2022 IEEE 7th International Conference on Recent Advances and Innovations in Engineering (ICRAIE) | 978-1-6654-8910-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICRAIE56454.2022.10054341

Semantic Segmentation on Martian Terrain for Navigation using Transformers

Gagan Goutham[*]
gagangoutham.n03@gmail.com

Hita Juneja[*]
hitajuneja@gmail.com

Ankitha C[*]
ankithac45@gmail.com

V R Badri Prasad[*]
badriprasad@pes.edu

[*]*PES University, Dept. of Computer Science, Bengaluru, India*

***Abstract*—For the longest time, it has been known that Earth is the only sustainable human-living planet across the universe. Spacecraft have not discovered the existence of life outside of our solar system. Given Mars' proximity and similarities to Earth, the possibility of life there is a subject of interest. Scientists have been striving to learn more about Mars for decades. At this time, it is physically impossible for humans to spend an extended period of time on Mars to study its habitat. This requires relying on extraterrestrial machinery to serve as our eyes. It involves giving machines the best training available so that they can learn how to navigate the terrain. These machines capture and send forth data which is processed to morph into mapping the sphere, forecasting paths, analyzing the landscape and even improving the same spacecraft's capabilities on learning more about the planet it is on.**

The rovers' in-situ landscape images are the main subject of this study. In this research, a novel approach has been presented for multi-label semantic segmentation of the martian terrain for navigation using variants of a transformer model - SegFormer, which has resulted in a commendable accuracy of 90.86% and mIoU of 83.55% on the AI4Mars dataset.

***Index Terms*—Mars Rover Navigation, Semantic Segmentation, Transformers, SegFormer, AI4Mars.**

I. Introduction

NASA has been at the forefront of extraterrestrial exploration, which has been the primary focus of human science research, for decades. Machine learning has been a major factor in enabling autonomy to increase the duration, dependability, and cost-effectiveness of space missions. In this research, we focus on the topography of Mars for the navigation of rovers.

Every Mars rover's traversability study depends on the ability to identify the terrain type. *AutoNav* [1], which is completely based on long-established machine learning techniques, is used by NASA's Mars rovers to enable autonomous driving. It has been found that different types of terrain have adverse effects on traversability. Both Spirit and Curiosity rovers were halted, put out of action when driving on sandy slumps. It was also reported that Curiosity's wheels suffered the majority of their damage on Sol 450–515 when it drove over sharp, embedded rock surfaces. These illustrations highlight how crucial it is to identify and forecast the least harmful terrain to assure mobility and safer rover adventures[**1**].

Though both terrestrial and satellite photographs could be used to determine the terrain, satellite images simply cannot be taken at scales small enough to distinguish the finer textural features to classify distinct types of rocks. Identification of the topography on these planetary images has proven to be challenging since they are all different from those obtained on the earth in terms of transparency, hue, sharpness, and luminance[**2**].

Classification of martian topography shall be inspected as a subject of semantic segmentation, the process of allocating a category to every pixel in the input image. It would be difficult to pinpoint the boundaries of several categories, such as varying sorts of sand and sky, granular particles, mud and path signs, stone grades, machinery and their reflections[**3**].

An automated terrain categorization built on a semantic net named SPOC [2] (Soil Property and Object Classification) learned from a limited sample of instances supplied by domain analysts, which SPOC then assessed. Raw photos with matching topographical descriptions made up the training set. After receiving an ample amount of training, SPOC categorized each input pixel autonomously[**4**].

Although, a handful of methods present on the approach for image segmentation - using SimCLR[**5**] and DeepLabV3+[**6**], there has not been any approach using models based on Vision Transformers (ViT)[**7**]. In this study, a new method has been presented for autonomous mars rover navigation with Image Segmentation using Transformers.

The paper is organized into six sections. Following this introduction, Section 2 describes the underlying motivation, Section 3 explains significant concepts through a literature review of related work. We put forth the proposed methodology in Section 4 and provide results in Section 5. Ultimately, we've encapsulated our research with a conclusion in Section 6, followed by acknowledgements and a list of references.

