Semantic Segmentation for Mars Terrain Analysis

Hok Wai Chan

February 2025

1 Introduction

Autonomous planetary exploration requires advanced perception systems to navigate and analyze complex terrains effectively. Semantic segmentation plays a crucial role in terrain classification by distinguishing different surface elements. This project focuses on researching various methods for semantic segmentation and dataset integration to improve Mars terrain analysis. Multiple approaches, including dataset integration and state-of-the-art segmentation models, were explored before finalizing DeepLabV3 as the selected model for experimentation.

2 Research

The research phase of this project involved exploring multiple technologies relevant to semantic segmentation and Mars terrain analysis.

2.1 Infinigen

One of the key technologies investigated was Infinigen, a procedural generator of 3D scenes. Infinigen is designed to generate high-quality, diverse 3D training data optimized for computer vision applications. Built on Blender, it allows for the creation of realistic Mars-like terrains and features that can be used to augment existing datasets. In this project, our original plan was to utilize Infinigen to generate synthetic Mars-like environments, which were then used for training and evaluating semantic segmentation models. By incorporating procedurally generated terrains, the dataset diversity was significantly enhanced, helping the models generalize better to real-world Martian landscapes. The ability to create customizable terrains also allowed for targeted testing of model performance in specific conditions such as rocky surfaces, sand dunes, and cratered regions. This approach provided a flexible and scalable way to simulate diverse Martian landscapes, ensuring robust model evaluation across various environmental challenges.

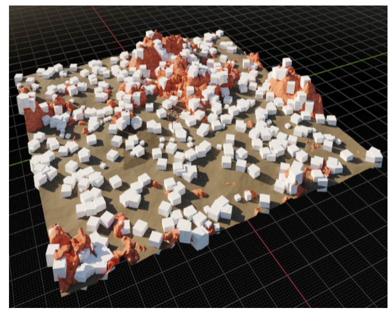


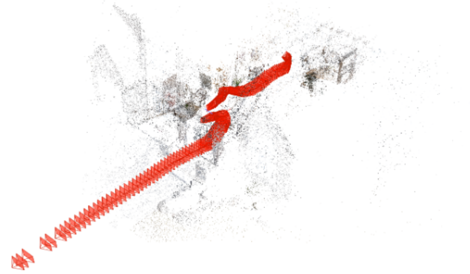
Figure 1: 3D prototype scene generated by Infinigen and Blender

2.2 COLMAP

Another critical tool explored was COLMAP, an end-to-end structure-from-motion (SfM) and multi-view stereo (MVS) pipeline. The primary purpose of using COLMAP in this project was to reconstruct 3D models from image frames, providing a deeper understanding of terrain structures and enhancing dataset generation for training semantic segmentation models. By leveraging COLMAP's SfM capabilities, we aimed to extract 3D information from 2D images, which could be used to improve model generalization by incorporating depth and structural cues into the dataset.

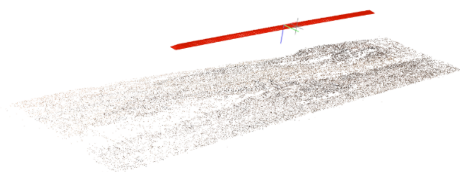
In our experiment, we used COLMAP to generate 3D maps from two different sets of image frames. The first set contained images of a residence, serving as a baseline test to verify the accuracy of the reconstruction process. The resulting 3D model of the residence demonstrated COLMAP's effectiveness in generating structured and well-defined scenes with clear geometric features. The second set consisted of Mars-like terrain images, aimed at assessing COLMAP's ability to reconstruct complex, unstructured surfaces characteristic of planetary landscapes. However, the reconstruction proved to be more challenging due to the relatively uniform and featureless nature of the Martian surface. The lack of distinct feature points in sandy and rocky regions led to difficulties in generating a coherent 3D structure, highlighting the limitations of tra-

((a)) Residence image frames



((b)) Residence COLMAP result

((c)) Mars terrain image frames



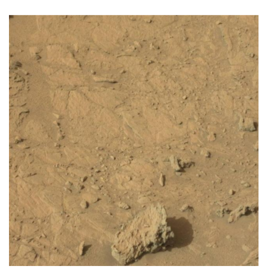
((d)) Mars terrain COLMAP result

Figure 2: COLMAP experiment results

ditional SfM approaches in environments with minimal texture variation. This experiment emphasized the importance of surface features in COLMAP's reconstruction process and suggested the potential need for additional preprocessing techniques, such as contrast enhancement or feature augmentation, to improve results on planetary terrains.

