Revolutionizing Martian Terrain Mapping: Precision Segmentation Through Deep Learning

Khadija Naveed[1], Muhammad Umair[2], Mishail Saif[2]
[1]Department of Computer Engineering, National University of Computer and Emerging Sciences, Lahore, Pakistan
[2]Department of Computer Science, FoITCS, University of Central Punjab, Lahore, Pakistan
***Correspondence**: khadija.naveed@nu.edu.pk
Citation | Naveed. K, Umair. M, Saif. M, "Revolutionizing Martian Terrain Mapping: Precision Segmentation Through Deep Learning", Special Issue pp 294-306, May 2026
Received | March 18, 2026 **Revised |** April 26, 2026 **Accepted |** May 02, 2026 **Published |** May 07, 2026.

Planetary mapping, autonomous rover navigation, geological analysis, and mission planning are all based on the Martian terrain segmentation. However, differentiating various surface characteristics is not easy because of the variations in illumination, imbalance in classes, and similarity of the appearance of different terrain types. To overcome these obstacles, this paper suggests a more accurate segmentation model, using an attention-enriched ResNet-50 framework, combined with two explainability methods, i.e., Grad-CAM and Integrated Gradients, to enhance the performance and explainability. The Mars_terrain_classification dataset, which consists of 6,153 high-resolution images of various terrain classes (craters, dunes, impact ejecta, slope streaks) was used to train and test the model. The experimental findings indicate that the proposed framework had an overall accuracy of 92.61, a mean Intersection over Union (IoU) of 84.7, and a total F1-score of 88.48, which is better than baseline models like U-Net and DeepLabV3+ given the same training conditions. In particular, the proposed model has led to a 6.41 percentage points F1-score improvement compared to U-Net and 2.59 percentage points improvement compared to DeepLabV3+, and an 8.7 and 3.2 percentage points IoU improvement, respectively. Class wise analysis revealed that Bright Dune (95.89% F1-score), Crater (91.28%), and other terrain (94.23%) have high scores, whereas Impact Ejecta (76.92%), and Slope Streak (80.62%) have lower scores, suggesting the still existing challenges in minority and visually ambiguous classes. The explainability results showed that Grad-CAM and Integrated Gradients always show geologically meaningful areas, which adds to the transparency and reliability of predictions. The results indicate that the presented framework can offer a correct and sensible solution to the high-precision mapping of the Martian terrain.

Keywords: Segmentation, Grad-CAM, Attention-Enhanced ResNet-50, Explainable Artificial Intelligence (XAI), Deep Learning for Planetary Mapping

Introduction:

The studies of the Martian surface are essential in studying the geological development of the planet, its surface features, and its possible life-supporting conditions. Proper identification of the features of the Martian surfaces is vital in the autonomous navigation of the rovers, geological study, as well as in planning the mission. Nevertheless, the terrain on the Martian surface can be described as highly tortuous, with a changing light intensity, and similar geomorphological formations, and therefore, it is not easy to map this terrain accurately. Conventional mapping methods, which rely on manual interpretation and the classical image processing methodology, are likely to be laborious, time-consuming and may produce inaccurate results due to the heterogeneous nature of such an environment.

Deep learning has become a potent instrument for semantic segmentation of planetary surfaces in recent years. The CNN-based architectures have proven to be highly effective in the extraction of hierarchical features and enhancement of segmentation. As an example, encoder-decoder-like models like MarsSeg were suggested to improve the multi-level feature representation to segment the Martian surface, which obtained better accuracy in comparison with the traditional models [1]. Nonetheless, even with such developments, CNN-based methods tend to have difficulties in modeling long-range interactions and multi-scale contextual interactions, required to differentiate geometrically complex and low-texture terrain features [2].

To overcome these constraints, more and more recent studies have aimed to consider transformer-based and hybrid architectures. Models with depth enhancement, including DepthFormer, have shown better segmentation results with depth geometric information to enhance their ability to deal with complex terrain structure [3]. On the same note, multimodal models that combine RGB and depth data have been developed to provide better understanding of the terrain and make the rover-based systems safer in navigation [4]. Newer methods such as dual stream attention networks and Swin Transformer fusion architectures have demonstrated high performance in capturing local details and global context and, therefore, much better segmentation of complex Martian terrains [5].

More recent works have delved further into deep learning-based segmentation to enhance rover-based terrain perception, and geological feature identification. As an example, methods based on convolutional neural networks have been effectively used to automatically detect and segment features on the Martian surface, like barchan dunes, illustrating the ability of deep learning to extract complex geomorphological features on planetary imagery [6]. Moreover, recent benchmarking initiatives, including Mars-Bench, emphasize the increasing importance of standardized evaluation models to evaluate the generalization of models on a variety of datasets and tasks on Mars [7]. Overall, regardless of these advances, there are still difficulties in the high-precision segmentation of limited labeled data, imbalance between classes and variability of domains across various environments on Mars. The other important challenge is that deep learning models are not easily interpretable. Although modern architectures perform well, their decision-making processes tend to be opaque and are therefore hardly applicable to scientific applications where transparency and trust are paramount. Grad-CAM and Integrated Gradients are explainable Artificial Intelligence (XAI) methods which have been utilized more often to gain insight into model predictions, but their joint usage in Martian terrain segmentation is underutilized. Therefore, a framework with not only high segmentation accuracy, but also interpretable and computationally efficient solutions that can be used in planetary exploration surroundings is required. Here, this paper suggests an attention-refined ResNet-50-based segmentation design that is combined with two explainability systems. The proposed solution will help to better feature discrimination, increase segmentation accuracy of different terrain classes and give a clear understanding of the model decision-making, which will lead to the high-quality and reliable Martian terrain mapping.

