POLITECNICO
MILANO 1863

AN2DL - Second Homework Report
AXON Biotech

Matteo Balice, Antonio Giuseppe Doronzo, Alessandro Masini, Marco Muraro

matteobalice, imtonio, jumpit, marcomuraro4

251648, 251320, 232814, 103391

December 14, 2024

1 Introduction

Semantic segmentation is pivotal in planetary exploration, enabling precise classification of surface features for tasks like rover navigation and geological analysis. Recent advancements have leveraged deep learning for Martian terrain segmentation with notable success [1] [2] [3] [4]. Our approach includes:

- **Data preparation**: The dataset was enhanced by removing incorrect images and masks and data augmentation to improve generalization.
- **Model design and training**: A custom neural network architecture was designed and optimized to effectively capture segmentation features.
- **Validation**: Model performance was evaluated using a custom validation set and the **Mean Intersection over Union (Mean IoU)** metric.
- **Post-processing**: Post-processing techniques were applied to analyze and refine segmentation results.

2 Problem Analysis

The dataset comprises 2615 (64×128) grayscale images with pixel-wise labelled masks. Data exhibits **significant class imbalance**, with underrepresented terrain types posing challenges for feature extraction and model generalization. This mirrors broader difficulties in Martian surface segmentation, which is complicated by **minimal inter-class differences**, **imbalanced categories**, and **limited data availability** due to the unique, non-structural nature of the Martian landscape.
The project requirement to build a **neural network from scratch** adds further complexity, as it precludes the use of pretrained architectures that could streamline development. Instead, this constraint necessitates a carefully designed model architecture, tailored to the problem.
Moreover, **data cleaning** and **augmentation techniques** are considered essential to *mitigate class imbalance* and *enhance generalization*. These strategies form the basis for developing a robust and effective segmentation approach [2].

3 Method

This work proposes a deep learning methodology to perform semantic segmentation of unstructured Mars terrain, featuring a **U-Net-based model built and trained from scratch**, strengthen with a dedicated post-processing pipeline. Considering the small size of the available dataset, class imbalance and poorly segmented images, the solution re-

quires a carefully designed method, whose macro steps are illustrated in Figure 1.

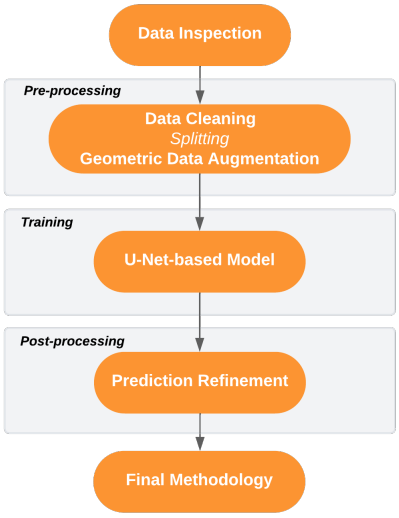

Figure 1: This high-level diagram describes the macro steps of the proposed methodology.

Firstly, accurate **data inspection** was conducted so as to extract relevant information about the data distribution and possible outliers. Subsequently, a **data cleaning** pipeline was designed accordingly to remove completely wrong images and masks, and increase the quality of information to be used later on for training. Moreover, after **data splitting** into *training* (80%) and *validation* (20%) sets, **geometric data augmentation**, including *Random Rotation, Translation, Zoom and Flip*, was applied to images and masks so as to improve model robustness and its ability in generalization.

The design of the **pre-processing chain** described above was driven by broad experimentation, starting from a standard U-Net, aiming to assess the most relevant **intra-class and inter-class discriminating features**.

The final model consists of a **U-Net architecture** featuring a **transformer-based bottleneck**. The network includes a **Data Augmentation Sequential chain**, whose output is then fed into the encoder structure. This augmentation pipeline involves transformations such as *Random Brightness, Contrast, Gaussian Blur, and Gaussian Noise*, affecting original images only, while keeping the corresponding masks unchanged.

The **compression path** is composed of two blocks, each one consisting of two **Convolutional (Conv2D) layers** with ReLU activation and padding *same*, followed by a **Max Pooling (MaxPooling2D) layer** performing downsampling operation. The first block employes 64 (3x3) filters in both the two convolutional layers, while the second one performs a deeper feature extraction by means of 128 (3x3) filters.

The **bottleneck** includes a Conv2D layer featuring 256 (3x3) filters with ReLU activation and padding *same*, whose output is fed into an **Encoder-only Transformer block**. This block implements **Multi-Head Attention mechanism**, followed by a light **Feedforward Neural Network**. Moreover, **Layer Normalization** and **Residual Connections** ensure stability during training. This kind of Attention mechanism allows the bottleneck to *focus on multiple aspects of the data simultaneously*, thus *extracting deeper information* based on a **context-dependent selective focus**.

The **expansion path** of the entire U-Net-based architecture, specular to the compression one, is formed by two blocks featuring **Transposed Convolution (Conv2DTranspose)**, performing upsampling operation, and **Adaptive Fusion**, computing a weighted sum of features extracted at the same level in the encoder and decoder paths respectively. This block is then followed by a pair of **Conv2D layers**.