II. Motivation

The purpose of this research is to develop and enhance a machine learning system that can recognise landscape features

[1]https://mars.nasa.gov/resources/26677/autonav-drives-perseverance-forward/

[2]https://nasa-jpl.github.io/SPOC/

978-1-6654-8910-2/22/$31.00 © 2022 IEEE

and predict a path from images, thereby enhancing the capabilities of Mars rovers. Despite having an *AutoNav* system, the previous MER Curiosity rover had significant navigational obstacles. This immediately stands out as a challenging task for future Mars rover expeditions. Therefore, developing a strong technique to identify the terrain on Mars is essential for the rovers' sustenance and ease of navigation on the red planet.

III. Related Work

Even though there have been several research studies on the application of images of the Martian surface, there are few studies that use Multi-Label Semantic Image Segmentation of the Martian Surface for rover navigation. Most of the related research works emphasize on categorizing and classifying images of Martian terrain from Mars imagery, with a keen interest in geological significance than navigation. While some research papers discuss unsupervised and semi-supervised training utilizing clustering techniques like K-means clustering[3], only a handful of work is based on semantic segmentation.

Convolutional Neural Network. The classification of both orbital and martian surface images has been the subject of many studies. A deep learning network called MRSCAtt (Mars Rover Spatial and Channel Attention), developed by Anirudh S Chakravarthy, successfully classifies images of the Curiosity rover with an accuracy rate of 81.53 percent using both spatial and channel attention[**2**].

A study by Kiri L. Wagstaff explains how Transfer learning was used to apply the AlexNet CNN (Convolutional Neural Network), which was trained on Earth images, to Mars images[**8**]. Despite substantial variations in picture qualities, imaging environments, and classes of interest, the transfer learning process was successful.SPOC can visually recognise different types of terrain and terrain characteristics on a planetary surface[**4**]. SPOC, which is based on a Deep Convolutional Neural Network (DCNN), uses a machine learning approach in which it learns from a limited number of examples provided by human experts and then efficiently applies the learnt model to a substantial amount of data.

Deep Convolutional Neural Networks have shown to perform well when classifying images[**15**]. Classifiers naturally integrate with low to high level characteristics in an end-to-end multilayer fashion, and their levels can be improved by their depth[**16**]. The NORB, CIFAR-10, ImageNet 2012, and PASCAL VOC datasets show a dramatic performance gain in visual classification tasks respectively due to larger annotated training sets, availability of high compute power like GPU, and improved model strategies[**17**]. The driving approach in a recursively optimized semi-supervised mars terrain classification is pretraining on an ImageNet followed by MoCo V2 using and Adam optimizer, fine tuned on the MSL dataset[**18**].

[3]https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html

Shuang Hu's DeepLabV3+/Efficientnet Hybrid Network-Based Scene Area Judgment for the Mars Unmanned Vehicle System[**6**] is interrelated to our study. Shuang Hu proposes a DeepLabV3+/Efficientnet hybrid network, a derivative of Deep CNN.

AI4Mars, which was crowdsourced, has over 35K EDR images from the Curiosity, Opportunity, and Spirit rovers. It is by far the largest dataset with full image labels for training and evaluating models for classifying terrain of Mars. On the AI4Mars training dataset, a DeepLabv3 model was trained and has a classification accuracy of over 96 percent[**1**].

Edwin Goh presents an implementation of SimCLR on the AI4Mars dataset with a comparably very less number of labels, yielding segmentation accuracy of 91.1% [**14**]. SimCLR is a module of the stochastic data augmentation framework that introduces a deep neural network similar to ResNet.

Support Vector Machines. Changjing Shang describes a special use of support vector machine (SVM) based classifiers for the classification of Mars McMurdo panoramic images[**3**]. The resulting SVM-based classifiers perform better than those using PCA (Principal Component Analysis)-returned features, MLP (Multi Layer Perceptron), and KNN (K-Nearest Neighbour) classifiers. His study specifically covers fuzzy-rough feature selection (FRFS) in conjunction with Support Vector Machines, which has been used and adapted (SVMs)[**9**]. In contrast to transformation-based dimensionality reduction strategies, this approach keeps the underlying semantics of the given feature subset.

