Mars Terrain Semantic Segmentation

M. Alipoor*, G. Annaloro*, M. Tirabassi*

*All authors contributed equally; authors are listed alphabetically.

1 Introduction

This work focuses on semantic segmentation of grayscale Mars terrain images using the U-Net architecture. The study addresses challenges such as class imbalance and texture variability, exploring techniques like data augmentation, tiling, and custom loss functions to optimize model performance and improve segmentation accuracy.

2 Dataset Analysis

The dataset consisted of 2,615 (64, 128) grayscale images with 5 distinct labels, including the background. Significant class imbalance was observed (further discussed in following sections).

2.1 Outliers

Initial visual inspection revealed the presence of images containing "gray aliens" with corresponding artificial segmentation masks. These outliers accounted for 110 samples and were removed by eliminating all image/label pairs with identical artificial segmentation masks.

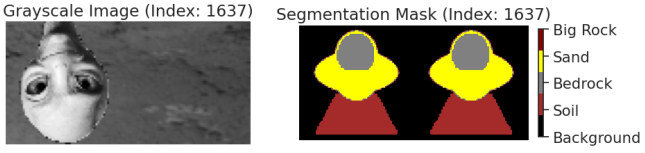

Figure 1: Example of an outlier image (left) and its corresponding segmentation mask (right).

2.2 Texture Analysis

Initial visual inspection highlighted the variety of textures representing different classes. It was hypothesized that this variability could pose challenges, mitigated by having sufficient data to capture it [1].

Texture variability was assessed using GLCM features: contrast, correlation, energy, and homogeneity. Sand and Background exhibited high variability, complicating generalization, while Soil displayed consistent textures with narrower feature distributions. Correlation effectively differentiated classes like Soil and Bedrock, whereas contrast showed overlap.

The underrepresentation of "Big Rocks" further hindered its categorization despite consistent textures. These findings underscored the importance of refined feature selection and addressing class imbalance [1]. Ultimately, the classes with the highest variability turned out to be those with the greatest representation, which helped balance the problem.

3 Data Preprocessing

To mitigate the impact of underrepresented classes, a conservative train-test-validation split was used.

During development, Mean Intersection over Union (IoU) was adopted as the primary evaluation metric, where the background class was not considered. It was observed that this setup consistently yielded marginally better online performance (reported in the tables as "Score") compared to offline testing, providing a reliable reference for offline evaluation.

Preprocessing techniques were compared using a traditional U-Net with skip connections, rather than the simpler model without them. This decision was based on the fact that skip connections preserve spatial information lost during downsampling, making the model more sensitive to fine details and allowing for a more accurate evaluation of preprocessing effects [4].

Tiling was chosen due to its benefits in handling large images [1]. By breaking high-resolution images into smaller, more manageable chunks, tiling improves memory efficiency and allows for faster processing. This approach also aids in model generalization, as it provides a diverse set of localized examples, helping the model learn more effectively from different regions of the image.

4 Class Imbalance

Severe underrepresentation of one class was identified.

Class	Pixels	Frequency (%)
Background (Class 0)	4,988,826	24.31
Soil (Class 1)	6,957,538	33.90
Bedrock (Class 2)	4,776,810	23.28
Sand (Class 3)	3,770,823	18.38
Big Rock (Class 4)	26,963	0.13

Table 1: Class distribution with pixel counts and frequencies.

Class imbalance was identified as a primary concern during data preprocessing [3]. It was hypothesized that severe underrepresentation (0.13% of the dataset) could result in the model ignoring it or failing to learn its features effectively. Techniques such as oversampling were preemptively considered to address the issue.

4.1 Data Sampling

Oversampling (OS) techniques involved augmenting the underrepresented class using traditional data augmentation methods.

Setup	Test Loss	Mean IoU (%)
Without OS	0.76	41.76
With OS	**0.71**	**43.51**

Table 2: Comparison of Test Loss and Mean IoU with and without oversampling.

The performance improvement was marginal.

