M. Ono, S. Tamaki, and S. Adachi, "Active localization for planetary rovers," *IEEE Aerospace Con-*

ference, pp. 1–7, 2016.

[6] J. Strader, K. Otsu, 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] M. Maimone, Y. Cheng, 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] A. Johnson, S. Goldberg, Y. Cheng, 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. P. Grotzinger, J. Crisp, A. R. Vasavada, R. C. Anderson, C. J. Baker, R. Barry, and D. F. B. et al., "Mars science laboratory mission and science investigation." *Space science reviews*, vol. 170, no. 1, pp. 5–56, 2012.

[11] D. M. Helmick, Y. Cheng, D. S. Clouse, L. H. Matthies, and S. R. et al., "Path following using visual odometry for a mars rover in high-slip environments." *IEEE Aerospace Conference*, vol. 2, pp. 772–789, 2004.

[12] 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.

[13] L. Rongxing, H. Shaojun, C. Yunhang, T. Min, T. Pingbo, D. Kaichang, 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.

[14] J. Hidalgo-Carrió, P. Poulakis, 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] M. Palieri, B. Morrell, A. Thakur, K. Ebadi, J. Nash, A. Chatterjee, C. Kanellakis, L. Carlone, C. Guaragnella, , and A. akbar Agha-Mohammadi, "Locus: A multi-sensor lidar-centric solution for high-precision odometry and 3d mapping in real-time." *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 421–428, 2020.

[19] 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.

[20] D. Geromichalos, M. Azkarate, E. Tsardoulias, L. Gerdes, L. Petrou, 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[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] 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

[34] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," 2017.

[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] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 12, pp. 2481–2495, 2017.

[43] 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

[44] 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.

[45] 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

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

[47] 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.

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

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

[50] 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.

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

[52] T. K. Vintsyuk, "Speech discrimination by dynamic programming," *Cybernetics*, vol. 4, pp. 52–57, 1968.

[53] H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43–49, 1978.

[54] S. Salvador and P. Chan, "Toward accurate dynamic time warping in linear time and space," *Intell. Data Anal.*, vol. 11, no. 5, p. 561–580, oct 2007.

[55] A. Ignatov, R. Timofte, W. Chou, M. Wu, T. Hartley, and L. Van Gool, "Ai benchmark: Running deep neural networks on android smartphones," in *Computer Vision – ECCV 2018 Workshops*, L. Leal-Taixé and S. Roth, Eds. Cham: Springer International Publishing, 2019, pp. 288–314.

[56] 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.*