The purpose of this study is to create a precise and interpretable framework of Martian landscape segmentation. The specific objectives are:

To construct an attention-enhanced ResNet-50 model, to effectively segment Martian terrain classes.

To enhance the discrimination between visually similar terrain features.

To assess the performance of the model in terms of accuracy, F1-score, and IoU.

To compare the proposed approach with base models (U-Net and DeepLabV3+).

To use Grad-CAM and Integrated Gradients to enhance interpretability.

The main contributions of this study are as follows:

A Martian terrain segmentation ResNet-50 framework with attention to enhance the segmentation.

Grad-CAM + Integrated Gradients: Adding Grad-CAM and Integrated Gradients to models to enhance interpretability.

Evidence of the improvement of performance compared to the baseline models.

A trade-off between the accuracy of segmentation and explainability to planetary applications.

Section II considers the related work performed on Martian terrain segmentation based on deep learning. The methodology is described in Section III. Section IV gives experimental findings and comparative analysis, and Section V describes conclusions to the study.

Literature Review:

[8] Introduced a novel multi-view Martian Terrain Segmentation framework (MTSNet) to aid in tackling the problem of unstructured and highly variable Martian terrains. It consists of the Martian Terrain text-Guided Segment Anything Model (MTG-SAM) equipped with a Terrain Context Attention Adapter Module (TCAM) and Local Terrain Feature Enhancement Network (LTEN). Further exploration is required to assess the computational efficiency of the model, although such discussions did occur, particularly in the context of Mars rovers as resource-constrained environments. The adaptability of text prompt-based segmentation approaches in ambiguous text or incomplete text descriptions is not fully validated as well. [9] Proposed a semi-supervised learning framework to mitigate the issues posed in the classification and segmentation of Martian images (i.e., limited number of images for annotation, train-test gap, and bad quality of images.) The scalability of the method and application to other Martian vision applications, such as object detection and tracking, would remain to be established with further research. [10] developed a holistic framework for improving Mars terrain segmentation through style transfer, enhancement, and semi-supervised pseudo-labeling techniques for thing-and-stuff-class perception. The researchers introduced a style transfer neural method, SA-CCPL, to solve the adaptive transformation of image terrains for better model generalization without additional annotations. The computational efficiency of the proposed methods was not probed in terms of resource-constrained environments such as Mars rovers. [11] Developed an automated multi-class crater segmentation framework that incorporates Mars orbital images. The study focused on counting craters and estimating surface ages. It presented a pipeline that comprised U2-Net and U-NetFormer models for both binary and multi-class semantic segmentation with template matching for crater counting. The count of craters depends on the template matching that, in the case of close or irregularly shaped craters, is very likely to provide erroneous counts. Also, the behavior of the framework has not yet been evaluated with regard to different Martian terrains and extremely high-resolution datasets like HiRISE. An automated methodology for segmenting Martian dust storms, using a combination of PCA and multilayer perceptron classifier, was proposed by [12]. It was also static for all seasonal and regional

varieties and terrain features in Mars because the parameters used by the model in relation to the existing MLP classifier were not tuned. There was no mention of how computationally expensive it would be to train and run the classifier either in real or large-scale applications. Liu et al. developed MRISNet as a deep-learning-based framework for instance segmentation of Martian rocks, which aims to address problems of motion blur and defocus blur caused by the rover's navigation cameras. The framework employs image decommissioning via DeblurGAN, feature pyramid, and attention mechanism along with the superpixel segmentation algorithm SLIC for enriched pixel-level classification. Furthermore, evaluation based on a single dataset raises issues regarding the generalizability of MRISNet with varied terrain found on Mars. Future efforts should focus on investigating efficiency optimization and validation with various datasets [2]. [13] Built a unique kind of convolutional neural network, or Crater U-Net, for automatic detection of Martian craters from THEMIS thermal infrared imagery. This work addresses the memory-hogging, manual past, and labor-intensive tradition of crater counting by employing U-Net-like segmentation mechanisms optimized for planetary data. The computational feasibility of deploying Crater U-Net on resource-constrained planetary missions was not considered. To address challenges such as the diverse appearance of Martian rocks and the deficiency of CNNs in modeling global dependencies, [14] introduced RockFormer, a U-shaped Transformer framework for Martian rock segmentation. The action adopts a Vision Transformer (ViT), improves multiscale feature extraction, and uses a Feature Refining Module (FRM) for enhancing interscale feature connections. The study, however, does not study the extent of its adaptability to very new terrains and extreme conditions, thus emphasizing further research on improving efficiency and generalizability.