Wenjing Wang presents a semi-supervised learning framework in his research work and extends contrastive learning into supervised inter-class and unsupervised similarity-only versions[**10**].

K-means Clustering. Tejas Panambur describes a self-supervised technique that can group sedimentary textures in images captured by the Curiosity rover's Mast camera. The approaches used in this work are largely focused on geological significance[**11**], and contribute to quickly classifying significant geological features. Classification of martian terrains via deep clustering of mastcam images [**12**] and its follow-up, Improved deep clustering of mastcam images using metric learning[**13**], both by Tejas Panambur, emphasize on unsupervised training of images using conventional K-means clustering algorithm with Deep clustering using metric learning as a triplet network.

Vision Transformers. Previous studies have used the AI4Mars dataset for various purposes such as rock detection, geological analysism, object classification and navigation. This is the first study that extends use of a vision transformer model (ViT)[**7**] to semantic segmentation as an alternative to deep convolutional networks on the AI4Mars dataset for navigation. Prior to being trained on the dataset for semantic segmentation, SegFormer[**19**], a hierarchical Transformer encoder, is pretrained on Imagenet-1k and given an MLP decode head.

The fact that the raw images were unclassified suggested the use of unsupervised learning approaches. Our research mainly transitioned from testing unsupervised techniques like

K-means clustering-based segmentation to semantic segmentation using manually created labels for corresponding images.

TABLE I: Final Comparision

Author	Features on the Data	Classifier used	Results
Anirudh S Chakravarthy et al.	Curiosity Rover Images	MRSCAttt (Mars Rover Spatial and Channel Attention)	Test set Accuracy of 81.53 on the MSL Surface Dataset%
Kiri L. Wagstaff et al.	Transfer Learning on Earth to Mars (NASA's Mars images)	AlexNet CNN	Accuracy: HiRISENet-94.5% MSLNet-84.0%
Changjing Shang et al.	Mars McMurdo panoramic images	SVM, FRFS with SVM	Better than PCA, MLP and KNN
Wenjing Wang et al.	Classification and segmentation on MSL Data	Semi Supervised MoCoV2	Final Top-1 (%)Accuracy of 95.86%
Tejas Panambur et al.	Geological Validation on Curiosity Rover Images	Self Supervised Learning technique	Achieved a Precision of 83.6%
Tejas Panambur et al.	Deep Clustering on MSL Mastcam	Conventional K means	Best performance on 50-70 clusters
Brandon Rothrock et al.	SPOC on MRO,HiRISE and MSL Imageries	DCNN	RMSE on Earth-Calibrated and Gaussian Process compared
R. Michael Swan et al.	AI4Mars dataset	DeepLabv3	Overall Accuracy of 96%
Edwin Goh et al.	AI4Mars dataset	SimCLR on deep segmentation network	Segmentation accuracy of 91.1% on training images compared to 81.9% with plain supervised learning
Jayaram M.A et al.	ConvexHull on MSL dataset	ImageNet followed by MoCo V2 using and Adam optimizer	Convex hull-based multi objective genetic programming outperformed the other models
Shuang Hu et al.	Mars Unmanned Vehicle	DeepLabV3+ /Efficientnet hybrid network	Accuracy of 99.84%

IV. Proposed Methodology

A. Dataset

NASA's AI4Mars[4] collection includes 35,000 images from the Planetary Data System3 (PDS), a gray shaded navigation camera NAVCAM from Curiosity, Opportunity and Spirit(MER) rovers, the Mast Camera, also called Mastcam images from Curiosity (MSL). The objective of AI4Mars is to train deep neural networks for sustainable self-driving on Mars, not to study interplanetary science. From MSL data, 18,130 EDR grayscale surface images were taken into consideration. These raw images excluded radiometric correction or camera model linearization. A substantial part consisted of low quality images (typically taken near sundown), comparable ones, with horizon, and those with enormous volumes of rover components.

We require an accurate labeled dataset, using pixel and feature information in an image segment. Since the dataset is broad in count, manually labeling all of them is not practical, which led to refining and annotating the most distinguishable, suitable and clear images - from each category. Only high confidence of labels were generated, hence only parts of the image that belonged to a category with conviction were labeled as sand, soil or bedrock.

