AN2DL - Second Homework Report
OverfittingExorcists

Daniele Laganà, Marcello Martini, Gianluigi Palmisano, Samuele Pozzani

danielelagana, marcellomartini, gianlupalmisano, samuelepozzani

252548, 243880, 244774, 242694

December 12, 2024

1 Introduction

This project focuses on semantic image segmentation using deep learning techniques. The goal is to maximize the accuracy of the model for the segmentation of five classes of Mars terrain.

2 Problem Analysis

Semantic segmentation of grayscale Mars terrain images presents several challenges:

- **Dataset characteristics:** the absence of color information makes it harder to distinguish classes with subtle texture or intensity differences. The dominance of the background class can bias the model toward overpredicting it, neglecting smaller or underrepresented classes. Terrain features often exhibit similar intensity levels, which complicates boundary detection.
- **Limited annotated data:** small datasets due to expensive and time-consuming annotation processes increase the risk of overfitting. Noisy or ambiguous labels may reduce the reliability of ground truth data.
- **Evaluation challenges:** dominance of the background class can skew metrics such as Intersection over Union (IoU), masking poor performance on minority classes.

3 Method

3.1 Data exploration

The dataset comprises 2615 grayscale images, each with a resolution of 64x128 pixels. Each of the five classes represents a specific object category.

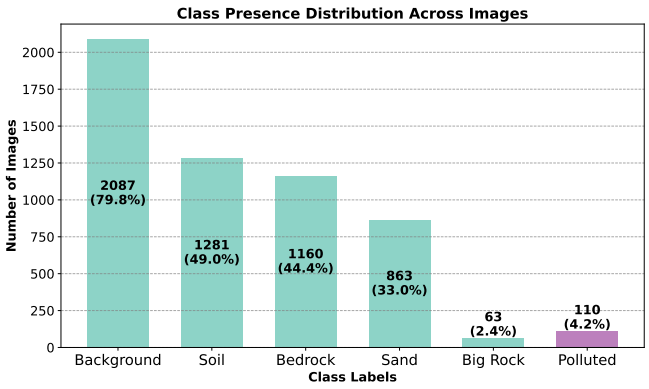

Figure 1: The plot shows the distribution of the predominant class labels across the dataset. 'Background' is the most common class, present in 2,087 images (79.8%). The dataset also includes 110 'Polluted' images (4.2%) containing aliens.

3.2 Data cleaning

The provided dataset is polluted with aliens, which are identified based on their corresponding mask indices. All polluted images share the same mask,

making it possible to detect and remove them systematically. The dataset is cleaned by identifying these polluted mask indices and deleting the associated images.

3.3 Data preprocessing

The cleaned dataset is split into training, validation, and test sets, with image pixel values scaled to the range of $[0, 1]$ for consistent input **normalization**. To tackle class imbalance, **class weights** are calculated and applied during training, giving greater importance to minority classes. This ensures balanced loss contributions, improving the model's performance on underrepresented categories. The weight of the "Big Rock" class is capped due to its strong underrepresentation.

3.4 Image augmentation

Data augmentation is employed to enhance the diversity of the training data by applying various **transformations** to the images and their corresponding labels. Since this is a semantic segmentation task with dense labels, all augmentation functions are implemented manually to ensure that the same transformations are applied consistently to both the images and the labels.

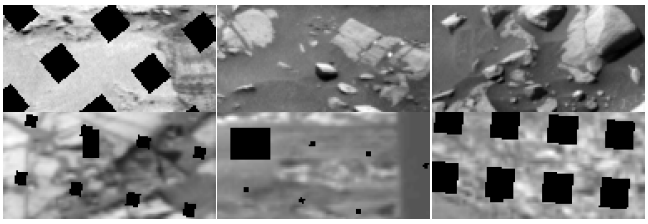

Figure 2: Example of the augmentations performed during training phase.

3.5 Ensemble prediction

Ensemble model prediction is used to improve segmentation accuracy and robustness by combining the strengths of models created with different strategies. Each model contributes unique insights based on its training approach, such as variations in data augmentation, loss functions, or architecture tweaks. By aggregating their predictions, the ensemble mitigates individual model biases, reduces errors, and enhances performance, particularly for complex and ambiguous regions.

4 Experiments

This section shows the evolution of the developed model, with all the related features. The corresponding results are shown in Table 1.

4.1 U-Net

U-Net [1] Ⓐ is a deep learning architecture for semantic segmentation, implemented from scratch in TensorFlow without pretrained weights. It uses a **symmetric encoder-decoder** structure: the encoder extracts features through convolution and **pooling**, while the decoder restores resolution with **upsampling** and convolution. **Skip connections** enhance spatial detail and boundary accuracy, enabling effective segmentation of complex structures. The first version leverages the augmentations illustrated in Section 4.5.

4.2 Combined Losses

U-Net Ⓑ introduces a **weighted combined loss** function to enhance segmentation performance. It uses **Dice Loss** [2] with inverse frequency weighting for rare classes, **Focal Loss** [3] to handle challenging cases, and **Boundary Loss** [4] for precise edge predictions. This approach balances class representation, improves accuracy in difficult regions, and sharpens boundary delineation.