[15] Introduced an incremental domain adaptive segmentation technique termed Cov-DA specifically aimed at addressing covariate shifts in remote sensing imagery (RSI). It has mostly been evaluated on datasets, which puts into doubt its generalization ability to unseen Martian terrains or extreme lighting conditions. CloverNet, proposed by [16], is a real-time CNN optimized for semantic segmentation for use on resource-limited planetary exploration rovers. The curated dataset currently offers only a small number of classes, which restricts its application to broader planetary contexts. [17] Produced a complete verticalized convolutional neural network framework that detects and segments multiple Martian geological features-such as impact craters, dunes, transverse aeolian ridges (TARs), and volcanic rootless cones (VRCs)-from high-resolution images captured by Mars Reconnaissance Orbiter (MRO). The tests were carried out mostly on the particular datasets, which raises the generalizability issues on other planetary environments or datasets with low resolution. The research conducted by [18] has repeatedly addressed semantic segmentation in the classification of Mars terrains with a goal of improving autonomous navigation and pathfinding for Mars rovers. The research limited its focus on certain datasets, thus constraining its breadth and scope concerning generalization to other Martian terrains. The method of instance segmentation, such as Mask R-CNN, introduced by [6], is used to detect and demarcate barchan dunes on Mars and Earth, which was quite a laborious task of manual mapping and often prone to errors through conventional methods. The fact that a single dataset is being used for Mars raises questions about how generalizable it is to various Martian terrains and other planetary bodies. Liu et al. presented a new framework called UDAFormer, a transformer-based unsupervised domain adaptation (UDA) approach for Martian terrain segmentation, to deal with such challenges resulting from domain shifts in datasets and sparse labeled high-quality data. However, as much as UDAFormer very well suits domain adaptation and segmentation accuracy, the reliance on such costly computation transformer-based architectures makes it less feasible for deployment on resource-limited systems such as Mars rovers [19].

Recent works on Martian terrain segmentation take the approach of deep learning: multi-view, semi-supervised, and transformer-based techniques. In spite of encouraging performance, there still exist some concerns about the efficiency of the computations, scaling, and generalization in constrained rover environments.

Materials and Methods:

To identify and classify surface features on Mars like craters, valleys, and plains, new data-acquisition methods are required to generate digital terrain models. The currently used technique involves satellite imagery interpretation and traditional recognition methods that is labor intensive, time-consuming, and prone to human error.

Dataset Collection and Preparation:

High-resolution images of Martian terrains were downloaded from the NASA Planetary Data System. These datasets comprise satellite imagery captured during Mars exploration missions such as the Mars Reconnaissance Orbiter and Viking Orbiter missions. The images represent a variety of Martian geological features, including craters, ridges, dunes, and plains under varying lighting and environmental conditions. The dataset was organized in a hierarchical structure where each subdirectory corresponds to a certain type of terrain (e.g., craters, plains, etc.).

Data Preprocessing:

Resize and Cropping:

All images were resized to 224×224 pixels. The resizing operation allowed all images to be compatible with the architecture of the ResNet-50 network. Cropping cut out unclear boundaries or irrelevant regions that caused the model to be distracted during training. The dataset was randomly split into 80% training data and 20% testing data.

Data Augmentation:

Data augmentation enhances the ability of the model to generalize by artificially raising the volume of the data. Its importance increased with Martian terrain mapping, where data sizes may be small since very few labeled images could be made available. Hence, data augmentation techniques, including random rotation and horizontal flipping, were applied.

Normalization:

Normalization was one of the most important preprocessing steps to rescale the pixel values consistently to aid convergence during training. Normalization was done using the mean and standard deviation values from the ImageNet Dataset (i.e., Mean: (0.485, 0.456, 0.406), Standard Deviation: (0.229, 0.224, 0.225)). This was due to the fact that ResNet-50, the deep learning model used in this study, was pre-trained on ImageNet, necessitating alignment of statistics of the input image with the weights pre-trained. Normalization ensured that the pixel intensities were scaled to similar ranges and minimized the influence of great brightness or color variations of images. It did so to stabilize the learning process and help the model converge more quickly, hence mitigating issues such as exploding gradients.

Class Imbalance Handling:

There is an imbalance in the classes of the Martian terrain dataset. An augmentation method of data, such as random rotations and horizontal flips, was implemented to enhance the diversity of samples. Moreover, sampling was used to balance classes during the training to make sure that the minority classes were adequately represented in every batch.

Model Architecture:

Backbone Selection:

The heart of the model architecture was based on ResNet-50, a very famous CNN architecture for its number of layers and efficiency in hierarchical feature extraction from images. ResNet-50 introduces the concept of residual connections for addressing the vanishing gradient problem in deep networks. These residual connections allowed gradients to flow more smoothly in the direction of backpropagation through the network, permitting deeper architectures without loss in performance. ResNet-50 consists of 50 layers composed of convolutional layers, batch normalization, ReLU activation, and skip connections.

Attention-Enhanced Feature Learning:

To enhance the discriminative features of the ResNet-50 backbone, an attention mechanism was adopted for feature representation. The attention module enables the network to focus on informative regions of the Martian terrain of the Martian terrain as the background factors are suppressed. A channel spatial attention model in this study was incorporated in the middle layers of the ResNet-50 as a feature extraction model. Particularly, the attention blocks were placed between the residual stages to trim feature maps prior to forwarding to sub-layers. The channel attention module is trained to value the significance of every channel of features by implementing global average pooling followed by fully connected layers and a sigmoid activation. This mechanism enables the model to highlight the feature channels, which reflect significant terrain features like craters, dunes, and slope streaks. Furthermore, there is also a spatial attention mechanism that helps to learn the spatial association in feature maps. Another convolutional layer with sigmoid activation generates a spatial attention map, which identifies the areas on the terrain image that have information. Through the integration of channel and spatial attention, the model is also able to capture features that are important along with their location in the image.