2.3 State-of-the-art models

During our research, we evaluated the Segment Anything Model (SAM) as a potential solution for semantic segmentation. While SAM has demonstrated strong general-purpose segmentation capabilities, its performance on Mars terrain imagery revealed several limitations. The model struggled with distinguishing small objects, often missing or misclassifying features due to low contrast or surrounding noise. Another challenge was its limited understanding of contextual relationships between terrain features, making it difficult to differentiate foreground from background when boundaries were subtle. Furthermore, SAM is primarily designed to generate segmentation masks rather than assigning specific class labels, making it difficult to compare its outputs with the structured ground truth data available in our dataset. While SAM performs well in general object segmentation, it does not always achieve high accuracy in semantic segmentation tasks that require precise categorization of visually similar terrain elements. Given these constraints, SAM was not incorporated into our final approach, as DeepLabV3 demonstrated better adaptability and accuracy for Mars-specific segmentation tasks.



(a) Input

(b) Output

Figure 3: SAM Results

2.4 Combining datasets

Another focus was combining datasets with new union labels to enhance segmentation accuracy. The datasets examined included MarsData-V2, S5Mars, AI4MARS, GOOSE (the German Outdoor and Offroad Dataset), RELLIS-3D, and the Artificial Lunar Landscape Dataset. By analyzing these datasets, we aimed to develop a robust dataset integration strategy for improved model training. However, this methodology was abandoned due to several difficulties encountered, such as datasets containing incorrect or missing labels, which hindered the integration process and affected the overall accuracy.

3 Methodology

Our initial strategy involved integrating multiple datasets to create a diverse and comprehensive training set for semantic segmentation. To ensure consistency across the datasets, we applied various image processing techniques such as normalization, augmentation, and color correction. We tested different architectures, including transformer-based models and convolutional neural networks (CNNs), before determining that DeepLabV3 was the most suitable model for our task. The implementation was based on the DeepLabV3FineTuning repository from GitHub. As noted earlier, the attempt to combine datasets was ultimately abandoned due to the inconsistencies and challenges it presented, leading us to focus exclusively on the S5Mars dataset.

The DeepLabV3 model was trained for 20 epochs with a batch size of 16, using CrossEntropyLoss as the loss function. The Adam optimizer was selected for its efficiency in optimizing training performance. We fine-tuned DeepLabV3 on the S5Mars dataset and explored additional preprocessing techniques to further improve segmentation accuracy. Extensive experimentation was carried out to assess the impact of various hyperparameter settings, loss functions, and data augmentation strategies. Automated evaluation metrics were integrated throughout the training process to track performance improvements across multiple iterations.

4 Dataset Details

The S5Mars dataset served as the primary dataset for experimentation. It consists of 6,000 images, with 4,800 designated for training and 1,200 for validation. Each image is paired with a ground truth segmentation mask, classifying pixels into ten categories: hole, trace, rover, rock, bedrock, sand, soil, ridge, sky, and NULL. Extensive preprocessing steps were conducted, including normalization and data augmentation techniques such as flipping and rotation to improve generalization. To ensure compatibility with different models, color-mapped segmentation labels were converted into numerical class indices.

Table 1: Color Map

RGB Value	Color Sample	ID	Class Name
[0, 64, 0]		255	NULL
[128, 0, 0]		0	sky
[0, 128, 0]		1	ridge
[128, 128, 0]		2	soil
[0, 0, 128]		3	sand
[128, 0, 128]		4	bedrock
[0, 128, 128]		5	rock
[128, 128, 128]		6	rover
[64, 0, 0]		7	trace
[0, 0, 0]		8	hole

5 Results

The highest validation accuracy achieved by the model was around 0.74, which is a promising result given the inherent challenges of segmenting Martian terrain. The low contrast and similar textures across different surface elements make it difficult to differentiate between various features, complicating the segmentation process. By analyzing the prediction images, we observe that certain classes did not perform as well as expected, particularly rover [128, 128, 128], trace [64, 0, 0], and hole [0, 0, 0]. The model often misclassified these regions, mapping them to colors belonging to other classes. This issue may be partially due to dataset imbalance, as these classes are underrepresented in the training data. The model's performance was also significantly influenced by the quality of the dataset annotations, suggesting that improvements in labeling accuracy and consistency could have a substantial impact on the results. Furthermore, expanding the dataset size could provide more diverse training examples, potentially leading to enhanced model performance. While DeepLabV3 served as a solid baseline for the segmentation task, exploring more advanced model architectures, such as newer versions or other specialized segmentation models, along with improved dataset integration techniques, could lead to further gains in segmentation accuracy.