B. Exploratory Data Analysis

Unsupervised clustering techniques were used to gain a grasp of the type of classified data we were looking at due to the enormous volume of the dataset. We used a K-means clustering algorithm to examine the texture, structure, and kind of rocky outcrops, grains, and sand dunes[**12**]. We noted general concentrations of stratified rock, flat/grainy ground, and fine, smooth sand. Convex hull method[**13**] was used on the clustered pictures, then edge detection. Sandy areas showed very few edges, but bedrock and soil with a grainy texture had more edges and contours. The output of the images are displayed as in Fig.1.

The dataset consisted of a large number of images. However, the performance of clustering techniques was degraded by the raw images' obscuration of rover segments, suggesting other methods of computation was required. In several of the sections, margins were difficult to discern, including those isolating distinct forms of sand and gravel, sand from the sky, mud from rover trail marks and rock grades as well as machinery and shadows. The images being in various sizes ought to be taken into consideration in order for the model to function.

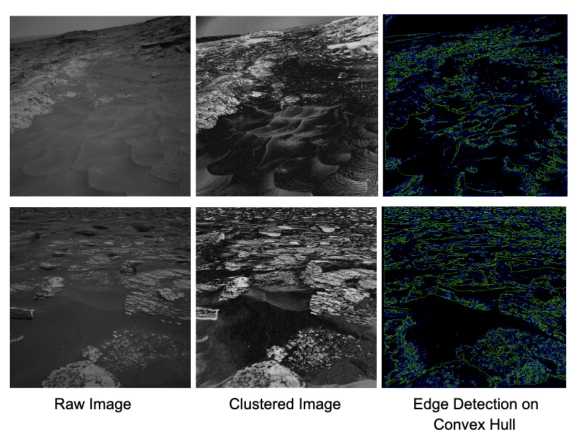

Fig. 1: Unsupervised Methods Image Output

[4]https://data.nasa.gov/Space-Science/AI4MARS-A-Dataset-for-Terrain-Aware-Autonomous-Dri/cykx-2qix

C. Pre-processing

Pre-processing the dataset was a very serious challenge because it presented several complications. First, it required a lot of computational power to examine the roughly 18,000 images. Secondly, most of these raw images featured the rover as the main subject and covered the actual terrain that was the subject of interest. The fact that some of the images were either overexposed or included the sky—again, not what we were aiming for—was another concern. Finally, the images had to be normalized due to its varied features like contrast, brightness and hue.

In order to tackle this problem, the entire dataset was split into 4 batches and a sequence of steps were implemented - Object Detection, Masking, Resizing and Labeling (Fig.4). Each batch of images consisted of nearly equal numbers of images (approximately 4500) and each of the four batches underwent these steps one at a time. A few tools and packages were required for these implementations. 'Detecto,' 'OpenLabelling,' and 'Segments.ai' were among them.

a) Object Detection: An object detection algorithm was implemented using the Python package "Detecto"[5], which is based on the pre-trained model "Faster R-CNN ResNet-50-FPN"[6]. A browser-based tool called "OpenLabelling"[7] was used to build annotation files from samples of images that included the Rover. If the area of the greatest bounding box formed around the detected rover is greater than 1/4th of the image size, the image was rejected from the batch. In the end, the only images retained were those in which the rover either covered less than 1/4 of the image size or was completely absent. Fig.2 represents Rover detected image.

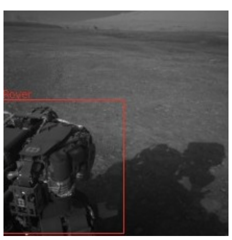

Fig. 2: Rover Detected Image

b) Masking: Following that, each retained image was masked using the associated mask file from the AI4Mars dataset. Unwanted areas, such the sky and the rover components, were masked as a result. Thereby, only the terrain that was of interest was present. Fig.3 represents the processing of masking and the output.