Custom Segmentation Head:

For pixel-wise segmentation, a custom segmentation head was designed and attached to the main backbone. This head receives the high-dimensional feature maps provided by the backbone and outputs the final segmentation masks having the same resolution as the input images. The segmentation head consists of convolutional layers followed by upsampling operations to restore spatial resolution. The architecture includes a 1x1 convolution layer to reduce channel dimensionality of the backbone feature maps, a 3x3 convolution layer for spatial feature refinement, bilinear upsampling to recover spatial resolution, and a final softmax activation layer to produce class probabilities for each terrain category. This design enables the network to combine high-level semantic features extracted by the ResNet-50 backbone with spatial localization required for terrain identification. The overall process of the proposed attention-based ResNet-50 Martian terrain segmentation framework is demonstrated in figure 1.

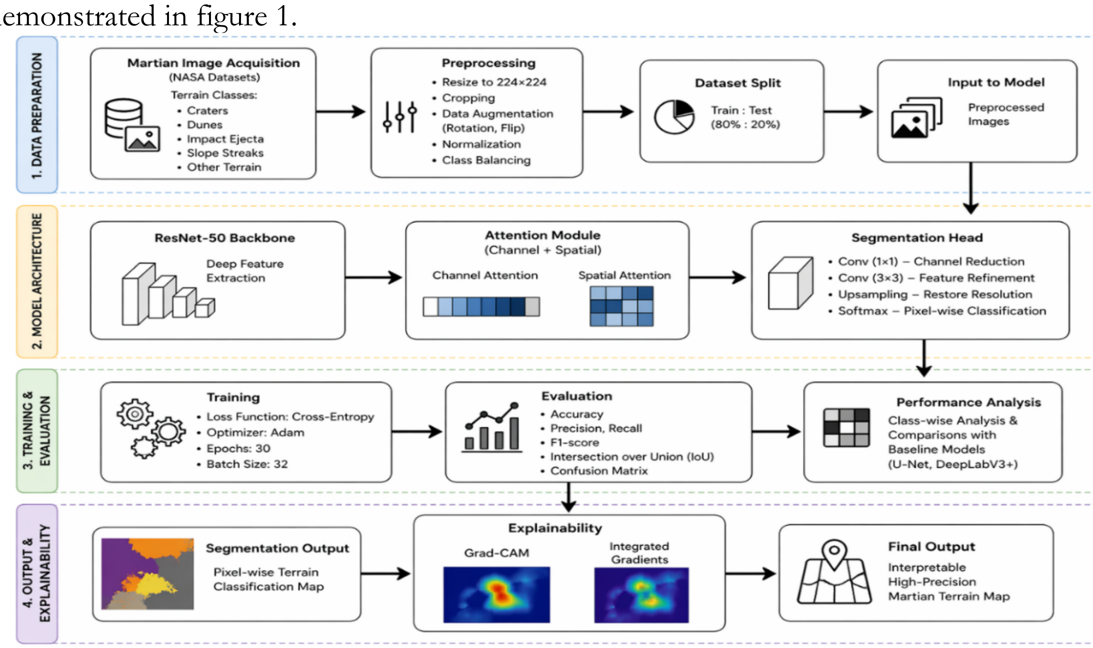

Figure 1. Workflow of the proposed attention-enhanced Martian terrain segmentation framework

Problem Formulation:

The Martian terrain segmentation can be formulated as a supervised learning problem where a deep learning neural network learns a mapping from input images to terrain class predictions. Let X = $\{x_i\}_{i=1}^{N}$ represent the set of input Martian terrain images and Y = $\{y_i\}_{i=1}^{N}$ denote the corresponding ground-truth terrain labels. The objective of the model is to learn a function f_θ parametrized by θ such that

$$y_i = f_\theta\ (x_i)$$

Where $x_i \in R^{H \times W \times 3}$ represents an RGB image and y_i corresponds to the terrain category label. The network parameters θ were optimized during training to minimize the classification loss between predicted labels and ground truth annotations.

Explainability Methods:

Explainability methods were among the core components of contemporary deep learning systems, especially in applications with critical trust, transparency, and validation of predictions.

Grad-CAM (Gradient-weighted Class Activation Mapping):

Grad-CAM is a method to produce attention maps of the most salient regions of an input image for a prediction. It offers qualitative explanations of the most important terrain features for classification. It involves calculating gradients of the target class score with respect to the feature maps of a convolutional layer, and then pooling them to generate a heatmap specific to the target class, which is upsampled to the original image size. Grad-CAM enhances interpretability by allowing the verification that the model is attending to relevant features such as crater rims and dune boundaries. It can also be used to check for anomalies, as attention on irrelevant areas could be the result of spurious correlations. Integrated Gradients is a useful supplement to Grad-CAM, with pixel-wise attribution based on integrated gradients, which satisfies the completeness property. It finds detailed feature attribution, showing important regions (such as the crater center) and less important regions. The combined analysis reveals that Grad-CAM focuses on coarse-scale spatial patterns, while Integrated Gradients offers fine-scale interpretation. In some instances, blurrier or more scattered heatmaps highlight model failure under low contrast and complex illumination, providing a basis to improve model performance.

Integrated Gradients:

Integrated gradients attribute the importance of individual pixels, or input features of the model's predictions. This gives an avenue to more quantitative interpretation of the model behaviors compared to what Grad-CAM provides with the qualitative heatmap. Integrated Gradients satisfies the completeness property; it states that the total attribution equals the output of the actual input minus that of the baseline. In Martian terrain segmentation, Integrated Gradients provides the importance of certain pixels or regions in identifying the terrain types. For example, it may assign a high importance value to the central pixels of a crater but a low importance value to more uniform background areas. This high degree of detail will be used to check if the model is focusing on scientifically relevant
Features.