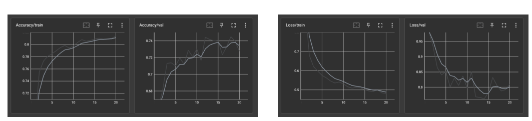


Figure 4: Accuracy and Loss

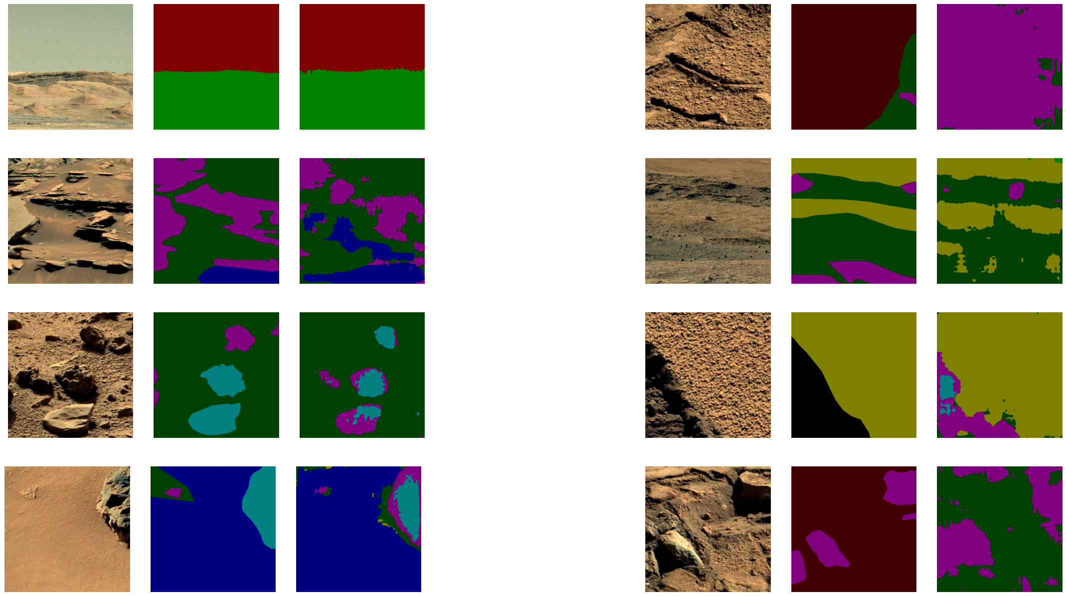


Figure 5: DeepLabV3 Results

6 Future Directions

Future research will focus on improving dataset quality and expanding model experimentation. This will include curating more high-quality images and enhancing labels through human annotation. Model optimization will explore cutting-edge architectures, such as vision transformers and multi-scale segmentation techniques. Additionally, integrating multiple datasets with consistent labeling schemes will be a key priority to boost model generalization. Ultimately, deployment on real-world rover systems and in simulated Martian environments will help validate the practical applications of these models in planetary exploration.

7 Conclusion

This research investigates various approaches to semantic segmentation for Mars terrain analysis, with an emphasis on dataset integration and model selection. While DeepLabV3 was chosen for experimentation, additional work is required to improve segmentation accuracy through better datasets and more advanced architectures. Moving forward, efforts will concentrate on dataset refinement, model optimization, and real-world deployment to push the boundaries of autonomous planetary exploration.

References

[1] ETH Zurich and UNC Chapel Hill, *COLMAP 3.12.0.dev0*, available at: `https://colmap.github.io/`, accessed: February 2025.

[2] Raistrick, Alexander and others, *Infinigen: Infinite Photorealistic Worlds Using Procedural Generation*, available at: `https://github.com/princeton-vl/infinigen`, accessed: February 2025.

[3] Piotr Skalski, *How to Use the Segment Anything Model (SAM)*, available at: `https://blog.roboflow.com/how-to-use-segment-anything-model-sam/`, accessed: February 2025.

[4] Jiahang Zhang, Lilang Lin, and others, *S5Mars: Semi-Supervised Learning for Mars Semantic Segmentation*, available at: `https://jhang2020.github.io/S5Mars.github.io/`, accessed: February 2025.

[5] Pytorch Team, *Deeplabv3*, available at: `https://pytorch.org/hub/pytorch_vision_deeplabv3_resnet101/`, accessed: February 2025.

[6] jnkl314, *DeepLabV3FineTuning*, available at: `https://github.com/jnkl314/DeepLabV3FineTuning/tree/master/sources`, accessed: February 2025.