# AN2DL - Second Homework Report
# The Bergers

Francesco Palma, Francesco Pellegrini, Luigi Raggi

palmfra, frap3lle, luigiraggi

243800, 247807, 217882

December 13, 2024

## 1 Introduction

The second homework of the 2024-2025 Artificial Neural Networks and Deep Learning course consisted in a **Mars terrain semantic segmentation task**. Given a dataset of images of Mars terrain and relative ground truth labels, the goal was to analyze the data and train a *Convolutional Neural Network* based model able to achieve the best possible mean Intersection over Union on a provided test set. The test set scores were computed by Kaggle, an external deep learning challenges platform that hosted the competition.

## 2 Problem Analysis

The first step taken to tackle the challenge was to load the dataset, stored in a NumPy archive file, and plot some of the contained images and relative labels 1.

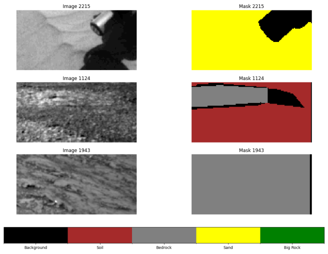

Figure 1: Dataset Mars terrain images with labels

Plotting some random images it was noted that there were outliers across all dataset. Since all outliers shared the same mask it was straightforward to delete them. Then, the distribution of classes across the dataset was plotted to check for an eventual class imbalance.

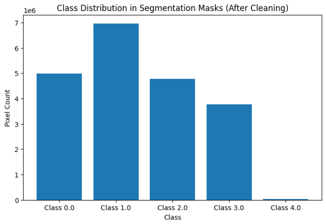

Figure 2: Classes distribution across dataset

The graph 2 confirmed that the dataset was indeed imbalanced, especially regarding the class representing ”big rock”.

## 3 Method

Since the rules of the competition did not allow the use of pretrained models, a custom model was designed (taking inspiration from literature architectures) and trained from scratch. **Model Architecture**: the model that was selected was *MarsSeg*. The reasons are suggested from the paper [6] which includes the model capacity to tackle the complex topography, the similar surface features, and the lack of extensive annotated data of the Martian surface. The MarsSeg framework, as shown in the diagram 3, primarily comprises three integral components: an encoder, feature enhancement connections, and a decoder. The feature enhancement connections are further subdivided into the Mini-ASPP, PSA, and SPPM. The Mini-ASPP and PSA modules collectively form the enhancement connections for shallow features, whereas the SPPM is responsible for the enhancement connections for mid-level features. These enhancement connections are effectively adept at adapting to the complex and variable Martian terrain, as well as the minimal inter-class feature differences.

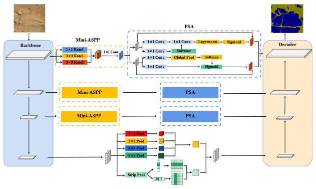

Figure 3: MarsSeg architectural view

**Data Augmentation**: from experiments it was highlighted that the small dataset of the competition was highly sensitive to aggressive augmentations. For this reason, offline augmentation was used and the dataset size was doubled. Using the library *Albumentation* [2], a fast and flexible library for image augmentations, an horizontal flip was applied to all images while a mask dropout (a transformation that randomly selects one label and erase it from both image and label) was applied in one fourth of the images. In addition to that, to address the class imbalance of the fourth class, oversampling of the images containing that class was applied with a rate of four.

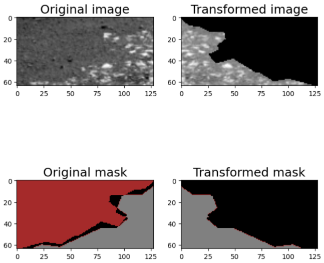

Figure 4: Augmentation with flip and maskdropout

**Optimizer**: to apply an efficient regularization starting from the optimizer *AdamW* was selected. According to the paper [7], it improves Adam's generalization performance. **Loss**: the loss function that yielded the best results overall is the *SparseCategoricalCrossentropy*. It was implemented a custom version of it that ignores the background. This was done to make the model learn better the foreground pixel. The following equation shows the CCE formula where $y_i$ are the ground truth labels and $f(x_i)$ are the predicted labels by the softmax function.

$$CCE = -\sum y_i \log f(x_i) \tag{1}$$

**Learning Rate**: according to the AdamW paper [7], the optimal learning rate is 3-10 times smaller then the one of Adam, so we set a fixed learning rate of 1e-4, with the addition of *ReduceLROnPlateau*, a callback that reduces the learning rate when the validation mean IoU stagnates. Then to prevent excessive overfitting on the training data the callback *EarlyStopping* was used.

## 4 Experiments

In the following section the steps to achieve the final results are are highlighted. The process of inquiry began with the design of the network framework. The analyzed models are essentially based on the same philosophy, they are all encoder-decoder.

U-Net [8] was the first used, the starting model was taken from the sixth laboratory of the course. Starting from this base model, other advanced structures were implemented, in particular: U-Net++ [13], a version with denser skip connections, and a so called Hierarchical Unet. The latter is a sequence of Unet architectures which attempts to resolve coarse and refined scales. The subsequently tested was DeepLabV3+ [3], a popular choise for segmentation tasks, for instance see [9]. The aim was to improve regularization increasing the model size and leaveraging ASPP modules and Dilated Convolutions. Also SegNet [1], EfferDeepNet [11] and Hierarchical MSA [10], other popular models, were implemented. Proceeding further, MarsSeg [6] was analyzed. At the end of this preliminary phase, MarsSeg was chosen as the basic structure.