4.3 Residuals and Feature Fusion

U-Net Ⓒ enhances the performance by adding **Residual Blocks** [5] and **Adaptive Feature Fusion** [6] to the standard architecture. Residual Blocks help mitigate the **vanishing gradient** problem, enabling better gradient flow and deeper feature learning. Adaptive Feature Fusion dynamically combines features from different layers, allowing the network to effectively leverage both low-level and high-level information. This approach improves the model's ability to capture fine details and contextual information, leading to more accurate and robust segmentation results.

4.4 Pyramid Pooling and Squeeze-and-Excitation

U-Net Ⓓ with a **Pyramid Pooling Module** [7] and **Squeeze-and-Excitation** [8] blocks aimes to improve feature representation by capturing global

Model	Parameters	Accuracy	Mean IoU	Mean IoU Kaggle
U-Net Ⓐ	1,399,877	67.92%	60.10%	66.68%
U-Net Ⓑ	2,205,061	64.49%	58.50%	57.87%
U-Net Ⓒ	5,907,315	68.18%	60.97%	66.43%
U-Net Ⓓ	4,509,573	67.04%	59.36%	61.46%
Ensemble model	**-**	**70.43%**	**71.53%**	**72.32%**

Table 1: Performance metrics computed using the test dataset. The last column presents the Mean IoU calculated by Kaggle using 25% of the evaluation dataset. The **Ensemble model**, created by combining Ⓐ, Ⓑ, and Ⓒ, outperforms the individual models.

context. However, these techniques proved ineffective for grayscale images due to the single channel limitation, leading to the version's exclusion from the ensemble.

4.5 Augmentation strategies

In addition to standard augmentation techniques, advanced algorithms are used probabilistically to introduce occlusions, distortions, and structured noise. These include **RandomCutout** [9], which masks random portions of an image to encourage robustness to occlusions; **GridMask** [10], which overlays grid-like masks to improve spatial awareness; and **Sobel filtering** [11], which can improve the model's ability to segment distinct regions in semantic segmentation tasks.

4.6 Training settings

As default loss, the **sparse categorical cross entropy** is used. The background class is excluded from the loss computation to focus the optimization on the meaningful foreground classes. The **AdamW optimizer** [12] is selected, with a **learning rate scheduler** that decays on plateau. A batch size of 32 ensures efficient memory use and gradient accuracy, while **early stopping** and the auto-tuned **TensorFlow data pipeline**[1] enhance training efficiency and GPU performance.

5 Discussion

Model performance is evaluated using Mean Intersection over Union (**IoU**), a standard metric for semantic segmentation that measures the overlap between predicted and ground truth regions across all classes. Certain data **augmentation techniques** perform poorly or are incompatible with the segmentation task. Modifications to the **U-Net structure** improve accuracy in specific aspects, such as underrepresented classes and boundary precision, but no single implementation achieves significant overall improvements. The **ensemble model** leverages these individual strengths by combining predictions from different models. The best results are obtained by integrating models Ⓐ, Ⓑ, and Ⓒ, which complement each other to enhance segmentation performance.

6 Conclusions

The project successfully developed and evaluated advanced U-Net-based models for semantic segmentation of Mars terrain, incorporating techniques such as Residual connections and Feature Fusion to enhance feature extraction and representation. Weighted loss functions, including class weights to address class imbalance, and an ensemble approach leveraging the strengths of multiple models, significantly contributed to robust performance.

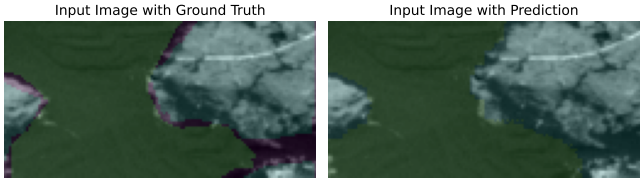

Figure 3: The figure shows a comparison between the ground truth and the predicted mask. The background class is not predicted by the network, as it is excluded from the loss calculation.

[1] TensorFlow Data documentation

References

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[2] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, pages 240–248. Springer, 2017.

[3] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection, 2018.

[4] Hoel Kervadec, Jihene Bouchtiba, Christian Desrosiers, Eric Granger, Jose Dolz, and Ismail Ben Ayed. Boundary loss for highly unbalanced segmentation. *Medical Image Analysis*, 67:101851, January 2021.

[5] Seokyong Shin, SangHun Lee, and HyunHo Han. A study on residual u-net for semantic segmentation based on deep learning. *Journal of Digital Convergence*, 19(6):251–258, 2021.

[6] Neelesh Mungoli. Adaptive feature fusion: Enhancing generalization in deep learning models, 2023.

[7] Hengshuang Zhao, Jianping Shi, Xiaojuan Qi, Xiaogang Wang, and Jiaya Jia. Pyramid scene parsing network. *CoRR*, abs/1612.01105, 2016.

[8] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. *CoRR*, abs/1709.01507, 2017.

[9] Terrance Devries and Graham W. Taylor. Improved regularization of convolutional neural networks with cutout. *CoRR*, abs/1708.04552, 2017.

[10] Pengguang Chen, Shu Liu, Hengshuang Zhao, Xingquan Wang, and Jiaya Jia. Gridmask data augmentation, 2024.

[11] N. Kanopoulos, N. Vasanthavada, and R.L. Baker. Design of an image edge detection filter using the sobel operator. *IEEE Journal of Solid-State Circuits*, 23(2):358–367, 1988.

[12] Ilya Loshchilov and Frank Hutter. Fixing weight decay regularization in adam. *CoRR*, abs/1711.05101, 2017.