[5]https://detecto.readthedocs.io/en/latest/

[6]https://pytorch.org/vision/main/models/generated/torchvision.models. detection.fasterrcnn_resnet50_fpn.html

[7]https://github.com/Cartucho/OpenLabeling

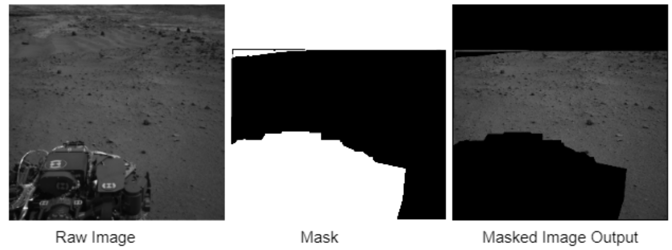

Fig. 3: Masking a Retained Image

c) Resizing: In order to shorten the training processing time and facilitate model fitting afterwards, each masked image was downsized from its original 1024x1024 size to 512x512.

d) Labeling: In order to generate classes with similar proportions, images from each batch were chosen. Priority was given to images with the most terrain covered and the least amount of mask in the selection of images. The chosen images were labeled into 3 classes- Rock, Soil, and Sand, using a browser-based tool called 'Segments.ai'[8]. Segments.ai was utilised in order to retain high labeling precision because it offers a super-pixel based clustered output, which is a very important characteristic of the tool.

The above four steps were implemented in order for all the 4 batches and a final dataset was created[9]. Overall, the original dataset with 18,130 images were reduced to 10,041 masked images. From the masked images, the final dataset was created which consisted of 932 samples with 582 mentions for the class Rock, 524 for the class Soil and 409 for the class Sand - amounting to 10% of the masked dataset.

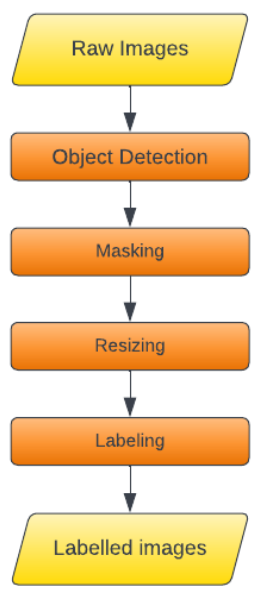

Fig. 4: Data Pre-processing

[8]https://segments.ai/

[9]https://huggingface.co/datasets/Ankhitan/1000Samples

D. Feature Extraction

The features of the images like brightness, contrast, hue and saturation varied a lot from each image due to the characteristics of the image. While some images were captured in bright light, there were others which fell under the shadow. This created a significant difference in the characteristic features of the image pixels. Hence, it was necessary to normalize all the images before feature extraction could be carried out.

Brightness, contrast and saturation were all set to 0.25, while hue was set to 0.1 in order to normalize the image's characteristics. After the images had been normalized, features were extracted for to process into the model.

SegFormer uses a hierarchical transformer encoder for feature extraction. The hierarchical component allows the encoder to generate high-resolution fine features and low-resolution coarse features instead of ViT encoder, which only produces single low-resolution feature maps with fixed resolutions. 1024 x 1024 x 3 images are divided into 4x4 patches, which are then combined using a hierarchical encoder to create a hierarchical feature map F_i with a resolution of $1024/2^{i+1} \ * \ 1024/2^{i+1} \ * \ C_i$, where $i \in \{1, 2, 3, 4\}$. A smaller size is more suitable for a dense prediction task. The top regional layers' features are created using the feature maps of the finer levels.

E. Models

There is a strong relation between classification and semantic segmentation. Many state of-the-art semantic segmentation frameworks are variants of popular architectures for image classification on ImageNet. Variants of Vision Transformers, Dilated Convolutions and Deep Convolutions have seen a recent surge of interest for semantic segmentation.

SegFormer[10] is a powerful semantic segmentation framework which comprises hierarchically structured Transformer encoder and Multilayer perceptron (MLP) decoders, pre-trained on ImageNet-1k. In this study, we have implemented three SegFormer variants and obtained the variation from the outputs on our final created dataset, interpreting results. The training set contained 746 photos, and the testing set contained 186 images after the dataset was split in half in the ratio 80:20 for the train and test sets, respectively. Using the Pytorch framework, we implement the Segformer - B0, B1 and B2 [11]. With a batch size of 8 and a learning rate of 0.0006, we train our network over 15 epochs.

