# AN2DL - Second Homework Report
## Spritzers

Federico Lombardo, Camilla Magnelli, Edoardo Margarini

federicoozzz, camillamagnelli, edoardomargarini

252705, 226152, 252733

December 14, 2024

## 1 Introduction

This project tackles the problem of *semantic segmentation* of 64x128 grayscale images of Martian terrain, aiming to classify each pixel into one of five classes, including the background. The objective is to develop models using a **scalar approach**, beginning with basic solutions and incrementally incorporating advanced techniques to achieve precise segmentation.

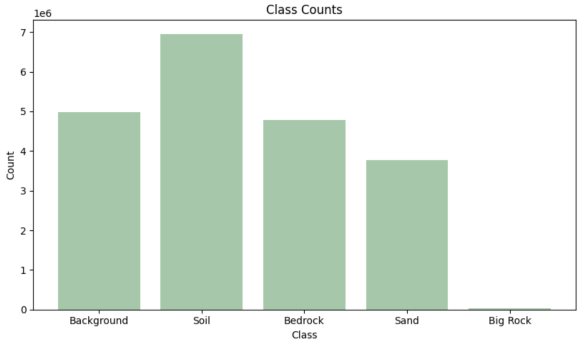


Figure 1: Class Distribution

## 2 Problem Analysis

### 2.1 Dataset Inspection

The dataset comprises gray-scale segmented images of Martian terrain, each paired with a mask that assigns pixels to five categories: background, soil, bedrock, sand, and big rocks. Each image has a resolution of $64 \times 128$ pixels, resulting in an input shape of $(64, 128)$. The **training set** includes 2,615 labeled samples, while the **test set** contains 10,022 unlabeled images reserved for evaluation.
An analysis of the dataset revealed a significant class imbalance (Figure 1), with the background class dominating most images, while the class big rock occurs less frequently. This imbalance necessitates strategies to prevent the model from favoring majority classes over underrepresented ones.

Additionally, the gray-scale nature of the images limits the information available for segmentation, forcing the model to rely solely on texture and intensity variations to differentiate between classes.

Figure 2: Outlier sample and mask

Finally, several anomalies were identified in the masks (Figure 2). These were mislabeled or noisy masks that differed from the expected structure of the data. These outliers were removed during preprocessing to improve the quality of the dataset, ensuring cleaner inputs for model training.

### 2.2 Main Challenges

The primary challenges of this project centered on balancing effective learning with generalization while addressing several key issues, outlined below:

1. **Avoiding Overfitting**: Given the relatively small size of the training set, there is a risk of overfitting, where the model learns to memorize the training data rather than generalizing to unseen images.

2. **No Transfer Learning**: The inability to utilize pre-trained models required training the network from scratch, increasing the complexity of the task

3. **Class Imbalance**: The inherent class imbalance and terrain complexity made it challenging to achieve consistent segmentation accuracy, particularly for underrepresented classes.

By addressing these challenges, the project aims to develop a robust segmentation model capable of effectively analyzing Mars terrain.

## 3 Method

The proposed approach to semantic segmentation follows a modular pipeline, consisting of data preprocessing, a custom U-Net-based model architecture, and an advanced training strategy.

### 3.1 Data Preprocessing

Preprocessing included **normalization** of pixel values to the range $[0, 1]$, random horizontal and vertical flips for **data augmentation**, and the application of **class weights**, computed as the reciprocal of pixel frequency for each class to address class imbalance, with reduced weight assigned to the background class to prioritize terrain features.

### 3.2 Model Architecture

The architecture is based on a **dual-branch U-Net** structure with enhancements for improved feature learning and leveraging over 950000 trainable parameters:

- **Downsampling Path**: Each branch contains two convolutional blocks, followed by max pooling layers. Each block employs `Conv2D` layers, batch normalization, ReLU activations, and residual connections.

- **Bottleneck**: A bottleneck with dropout layers and a squeeze-and-excitation block is used to improve channel-wise feature importance.

- **Upsampling Path**: Skip connections link corresponding downsampling and upsampling layers. Two upsampling layers per branch were used to reconstruct spatial dimensions.

- **Final Output**: Outputs from the two branches were concatenated and passed through a `Conv2D` layer with a softmax activation [1] for pixel-wise classification.

$$\sigma(x)_i = \frac{\exp(x_i)}{\sum_{j=1}^{N} \exp(x_j)} \tag{1}$$

where: $x_i$ is the $i$-th element of the input vector $x$, $N$ is the total number of elements in the vector $x$, $\exp(x)$ is the exponential function applied element-wise.