Results and Discussion:

The dataset provided diverse training samples derived from various Martian terrains to thoroughly test the suggested deep learning segmentation model. The experiment was conducted in several training sessions to test the stability of performance. The following sections discuss the results in detail.

Quantitative Performance Assessment:

As summarized in table 1, the segmentation model performed exceedingly well in various terrain classes, yielding an overall accuracy of 92.61%. Bright dunes achieved the highest F1-score of 95.89%, which infers remarkable class performance in the detection of this terrain. The detection of craters also proved to be successful with an F1 score of 91.28%, thus establishing the reliability of the model in recognizing impact structures. With an F1 score of 76.92%, the impact ejecta class yielded a lower recall, indicating some misclassification owing to the similarity of this terrain with others adjacent to it. Strong overall model performance was reflected in the weighted macro-average value of 87.51% for the F1 score.

Table 1. Class-wise Performance Metrics

Terrain Type	Precision (%)	Recall (%)	F1-score (%)	Support (Samples)
Bright Dune	99.06	92.92	95.89	113
Crater	95.19	87.68	91.28	203
Dark Dune	97.44	77.55	86.36	49
Impact Ejecta	100.00	62.50	76.92	8
Other	91.04	97.66	94.23	728
Slope Streak	89.66	73.24	80.62	71
Spider	88.89	80.00	84.21	10
Swiss Cheese	93.48	87.76	90.53	49

To understand the overall performance under class imbalances, the macro-averaged scores were considered along with class-specific scores. Reduced recall for impact ejecta and spider formations is due to the lack of training samples and visual ambiguity in the classes. This highlights the need for future research to balance and augment data to enhance the detection of minority classes.

IoU Performance:

The intersection over union measured directly quantifies how close the predicted segmentation mask was to the corresponding ground truth annotations. The model recorded an average IoU of 84.7%, which was significantly better than the baseline performances of U-Net (76%) and DeepLabV3+ (81.5%).

Confusion Matrix Analysis:

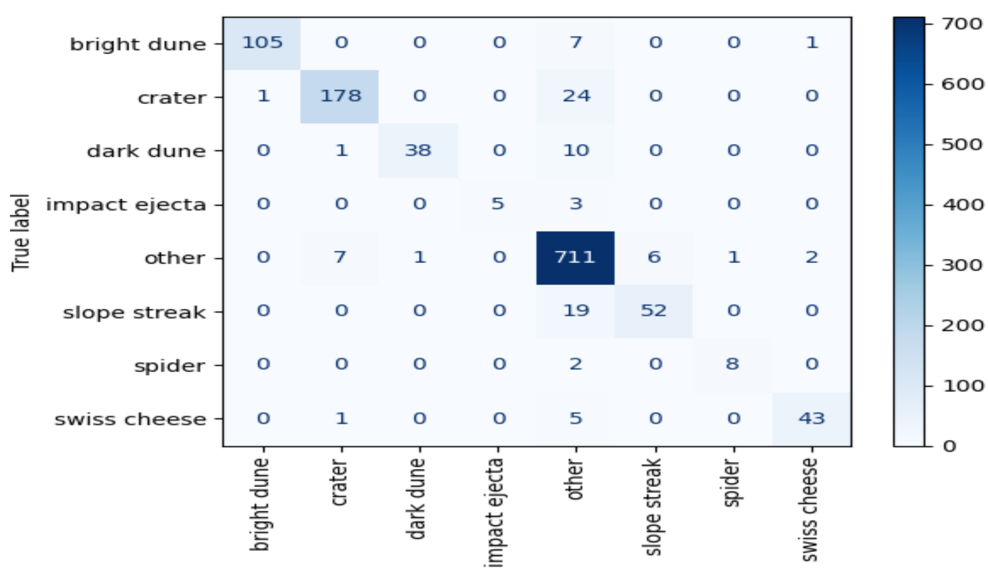

Figure 2. Confusion matrix illustrating classification performance across Martian terrain categories

A confusion matrix was obtained to study the misclassification patterns in order to further assess the model performance on terrain categories. As illustrated in figure 2, most of the misclassifications are between the terrain structures that appear visually similar i.e., dark dunes and bright dunes and between the edges of a crater and the terrain around it. In spite of the fact that there is an imbalance of classes in the dataset, the given model proves to have a high level of classification among most different terrains. The confusion matrix also points out that the classes as minority as impact ejecta and spider formations have somewhat smaller recall values because they have few training samples.

Computational Efficiency Analysis:

Along with segmentation accuracy, computational efficiency is another factor of interest with the usage of deep learning models in planetary exploration systems like Mars rovers. The backbone that was used in the proposed architecture is the ResNet-50, which offers the trade-off between the power to extract features and the cost of computation. As summarized in table 2, the suggested architecture has about 25 million learnable parameters, which match typical deep learning networks that are trained as image classifiers and segmentation systems. It is estimated that the model has a computational complexity of around 4 GFLOPs in one forward pass, but with an input image size of 224 x 224 pixels. The inference experiments have revealed that the model could process terrain images under near real-time circumstances when it operated on systems that were powered by a GPU. The mean inference time per image is around tens of milliseconds and this shows that the architecture is computationally viable when used in automated terrain analysis pipelines.