The model uses a combination of **Weighted Categorical Cross-Entropy**, **Focal Loss**, and **Dice Loss**. **Weighted Categorical Cross-Entropy** aims to maximize the probability of pixels to belong to a specific class. **Focal Loss** concentrates on the most difficult classes, in our case class 2 (Bedrock) and class 4 (Big Rock), which are often misclassified. **Dice Loss** focuses its attention on class 4, which is underrepresented in the dataset.

The methodology also includes a dedicated **post-processing** stage, which consists of a **lightweight U-Net**. This additional model takes as input the output masks from the first architecture and aims to improve the predictions by employing a loss function which is the weighted sum of **Intersection over Union Loss** for the classes from 1 to 4. Class 0 (Background) is excluded since it is not considered in the model performance evaluation. This network was trained with only 30 epochs on the validation set so as to prevent from overfitting.

For parameters and hyperparameters values, and additional details refer to the attached notebooks.

4 Experiments

The proposed methodology is the result of **broad experimentation** with different kinds of architectures and various arrangements for pre-processing and post-processing pipelines.
The very *first experiment* was conducted employing a simple pre-processing chain with rough data cleaning and basic geometric transformations and a **standard U-Net architecture** featuring *standard convolutional block configuration (Conv2D, Batch Normalization, ReLU Activation)* for downsampling, bottleneck and upsampling part.
Subsequently, the pre-processing pipeline was refined, while the model customized with a **dual-branch bottleneck** focusing simultaneously on *global context* (chain of convolutional blocks having the same structure mentioned above) and *fine-grained details* (squeeze-and-excitation chain).
These models led to a **Mean IoU** ranging in between **0.41** and **0.47** on Kaggle Leaderboard. On a later stage, an **Attention U-Net** was also trained and validated without significant improvements.
The **final model architecture** achieved **0.50483 Mean IoU** *without post-processing* and **0.62904** *with post-processing* (on Kaggle).

Table 1: This table evidences the resulting Mean IoU for each Model.

Model	Mean IoU
Standard U-Net	0.43280
Advanced U-Net	0.46183
Transformer-based U-Net	0.50483
Transformer - Post-processing	**0.62904**

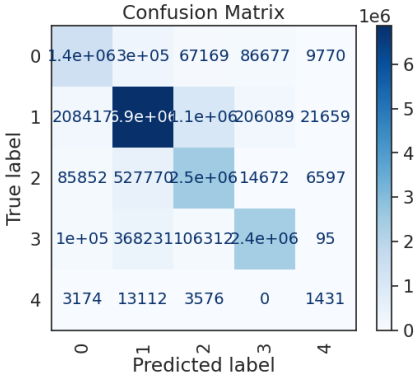

Figure 2: Confusion Matrix computed from the results of the final model.

5 Results

Compared to baselines in the literature, such as SegMarsViT [1] and the Mars Terrain Segmentation Network [2], our approach delivered competitive results despite simpler architecture and less extensive pre-processing. Unexpectedly, class 4 proved challenging to classify due to its severe underrepresentation and errors in the ground truth masks. Similarly, frequent mislabeling of the background class in the dataset led to segmentation inaccuracies. Despite these issues, the redefinition of the background class significantly enhanced overall performance.

6 Discussion

Our best model achieved a good Mean IoU, benefiting from the use of **transformer-based architecture** and the application of the additional **post-processing network**. This result highlights the model's capacity to accurately segment Martian terrain features. Also, in a *real world scenario*, we expect to deal with well-segmented images and, in case of raw data, re-label noisy samples to improve information quality (useless for this challenge).

7 Conclusions

Our **final segmentation model** achieved **Mean IoU** of **0.50483** and **0.62904** for the two best-performing configurations. These results can be further improved by means of a heavier model trained on a larger and well-segmented dataset. Also, a better post-processing network can enhance the final model performance even more.

Contributions

Team working and meticulous organization of activities among teammates was fundamental to achieve the final goal.
Matteo Balice - Data Cleaning, Transformer and Post-Processing, Loss Functions
Antonio Giuseppe Doronzo - Data Inspection, Advanced U-Net with Attention Mechanisms
Alessandro Masini - Data Insepction and Cleaning, Transformer-based U-Net
Marco Muraro - State-of-the-Art Techniques, Dual-Branch Bottleneck, Attention U-Net

References

[1] Y. Dai, T. Zheng, C. Xue, and L. Zhou. Segmarsvit: Lightweight mars terrain segmentation network for autonomous driving in planetary exploration. *Remote Sensing*, 14:6297, 12 2022.

[2] J. Li, K. Chen, G. Tian, L. Li, and Z. Shi. Marsseg: Mars surface semantic segmentation with multi-level extractor and connector. *arXiv preprint arXiv:2404.04155*, 2024.

[3] H. Liu, M. Yao, X. Xiao, and H. Cui. A hybrid attention semantic segmentation network for unstructured terrain on mars. *Acta Astronautica*, 204:492–499, 2023.

[4] Y. Wang, L. Yang, and G. Huang. Semantic segmentation of martian terrain based on dual-branch. In *ICETIS 2022; 7th International Conference on Electronic Technology and Information Science*, pages 1–5, 2022.