### 3.3 Training Strategy

The model was trained using the Adam optimizer with an initial learning rate of $10^{-3}$ and a custom loss function combining ***DiceLoss*** and ***FocalLoss*** [2] to effectively address class imbalance and improve segmentation accuracy.

$$\text{DiceLoss} = 1 - \frac{2 \cdot \sum_i (y_{\text{true},i} \cdot y_{\text{pred},i} \cdot w_i)}{\sum_i (y_{\text{true},i} + y_{\text{pred},i}) \cdot w_i} \tag{2}$$

$$\text{FocalLoss} = -\alpha \cdot \sum_i w_i \cdot (1 - \hat{y}_i)^{\gamma} \cdot y_{\text{true},i} \cdot \log(\hat{y}_i) \tag{3}$$

The training process incorporated several callbacks to optimize performance:

- **Early stopping** was used to monitor validation performance, halting training after 30 epochs of stagnation and restoring the best weights.

- **Reduce LR on Plateau** dynamically reduced the learning rate by 50% when improvements plateaued, with a minimum threshold set at $10^{-5}$.

- **Visualization callback** provided periodic insights into model predictions on validation images, enabling real-time qualitative assessment.

The training was performed with a batch size of 64 to balance computational efficiency and performance. The segmentation performance was evaluated using the **Mean Intersection Over Union (Mean IoU)** metric, excluding background pixels.

# 4 Experiments

To evaluate the performance of our model, we conducted a series of experiments on two parallel tracks, Single U-Net architectures and Dual U-Net architectures, each of which aimed to address specific challenges.

The following experiments regard the **Single U-Net** track. We started with a simple U-Net architecture, which served as a baseline and achieved a val_mean_IoU of **44.82%**. To improve feature extraction, we expanded the bottleneck layer, resulting in a val_mean_IoU of **47.12%**. Data augmentation techniques, such as random flips, were then applied, but provided only marginal improvement, with a val_mean_IoU of **47.41%**. We further introduced weighted Dice and Focal loss, reducing the weight for the background class to prioritize terrain features, which notably improved the val_mean_IoU to **51.63%**. Lastly, the integration of squeeze-and-excitation (SE) blocks brought notable improvements in feature emphasis, achieving a val_mean_IoU of **61.7%**.

The **Dual U-Net** experiments focused on combining two branches to capture features at multiple scales. In one configuration, a fine branch for detailed feature extraction and a coarse branch for global context were fused, achieving a val_mean_IoU of **48.02%**. We also implemented various types of normalization techniques, such as Group and Layer Normalization, applied differently across the network's branches. However, these approaches did not lead to relevant performance improvements.

The final model incorporated single U-Net features within the dual-branch architecture, balancing segmentation accuracy and computational efficiency, and achieving val_mean_IoU of **70.24%**.

# 5 Results

The table summarizes the experimental results, focusing on validation accuracy and mean IoU, providing a clear overview of model performance across configurations. The best model resulted in the final Dual U-Net mentioned in the section experiments.

| Model/Technique | Mean IoU | Accuracy |
|---|---|---|
| **Single U-Net** | | |
| Baseline | 44.82% | 72.97% |
| Expanded bottleneck | 47.12% | 77.71% |
| Data augmentation | 47.41% | 76.50% |
| Weighted Dice/Focal loss | 51.63% | 78.80% |
| SE blocks | 61.70% | 76.43% |
| **Dual U-Net** | | |
| Fine/coarse fusion | 48.02% | 76.90% |
| Normalization techniques | 48.00% | 77.06% |
| Weighted Classes | **70.24%** | 68.18% |

# 6 Discussion

The model's success highlights that simplicity and targeted enhancements, like SE blocks, outperform overly complex designs. While effective in balancing accuracy and efficiency, some techniques, such as group and layer normalization, provided minimal benefits. Although the background class was assigned minimal weight to reduce its influence, this adjustment subtly affected the overall predictions, occasionally leading to inconsistencies in the classification of underrepresented land types.

# 7 Conclusions

This project developed a U-Net-based model for semantic segmentation of Martian terrain, addressing challenges such as varying terrain features and the need of precise delineation of different surfaces. The results were promising, demonstrating the model's capability to classify terrain types accurately. The project was a cooperative effort in which each team member contributed to its success in a balanced way.

## 7.1 Future Directions

In the current approach, the background class is given a lower weight to reduce its influence. A more effective strategy can involve the use of standard background weights, followed by a secondary model to refine the predictions. Future work could also involve adjusting model parameters or adding pretrained encoders to improve performance.

## References


[1] K. A. Documentation. Activations. https://keras.io/api/layers/activations/.

[2] M. V. Fischer. Weighted dice and focal loss. https://github.com/maxvfischer/keras-image-segmentation-loss-functions/blob/master/losses/multiclass_losses.py.