Table 2. Computational Complexity of the Proposed Model

Metric	Value
Backbone	ResNet-50
Input Resolution	224 × 224
Parameters	~25M
Computational Cost	~4 GFLOPs
Average Inference Time	~25–30 ms/image

Even though on-board deployment can be optimized further on resource limited rover hardware, the current architecture makes future edge-deployment planning, such as model pruning, quantization or lightweight backbone replacement, a good starting point.

Grad-CAM Heatmaps:

To visualize the most influential regions contributing to model prediction, the Grad-CAM implementation was applied to Martian terrain images. As shown in figure 3, the technique generated heatmaps for the relevant regions in the images that indicate places where the model focused on differentiating terrain types. The Grad-CAM results demonstrated that the model effectively learns spatial features for Martian terrain classification. While the method provides valuable insights into the model's decision-making process, some misclassifications indicate the need for further refinement in data augmentation and feature extraction strategies. These insights will be crucial in enhancing model performance for planetary exploration and automated terrain analysis. The Grad-CAM heatmaps were further compared with integrated gradients attribution maps to evaluate the consistency of highlighted terrain regions.

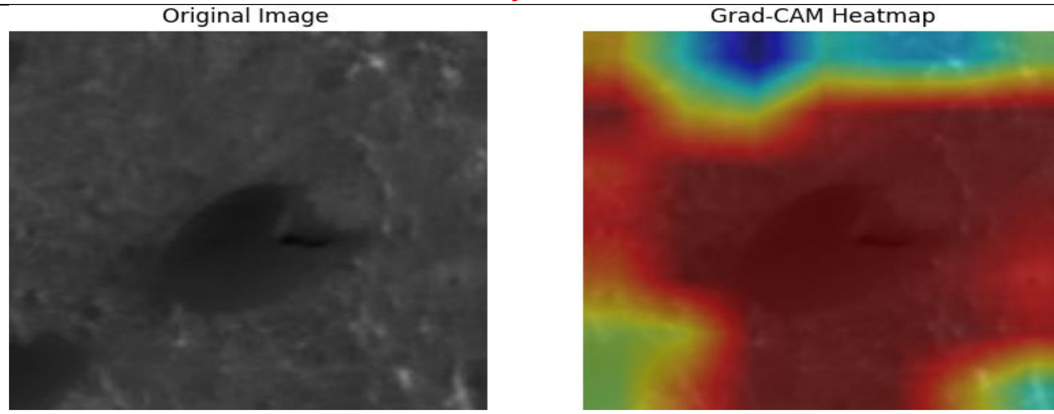

Figure 3. Original Image and Grad-CAM Heatmap

Integrated Gradients Heatmap:

The Integrated Gradients method was applied to analyze feature importance in the classification of Martian terrain. The resulting attribution maps shed light on the contribution of each pixel to the model's decision and provide information on how the neural network deals with visual data.

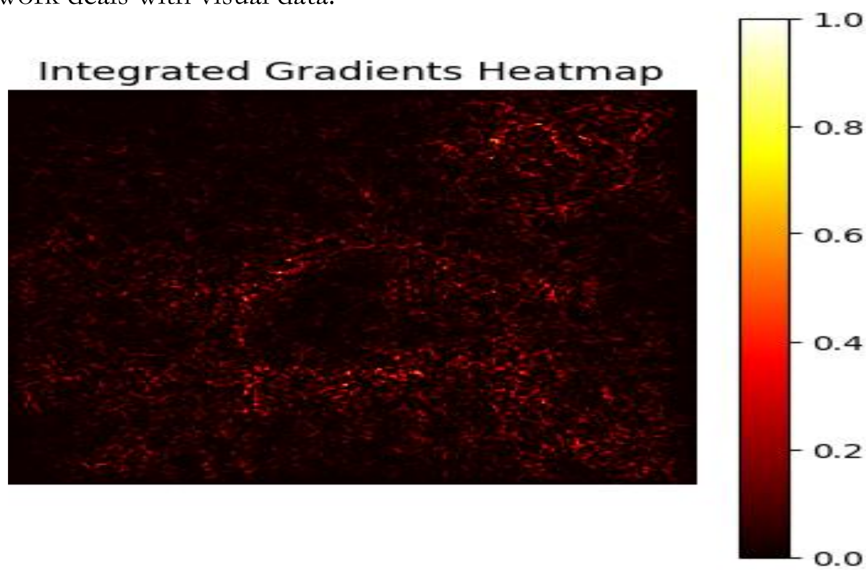

Figure 4. Integrated Gradients Heatmap

As presented in figure 4, integrated gradients have revealed vital regions in terrain classification concerning model decisions. While the technique shows promise in providing interpretability, further refinements, such as using multi-modal baselines or employing contrast-enhanced images, could potentially improve accuracy for low-contrast features. The insights gained through integrated gradients will aid in improving deep-learning-based models for planetary exploration.

