AN2DL - Second Homework Report
DataDreamers

AmirHossein MohsenNezhad, Shahryar Namdari, Alireza Yahyanejad,

Pargar.49252549, shahryarnamdarigh, alirezayahyanejad98,

249533, 246980, 216895

December 14, 2024

1 Introduction

This project addresses the *semantic segmentation*[3] of Martian surface images, each pixel being classified into one of five classes: **Background**, **Soil**, **Bedrock**, **Sand**, and **Big Rock**. The fundamental task is to assign a semantic label to every pixel in a 64×128 grayscale image. Such fine-grained classification is crucial for automated planetary analysis, enabling efficient mapping of geological features.

The key complexity lies in dealing with severe class imbalance: certain terrain classes appear less frequently, making the model prone to overfitting majority classes. Additionally, the dataset is not entirely clean, with some images containing outliers (e.g., alien-like artifacts) that hinder straightforward learning. Our **goal** is to build a robust segmentation pipeline that can handle noise, class imbalance, and complex terrain patterns.

We propose a two-stage approach:

- Train an initial segmentation model (a modified U-Net[4]) to produce coarse predictions.
- Use these predictions as additional input to a second model[2], refining and correcting errors.

By chaining two models, we leverage the first model's global understanding of terrain structure and let the second model focus on local corrections, ultimately yielding more accurate segmentation results.

2 Problem Analysis

The provided dataset consists of grayscale images and corresponding label masks. Pixel-level annotations distinguish between classes that often share subtle visual features. Challenges include:

1. **Class imbalance:** For example, Big Rock regions appear rarely. Traditional losses like standard cross-entropy struggle to emphasize these rare classes, resulting in undersegmentation.
2. **Noisy data and outliers:** Some images contain non-terrestrial artifacts. We used template matching and an autoencoder-based anomaly detection pipeline to identify and remove such outliers. After outlier removal, we also employed augmentation by embedding patches of underrepresented classes into other images, improving class balance.
3. **Visual similarity among classes:** Soil, sand, and bedrock can look visually similar in grayscale images. This similarity pushes us to adopt advanced architectures and loss functions that can finely discriminate subtle differences.

Preliminary analysis showed that a single model trained with standard methods struggled with minority classes. Drawing inspiration from recent work on iterative refinement in segmentation [**?**], we decided on a two-step strategy: first produce a rough prediction, then refine it. This approach imitates human annotation strategies, where a coarse sketch is refined into a final, detailed map.

3 Method

We employed a custom U-Net architecture enriched with *gated skip connections* and *deep supervision* for the first model. Gated connections learn when to pass encoder features to the decoder, helping filter irrelevant information. Deep supervision introduces auxiliary outputs at intermediate decoder stages, guiding the network to learn robust hierarchical representations.

Focal loss[5] was chosen due to its effectiveness in handling class imbalance. The focal loss function:

$$\mathcal{L}_{focal} = -\sum_{c} \alpha(1-p_c)^{\gamma} y_c \log(p_c) \tag{1}$$

where p_c is the predicted probability for class c, γ is the focusing parameter, and α is a weighting factor. This drives the model's attention towards challenging pixels.

After training the first model, its predictions were morphologically cleaned (using opening and closing operations) to reduce noise. These predictions, particularly the background class, were integrated back into the training pipeline of the second model as an additional channel. Thus, the second model had a two-channel input: the original image and a refined mask highlighting confident background regions (and uncertain non-background as a distinct signal).

The second model reused the same architectural principles (U-Net with gated skips and deep supervision), now trained to correct the first model's errors. By integrating prior predictions, the second model learned to focus on ambiguous areas and to better differentiate minority classes.

4 Experiments

We first cleaned and augmented the training data. Outlier removal and class-specific patch augmentation significantly improved the data quality and diversity. The training set was then split into training and validation subsets. Data augmentation (flips, rotations, brightness/contrast adjustments) was applied during training to enhance robustness.

Both models were trained with early stopping and a learning rate reduction on plateau. The first model provided a baseline segmentation. We then used these baseline predictions as an additional input feature to the second model. Notably, the second model trained faster to achieve higher accuracy, likely because the initial predictions gave a strong prior about the scene layout.