To implement deep learning semantic segmentation of images, SegFormer for image segmentation integrates each MLP decode head on top. On the imagenet-1k database, the transformer encoder is initially trained in advance to categorize pictures. All-MLP decode head is then used instead of the categorisation head. The prototype is then substantially fine-tuned on the sample for the classifier pictures and accompanying demarcation detailed charts are prepared using feature extractor. The Segformer architecture, unlike Vision Transformers, does not rely on positional encodings and thus improves inference on images of varying resolutions. The most vital pre processing stage is arbitrarily padding and trimming pictures and demarcation waypoints to the equal dimensioned in 512x512 and later normalising them by changing the brightness, contrast, saturation and hue of the images accordingly. Each variant differs in the parameters, hidden sizes, depths of the transformer encoder backbone introduced in SegFormer.

[10]https://github.com/NVlabs/SegFormer

[11]https://huggingface.co/docs/transformers/model$_{doc/segformer}$

V. Results

The training process took place on a system running GPU version of NVIDIA Tesla P100-PCIE [12] using GoogleColab [13] with Pytorch[14] framework, and it took nearly 7 hours to complete 1400 optimization steps for each model version. All three of the model versions were successful in achieving a commendable accuracy, with the B2 model performing the best with an overall accuracy of 90.86% and a Mean IoU of 83.55%.

Loss, accuracy, and IoU (Intersection over Union) are one the most effective metrics, and learning rates are fine-tuned to infer these. The overlap between expected and real masks is measured by IoU [20] to assess the performance of our network. It indicates the amount of overlap between predicted and ground truth images. The final results have been computed and tabulated in the Tables 1 and 2. The results are also plotted as in the Fig.5,6,7,8. The final segmented image output samples are displayed in Fig.9. The entire research has been open-sourced and documented in a Github repository[15].

TABLE II: Final Results

Model	Mean IoU	Mean Accuracy	Overall Accuracy
SegFormer - B0	81.15%	88.61%	89.92%
SegFormer - B1	83.02%	90.33%	90.66%
SegFormer - B2	83.55%	90.75%	90.86%

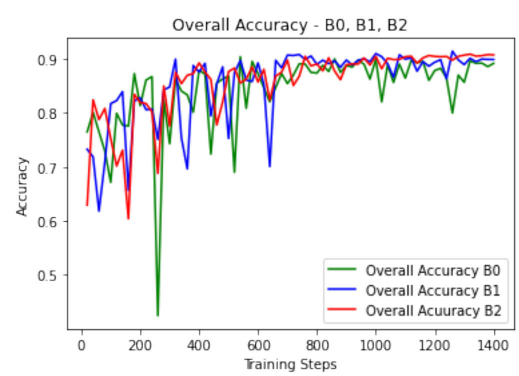

Fig. 5: Overall Accuracy

[12]https://www.nvidia.com/en-in/data-center/tesla-p100/

[13]https://research.google.com/colaboratory/

[14]https://pytorch.org/

[15]https://github.com/Ankithac45/Semantic-Segmentation-on-Martian-Terrain

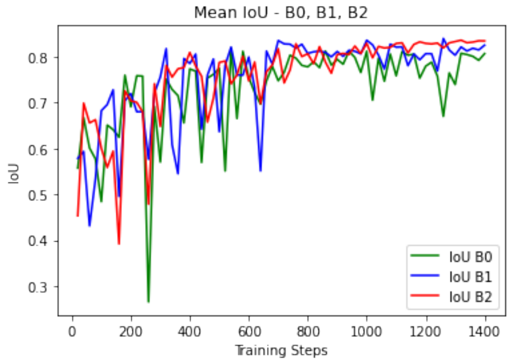

Fig. 6: Mean IoU

TABLE III: Per Class Results of SegFormer - B2

Class	Per Class IoU	Per Class Accuracy
Sand	85.67%	90.65%
Soil	79.35%	89.06%
Rock	85.64%	92.54%