Explainability Consistency Analysis:

To further examine the interpretability of the proposed model, the consistency check of the Grad-CAM and Integrated Gradients attribution maps was performed. Whereas Grad-CAM puts emphasis on the locations in the image that affect the prediction process by using convolutional activations of features, Integrated Gradients emphasizes pixel-level locations using gradient integration, using gradient integration. To measure their consistency, the overlap between the salient areas generated by the two methods was measured. The findings reveal that the two approaches consistently point out important terrain features like crater rims, dune boundaries as well as impact ejecta patterns. The overlapping highlighted areas indicate that the model concentrates on the geological features that are of scientific interest to arrive at predictions. It was also discovered that Grad-CAM gives more geological structures of large scale in terms of their spatial localization whereas the Integrated Gradients yield more fine-grained contributions on a pixel-by-pixel basis. The above-mentioned complementary quality of these explainability methods enhances better interpretability because it gives spatial and feature-level results on how the model has made decisions. This two-explainability system makes deep learning models in the study of planetary exploration more transparent.

Baseline Comparison:

The suggested model was compared with popular semantic segmentation models such as U-Net and DeepLabV3+. To ensure a fair comparison, all the baseline models were trained according to the same dataset, preprocessing pipeline, input resolution, and training protocol. This ensured that discrepancies in performance could be attributed to architectural disparities as opposed to training circumstances. The results of the evaluation are summarized in Table 3, which provides the overall accuracy of each model. The attention-enhanced architecture has been proposed and proves to be better than the baseline approaches.

Ablation Study:

Table 3. Ablation Study of Proposed Framework

Model Configuration	Attention Module	Segmentation Head	Accuracy (%)	IoU (%)	F1-score (%)
ResNet-50 Baseline	✗	✗	88.20	78.10	83.45
ResNet-50 + Segmentation Head	✗	✓	90.35	81.25	85.92
ResNet-50 + Attention	✓	✗	91.10	82.60	86.75
Proposed Model (Full)	✓	✓	**92.61**	**84.70**	**88.48**

Ablation study was carried out to determine the contribution of individual elements of the attention mechanism and segmentation head. The initial ResNet-50 model did not perform very well, but the addition of the segmentation head enhanced the spatial localization and pixel-level classification. Equally, the attention module also increased the ability to discriminate features by concentrating on the pertinent terrain areas. The overall model which combined attention and segmentation models recorded the best performance, which validates that the proposed architecture was effective in enhancing the segmentation of the Martian terrain.

Limitations:

The proposed method has some limitations. First, the model was only tested on one dataset, which may not be representative of all Martian missions and image capture conditions. Second, the model has lower accuracy on under-represented classes, as a result of class imbalance and similarities between classes. Third, while efficient, the model needs further optimization to run in real-time on an embedded system with limited resources. Lastly, illumination and terrain complexity variability may impact its robustness, suggesting the need for a more varied dataset and dynamic learning techniques.

Conclusion:

In this research work, an attention-enhanced deep learning architecture of the Martian terrain segmentation was suggested, which included a ResNet-50 backbone that was supplemented by channel-spatial attention and a self-designed segmentation head. The model had a high level of performance as it recorded an accuracy of 92.61, mean IoU of 84.7 and F1-score of 88.48 which was better than baseline models, including U-Net and DeepLabV3+. The findings indicate that the attention mechanisms enhance feature discrimination and segmentation accuracy in a variety of terrain classes. The analysis by classes revealed good performance of major terrain types, including dunes and craters, whereas the good performance of minor classes indicates the issue of class imbalance and image similarity. Grad-CAM and Integrated Gradients further validated that the model pays attention to important geological areas, improving interpretability. Practically, the suggested framework will be able to assist in the autonomous navigation of the rovers, analysis of the terrain and the planning of the mission, as it will be in a position to provide accurate and interpretable terrain maps. In general, the method represents a dependable means of mapping the terrain of Mars with high precision. The future work could involve enhancing generalization, efficiency of computation, and testing of more advanced architectures, like transformer-based models.

Future Work:

The proposed model can be further improved in a number of ways. First, testing on other high-resolution data (e.g., HiRISE) and multi-mission data can enhance model robustness over different Martian environments. Second, model compression, quantization and efficient models should be considered to allow real-time operation on rover processors. Third, the inclusion of multimodal data (such as depth, thermal or spectral) can improve performance across different lighting and terrain conditions. Finally, novel architectures (e.g., transformers, hybrid networks) can potentially enhance model performance and flexibility.

Author's Contribution: K.N.: Literature review, Methodology, Experiments, Result analysis, Writeup; M.U.: Administration, Study supervision, Reviewing, revising, and improving the manuscript; M.S.: Formatting and manuscript editing.

Conflict of Interest: The authors declare no conflicts of interest related to this study.

Project Details: This research was carried out as part of the Deep Space Exploration Group (DSEG) led by Dr. Muhammad Umair.

References:

[1] Junbo Li, Keyan Chen, Gengju Tian, Lu Li, Zhenwei Shi, "MarsSeg: Mars Surface Semantic Segmentation with Multi-level Extractor and Connector," *arXiv:2404.04155*, 2024, [Online]. Available: https://arxiv.org/abs/2404.04155

[2] Boyu Lin, Fei Wang, "LisseMars: A Lightweight Semantic Segmentation Model for Mars Helicopter," *Aerospace*, vol. 12, no. 12, p. 1049, 2025, doi: https://doi.org/10.3390/aerospace12121049.