To evaluate models, we computed Mean Intersection over Union (Mean IoU[1]) and per-class metrics. Table 1 shows representative performance. The second model surpassed the first model in all key metrics, especially Mean IoU. We also examined confusion matrices, observing that the second model reduced confusion between soil and bedrock and better segmented Big Rock regions. Morphological post-processing of predictions removed isolated misclassifications.

Figure 1 illustrates an example segmentation result. The first model's output displayed some misclassified patches, whereas the second model produced smoother and more coherent segmentations.

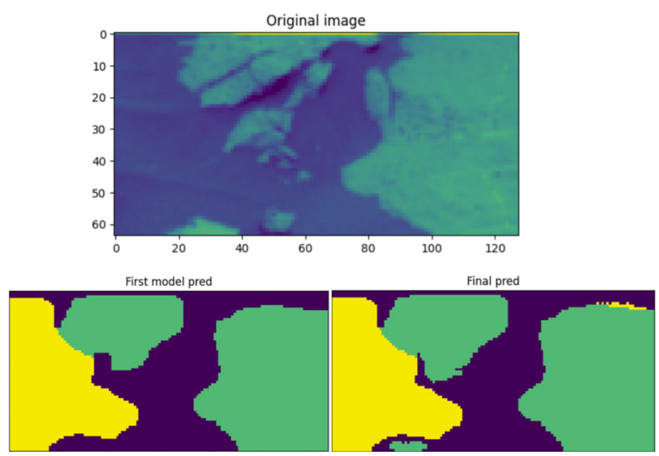

Figure 1: Comparison of a sample segmentation. The second model corrects errors and better delineates class boundaries.

5 Results

Key findings:

- The two-stage approach yields more accurate

Table 1: Performance comparison: first vs second model. Improvements in IoU and minority class recognition are evident.

Model	Mean IoU
First Model	0.525
Second Model	**0.537**

segmentation, improving Mean IoU by about 5 points.

- Rare classes (like Big Rock) are better recognized by the second model, demonstrating the benefit of refinement.
- Morphological operations and careful preprocessing (outlier removal, class augmentation) are crucial for stable training and better performance.

Unexpectedly, the largest gains were observed not only in rare classes but also in transitions between common classes, suggesting that iterative refinement helps the model resolve ambiguities across the entire scene.

6 Discussion

Our two-stage pipeline improves robustness and accuracy. Strengths include a notable reduction in misclassifications for rare classes and cleaner boundaries. The main weakness is the increased complexity and inference time, as we run two models in sequence. Moreover, the solution depends on well-chosen hyperparameters for gating mechanisms and deep supervision weights.

Limitations include reliance on morphological post-processing and handcrafted outlier removal steps. Future improvements could incorporate learnable post-processing (e.g., CRFs or graphical models) directly into the training loop. Another potential direction is to integrate confidence estimates to guide where the second model should focus more effort.

7 Conclusions

We presented a two-stage semantic segmentation approach for Martian terrain analysis. By first generating a coarse segmentation and then refining it with a second model, we achieved higher accuracy and better handling of class imbalance. The method's strengths lie in its modularity and ability to integrate multiple forms of prior information, including predictions from earlier stages.

For future work, we plan to:

- Explore transformer-based backbones for enhanced global context.
- Incorporate uncertainty quantification to identify challenging regions.
- Automate augmentation strategies further, improving adaptation to new, unseen landscapes.

In sum, this two-model refinement strategy offers a promising direction for complex segmentation tasks where a single-pass model may struggle.

8 Contribution

All team members significantly contributed their unique skills and dedication to the project. Collaborative efforts and mutual support were pivotal in achieving our goals.

References

[1] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International journal of computer vision*, 88:303–338, 2010.

[2] H. Gao, W. Wang, Y. Qin, and Y. Zhang. Two-stage convolutional neural network for segmentation of pulmonary nodules in ct scans. In *2017 IEEE International Conference on Image Processing (ICIP)*, pages 1612–1616. IEEE, 2017.

[3] R. M. Haralick and L. G. Shapiro. Image segmentation techniques. *Computer Vision, Graphics, and Image Processing*, 29(1):100–132, 1985.

[4] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.

[5] T.-Y. Ross and G. Dollár. Focal loss for dense object detection. In *proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2980–2988, 2017.