5 Loss Functions

Loss functions were crucial in guiding the training of the segmentation model, particularly in addressing class imbalance and improving boundary precision. Extensive testing ensured that the chosen loss function effectively aligned with the specific challenges of the task [2].

An extensive search across multiple phases was conducted to evaluate different loss functions.

Function	Mean IoU (%)
Sparse (Baseline)	42.39
Weighted Cross-Entropy	39.20
Focal	42.35
Dice	23.00
Dice + Weighted CE	22.23

Table 3: Comparison of Mean IoU scores for different loss functions.

Class weights for weighted cross-entropy were derived from pixel representation in the training set.

Focal loss was the primary choice during the development process until a breakthrough led to the adoption of sparse categorical cross-entropy without class weights for the final model.

5.1 Focal Loss

Focal loss was initially thought as the most promising approach for dealing with underrepresented classes.

$$\mathcal{L}_{\text{focal}} = -\alpha(1 - p_t)^{\gamma} \log(p_t)$$

where p_t is the model's estimated probability for the true class, α is a weighting factor for the class, and γ is the focusing parameter that adjusts the rate at which easy examples are down-weighted.

A grid search was conducted to optimize its parameters.

α	γ	Val Mean IoU (%)
0.25	**2.0**	**44.65**
0.50	1.0	43.92
0.50	2.0	43.60
0.25	1.0	43.27

Table 4: Grid search results for focal loss hyperparameters.

6 Data Augmentation

Data augmentation was employed to enhance overall performance by applying several types of transformations: geometric transformations (denoted as G), noise injection (denoted as N), and texture variations (denoted as T) [1]. Geometric transformations were also applied to the segmentation masks to ensure consistency.

Additionally, the increased dataset size resulting from augmentation contributed to improved training efficacy, as larger datasets generally lead to better model generalization and reduced overfitting.

Augmentation	Mean IoU (%)	Score (%)
G + N	**44.55**	**45.70**
N + T	22.98	22.33
T + G	35.02	35.33
G + N + T	35.39	36.31

Table 5: Performance of different data augmentation combinations based on Mean IoU and Score.

7 Architecture

The work focused on U-Net-based architectures and nonlinear variations of U-Net blocks.

Model Addition	Mean IoU (%)	Score (%)
Gated Skip	41.02	43.24
Attention Dec.	41.01	42.12
Residual Conn.	39.01	40.90
Deep Superv.	42.80	45.34

Table 6: Performance of different U-Net architecture modifications based on Mean IoU and Score.

Despite initial improvements, performance stagnated, suggesting that the primary bottleneck was the model's base architectural complexity, specifically the number of filters and depth of the network. As a result, efforts shifted toward overfitting the model to identify the optimal complexity that balanced generalization, though limitations were encountered.

A significant breakthrough was achieved by excluding the background class from the loss function, addressing its overrepresentation. This adjustment led to a notable performance improvement, after which a trial-and-error approach was used to further fine-tune the architecture.

8 Final Model

The architecture is a dual U-Net model combining standard and dilated convolutions. It includes separate low-scale and high-scale paths, each with two downsampling and upsampling blocks, followed by a bottleneck with a squeeze-and-excitation (SE) attention mechanism. The outputs of both networks are concatenated and passed through a softmax layer for classification. Spatial dropout is applied throughout to reduce overfitting.

The model is designed to capture multi-scale features effectively, leveraging both standard and dilated convolutions for improved segmentation performance.

Model	Mean IoU (%)	Score (%)
Final Model	60.51	69.06

Table 7: Performance of the final U-Net architecture based on Mean IoU and Score.

References

[1] A. K. Gupta, P. Mathur, F. Sheth, C. M. Travieso-Gonzalez, and S. Chaurasia. Advancing geological image segmentation: Deep learning approaches for rock type identification and classification. *Applied Computing and Geosciences*, 23:100192, 2024.

[2] S. Jadon. A survey of loss functions for semantic segmentation. In *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, page 1–7. IEEE, Oct. 2020.

[3] K. Liu. Semi-supervised confidence-level-based contrastive discrimination for class-imbalanced semantic segmentation, 2022.

[4] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.