[3] Yuan Ma, Zhaojin Li, "DepthFormer: Depth-Enhanced Transformer Network for Semantic Segmentation of the Martian Surface From Rover Images," *Earth Sp. Sci.*, vol. 12, no. 6, 2025, [Online]. Available: https://www.semanticscholar.org/paper/DepthFormer%3A-Depth-Enhanced-Transformer-Network-for-Ma-Li/8ad9d389573aef7a01faefcb327bf144e3f29fc0

[4] Raul Castilla-Arquillo, Carlos Perez-del-Pulgar, Levin Gerdes, Alfonso Garcia-Cerezo, Miguel A. Olivares-Mendez, "OmniUnet: A multimodal network for unstructured terrain segmentation on planetary rovers using RGB, depth, and thermal imagery," *2025 Int. Conf. Sp. Robot.*, 2025, [Online]. Available: https://arxiv.org/abs/2508.00580

[5] Shan Lin, Quanke Hu, "An interactive gated-attention network with dual streams for semantic understanding of Martian surface images," *Acta Astronaut.*, vol. 241, pp. 529–546, 2026, doi: https://doi.org/10.1016/j.actaastro.2026.01.002.

[6] L. Rubanenko, S. Pérez-López, J. Schull and M. G. A. Lapôtre, "Automatic Detection and Segmentation of Barchan Dunes on Mars and Earth Using a Convolutional Neural Network," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 14, pp. 9364–9371, 2021, doi: 10.1109/JSTARS.2021.3109900.

[7] Mirali Purohit, Bimal Gajera, Vatsal Malaviya, Irish Mehta, Kunal Kasodekar, Jacob Adler, Steven Lu, Umaa Rebbapragada, Hannah Kerner, "Mars-Bench: A Benchmark for Evaluating Foundation Models for Mars Science Tasks," *arXiv:2510.24010*, 2025, [Online]. Available: https://arxiv.org/abs/2510.24010

[8] Yang Fang, Xuefeng Rao, "MTSNet: Joint Feature Adaptation and Enhancement for Text-Guided Multi-view Martian Terrain Segmentation," *MM 2024 - Proc. 32nd ACM Int. Conf. Multimed.*, 2024, [Online]. Available: https://dl.acm.org/doi/10.1145/3664647.3681430

[9] Wenjing Wang, Lilang Lin, Zejia Fan, Jiaying Liu, "Semi-Supervised Learning for Mars Imagery Classification and Segmentation," *arXiv:2206.02180*, 2022, [Online]. Available: https://arxiv.org/abs/2206.02180
[10] L. Yang, G. Huang, and Y. Xing, "Improved Mars Terrain Segmentation in Terms of Style Transfer and Network Architecture," *Int. J. Aeronaut. Sp. Sci. 2024 253*, vol. 25, no. 3, pp. 1121–1134, Feb. 2024, doi: 10.1007/s42405-023-00702-4.
[11] Shrey Malvi, Hitansh Shah, "Automated Multi-class Crater Segmentation in Mars Orbital Images," *GeoAI 2023 - Proc. 6th ACM SIGSPATIAL Int. Work. AI Geogr. Knowl. Discov.*, 2023, [Online]. Available: https://dl.acm.org/doi/10.1145/3615886.3627748
[12] Ryusei Gichu & Kazunori Ogohara, "Segmentation of dust storm areas on Mars images using principal component analysis and neural network," *Prog. Earth Planet. Sci.*, vol. 6, 2019, [Online]. Available: https://link.springer.com/article/10.1186/s40645-019-0266-1
[13] D. M. DeLatte, S. T. Crites, N. Guttenberg, E. J. Tasker and T. Yairi, "Segmentation Convolutional Neural Networks for Automatic Crater Detection on Mars," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 12, no. 8, pp. 2944–2957, 2019, doi: 10.1109/JSTARS.2019.2918302.
[14] H. Liu, M. Yao, X. Xiao, and Y. Xiong, "RockFormer: A U-Shaped Transformer Network for Martian Rock Segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, doi: 10.1109/TGRS.2023.3235525.
[15] J. Li, S. Zi, R. Song, Y. Li, Y. Hu, and Q. Du, "A Stepwise Domain Adaptive Segmentation Network With Covariate Shift Alleviation for Remote Sensing Imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, doi: 10.1109/TGRS.2022.3152587.
[16] D. Gasperini, W. Masawat, S. Santra, and K. Yoshida, "CloverNet: A Real-Time Network for Semantic Segmentation Onboard Edge Devices Towards Planetary Exploration," *9th 2023 Int. Conf. Control. Decis. Inf. Technol. CoDIT 2023*, pp. 333–338, 2023, doi: 10.1109/CoDIT58514.2023.10284509.
[17] R. Mulvi, S. Gauns, A. Gaonkar, S. Naik, S. Aswale, and P. Shetgaonkar, "Detection and Segmentation of Different Structures from Satellite Images," *Proc. 3rd Int. Conf. Intell. Eng. Manag. ICIEM 2022*, pp. 384–389, 2022, doi: 10.1109/ICIEM54221.2022.9853106.
[18] F. Mohammad, Y. Gao, S. Kay, R. Field, M. De Benedetti, and E. V. Ntagiou, "Deep Learning based Semantic Segmentation for Mars Rover Terrain Classification," *2024 Int. Conf. Sp. Robot. iSpaRo 2024*, pp. 292–298, 2024, doi: 10.1109/iSpaRo60631.2024.10687827.
[19] H. Liu, M. Yao, X. Xiao, B. Zheng, and H. Cui, "MarsScapes and UDAFormer: A Panorama Dataset and a Transformer-Based Unsupervised Domain Adaptation Framework for Martian Terrain Segmentation," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, pp. 1–17, 2024, doi: 10.1109/TGRS.2023.3343109.