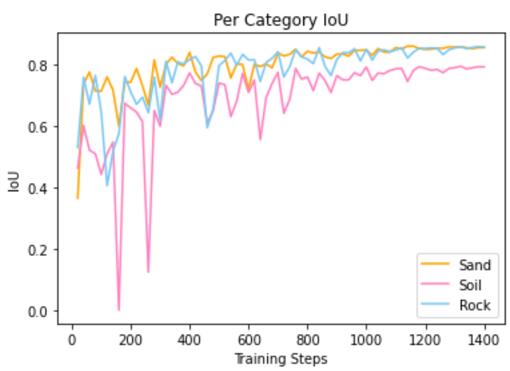

Fig. 7: SegFormer-B2 Per Category IoU

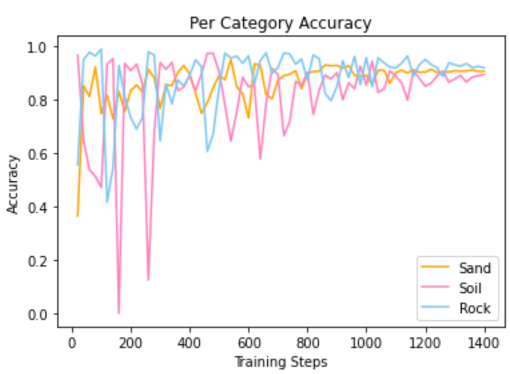

Fig. 8: SegFormer-B2 Per Category Accuracy

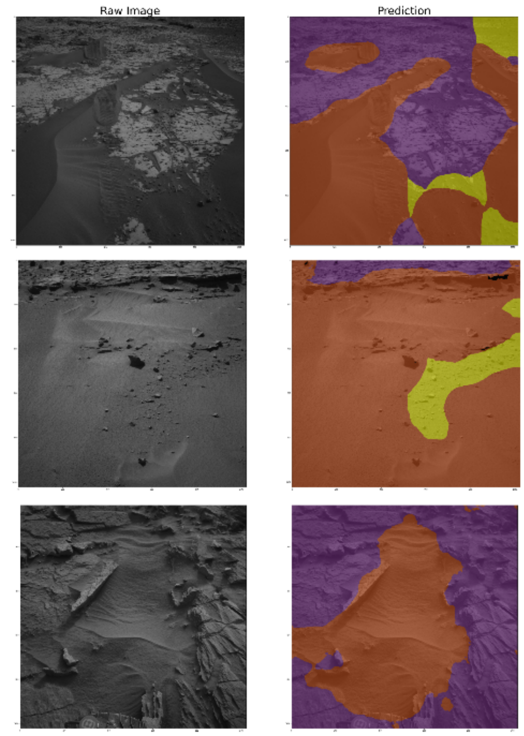

Fig. 9: Ground Truth vs Prediction

VI. Conclusion

Relevant research to date has concentrated on several approaches and strategies for creating intelligent autonomous navigation for Mars rover missions. This study has two major results: (1) A novel, alternative approach is presented on Multi-Label Image Segmentation for the martian terrain using Transformers, instead of Deep Convolutional Neural Networks; (2) The comparison of the performance the three SegFormer model versions implemented. The raw images were extensively pre-processed using a variety of tools and techniques in order to extract the features and fit the models. Further, the three models- SegFormer B0, SegFormer-B1 and SegForme-B2 were trained, and the results were computed. We conclude that the segmentation process was successfully accomplished by all the models, with the best overall accuracy of 90.86% by SegFormer-B2. As a result, we believe the model has the ability to address the issue of image semantic segmentation with regard to the martian terrain and can contribute in related research. Overall, this research emphasizes the significance of autonomous navigation for rovers and presents a new method to implement for the challenging martian terrain.

Acknowledgment

We thank Dr. Mamatha H.R for her support and guidance in structuring this manuscript. We thank IEEE CS PESU and Dr. Venugopal N, Faculty Advisor, IEEE CS PESU for their support in completing this research. We acknowledge Shreyas A Katti, Rishabh Oruganti and Parimala S for insightful discussion and guidance throughout our research.