Based on what was learned in the previous challenge, image augmentation seemed to be the best weapon to improve model prediction. Two strategies were considered. Firstly, The training dataset

was doubled in size with augmentation techniques, and then processed by the network. A second approach considered, was to pre-process the images during training time, developing an augmenting strategy similar to RandAugment [4].
As the last step we investigated different loss functions. Starting from a sparse categorical cross-entropy loss, we develop a weighted version to manage the class imbalance. Afterward, the background class was excluded in the computation of the loss function, being this class omitted evaluating the mean IoU. Then a series of losses were coupled, in particular: sparse categorical cross entrophy, dice loss [12] and boundary loss [5]. The optimal outcome was achieved combining offline augmentation and a sparse categorical cross-entropy loss function that ignored the background class.

## 5 Results

This table 1 shows the outcome of different models in terms of mean IoU on the Kaggle test set.

Table 1: Custom model used

| Model | Best test mean IoU |
|---|---|
| SegFormer | 0.53 |
| Unet | 0.63 |
| DeepLabV3 | 0.68 |
| **MarsSeg** | **0.69** |

## 6 Discussion

The most incisive features in the network were in the order: the custom model type, the loss function and the data augmentation method.
U-Net based models, like U-Net, hierarchical U-Net and U-Net++, being limited by a simple encoder-decoder architecture, were not sufficiently accurate on the Mars terrain dataset. To enhance the competition score the choice of DeepLabV3+, with elements like the Atrous Spatial Pyramid Pooling, was extremely helpful in tackling the challenge. Finally the choice of MarsSeg, a model directly tailored for extraterrestrial ground types like Mars, gave the final score boost. Its reduced complexity, ideal for small datasets, and components such Mini-ASPP (for multi-scale context) and Pixel Spatial Attention (for focusing on relevant regions) helped gain that improvement.
Besides the model, a crucial factor for the achievement of the final score was the loss function. Given how the models were scored in the kaggle competition, namely the metric used for the valuation, it was believed that a loss function that ignores the background pixel could have benefits in term of mean IoU. Regarding the augmentation, the techniques based on pixel-wise transformations were found not just useless but detrimental.
On the contrary, techniques based on geometric transformations had a good, even if slight, impact. The first one being a simple horizontal flip and the other one being maskdropout, a transformation that erase both from the image and from the labels a random selected mask.
The high class imbalance of the last label was highly taken into consideration by the mean IoU, even if the number of pixels where just a fraction of unit. For this reason, oversampling the minority class ”big rock” resulted in a slight final improvement of the score.
Lastly, after some analysis on the images and their relative masks, inconsistencies were seen on the provided dataset likely contained labeling errors, which may have affected the accuracy of the model's evaluation and generalization.

## 7 Conclusions

The results achieved by the group both in terms of the leaderboard and the work done were excellent. This was done thanks to a great cooperation by the team members as well as a great curiosity that pushed the members to go beyond the sufficient assignment. However, there is still room for improvement as the table 2 shows that the bottleneck of the mean IoU is the class ”big rock” that in the dataset is highly underrepresented. Addressing this paves the way for further improvement.

Table 2: Intersection over Union per class

| Label | IoU |
|---|---|
| 1 | 0.8701 |
| 2 | 0.8064 |
| 3 | 0.8755 |
| 4 | 0.2032 |

# References

[1] V. Badrinarayanan, A. Handa, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for robust semantic pixel-wise labelling, 2015.

[2] A. Buslaev, V. I. Iglovikov, E. Khvedchenya, A. Parinov, M. Druzhinin, and A. A. Kalinin. Albumentations: Fast and flexible image augmentations. *Information*, 11(2):125, Feb. 2020.

[3] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation, 2018.

[4] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le. Randaugment: Practical automated data augmentation with a reduced search space, 2019.

[5] H. Kervadec, J. Bouchtiba, C. Desrosiers, E. Granger, J. Dolz, and I. B. Ayed. Boundary loss for highly unbalanced segmentation. In *International conference on medical imaging with deep learning*, pages 285–296. PMLR, 2019.

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

[7] I. Loshchilov and F. Hutter. Fixing weight decay regularization in adam. *CoRR*, abs/1711.05101, 2017.

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

[9] R. M. Swan, D. Atha, H. A. Leopold, M. Gildner, S. Oij, C. Chiu, and M. Ono. Ai4mars: A dataset for terrain-aware autonomous driving on mars. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1982–1991, 2021.

[10] A. Tao, K. Sapra, and B. Catanzaro. Hierarchical multi-scale attention for semantic segmentation, 2020.

[11] Y. Wei, W. Wei, and Y. Zhang. Efferdeepnet: An efficient semantic segmentation method for outdoor terrain. *Machines*, 11(2), 2023.

[12] R. Zhao, B. Qian, X. Zhang, Y. Li, R. Wei, Y. Liu, and Y. Pan. Rethinking dice loss for medical image segmentation. In *2020 IEEE International Conference on Data Mining (ICDM)*, pages 851–860. IEEE, 2020.

[13] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang. Unet++: A nested u-net architecture for medical image segmentation, 2018.