REFERENCES

[1] Swan, R. Michael, Deegan Atha, Henry A. Leopold, Matthew Gildner, Stephanie Oij, Cindy Chiu, and Masahiro Ono. "AI4MARS: A dataset for terrain-aware autonomous driving on Mars." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1982-1991. 2021.

[2] Chakravarthy, Anirudh S., Roshan Roy, and Praveen Ravirathinam. "MRSCAtt: A spatio-channel attention-guided network for mars rover image classification." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1961-1970. 2021.

[3] Shang, Changjing, and Dave Barnes. "Combining support vector machines and information gain ranking for classification of mars mcmurdo panorama images." In 2010 IEEE International Conference on Image Processing, pp. 1061-1064. IEEE, 2010.

[4] Rothrock, Brandon, Ryan Kennedy, Chris Cunningham, Jeremie Papon, Matthew Heverly, and Masahiro Ono. "Spoc: Deep learning-based terrain classification for mars rover missions." AIAA SPACE 2016 (2016): 5539.

[5] Chen, Ting, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. "A simple framework for contrastive learning of visual representations." In International conference on machine learning, pp. 1597-1607. PMLR, 2020.

[6] Hu, Shuang, Jin Liu, and Zhiwei Kang. "DeepLabV3+/Efficientnet Hybrid Network-Based Scene Area Judgment for the Mars Unmanned Vehicle System." Sensors 21, no. 23 (2021): 8136.

[7] Dosovitskiy, Alexey, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani et al. "An image is worth 16x16 words: Transformers for image recognition at scale." arXiv preprint arXiv:2010.11929 (2020).

[8] Wagstaff, Kiri, You Lu, Alice Stanboli, Kevin Grimes, Thamme Gowda, and Jordan Padams. "Deep Mars: CNN classification of mars imagery for the PDS imaging atlas." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, no. 1. 2018.

[9] Shang, Changjing, and Dave Barnes. "Fuzzy-rough feature selection aided support vector machines for mars image classification." Computer Vision and Image Understanding 117, no. 3 (2013): 202-213.

[10] Wang, Wenjing, Lilang Lin, Zejia Fan, and Jiaying Liu. "Semi-Supervised Learning for Mars Imagery Classification." In 2021 IEEE International Conference on Image Processing (ICIP), pp. 499-503. IEEE, 2021.

[11] Panambur, Tejas, Deep Chakraborty, Melissa Meyer, Ralph Milliken, Erik Learned-Miller, and Mario Parente. "Self-Supervised Learning to Guide Scientifically Relevant Categorization of Martian Terrain Images." In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1322-1332. 2022.

[12] Parente, Mario, and Tejas Panambur. "Classification of martian terrains via deep clustering of mastcam images." In IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, pp. 1054-1057. IEEE, 2020.

[13] Panambur, Tejas, and Mario Parente. "Improved deep clustering of mastcam images using metric learning." In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 2859-2862. IEEE, 2021.

[14] Goh, Edwin, Jingdao Chen, and Brian Wilson. "Mars Terrain Segmentation with Less Labels." arXiv preprint arXiv:2202.00791 (2022).

[15] Zeiler, Matthew D., and Rob Fergus. "Visualizing and understanding convolutional networks." In European conference on computer vision, pp. 818-833. Springer, Cham, 2014.

[16] He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.

[17] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587. 2014.

[18] Jayaram, M. A., and Hasan Fleyeh. "Convex hulls in image processing: a scoping review." American Journal of Intelligent Systems 6, no. 2 (2016): 48-58.

[19] Xie, Enze, Wenhai Wang, Zhiding Yu, Anima Anandkumar, Jose M. Alvarez, and Ping Luo. "SegFormer: Simple and efficient design for semantic segmentation with transformers." Advances in Neural Information Processing Systems 34 (2021): 12077-12090.

[20] Giorgiani do Nascimento, Renato, and Felipe Viana. "Satellite image classification and segmentation with transfer learning." In AIAA Scitech 2020 Forum, p. 1864. 2020.