AN2DL - Second Homework Report
Deep Drive

Martina Riva, Lorenzo Rosati, Alberto Eusebio

martina6riva, slrose, albertoeusebio01

249408, 233465, 244659,

December 14, 2024

1 Introduction

This homework focuses on the **multiclass image semantic segmentation** of Mars terrain images. We employed a *U-Net neural network* [4] to address the challenge of segmenting the different terrains, and different processing techniques to achieve a maximun Mean Intersection Over Union of **0.69230** on the test dataset.

2 Problem Analysis

Our initial analysis of the problem involves understanding the dataset characteristics, identifying the main challenges, and stating our initial assumptions:

1. **Dataset characteristics**: The dataset includes 2,615 training images and 10,022 test images, each of size 64x128 pixels in grayscale format. The images come with associated masks that classify each pixel into one of five categories: Background, Soil, Bedrock, Sand, and Big Rock.

2. **Main challenges**:

   - *Terrain categories* in images are often indistinguishable, posing significant challenges for classification.
   - *The dataset contains many outliers* that must be identified and filtered to ensure quality training.
   - *Data augmentation is necessary* due to the limited dataset size, to enhance model robustness and generalizability.
   - *Model complexity* is limited by the constraint of not using pre-trained architectures and the limited train dataset size.

3. **Initial assumptions**: We assume that the provided labels are accurate and well representative of the test set distribution.

3 Methods

3.1 Preprocessing

We performed the following key steps:

- **Outlier Filtering**: Removed non-representative images.

- **Data Augmentation**: Enhanced the dataset's size through **custom preprocessing** functions, including *random rotation, shear, brightness adjustments, gaussian noise, flips, and zoom*, and utilized an **Albumentation** [1] pipeline for advanced online-training augmentations like *GridDropout, RandomBrightnessContrast, RandomRotate90, and RandomScale*.

3.2 Model

Our model utilizes a custom implementation of the U-Net architecture. We utilized Keras, and this is the model's architecture:

- **Input Layer**: Accepts (64, 128, 1) for grayscale images.
- **Contracting Path (Encoder)**: Made of 5 blocks with filters doubling from 64 to 1024. Each block includes convolutions, ReLU, batch normalization, and max pooling.
- **Bottleneck**: Processes the most abstracted features, connecting encoder and decoder with 2048 filters.
- **Expanding Path (Decoder)**: Features up-convolutions and skip connections, gradually refining output to match original image dimensions.
- **Output Layer**: Ends with a softmax layer for pixel-wise classification into five terrain types.

3.3 Training

Optimization strategies included:

- **Callbacks:** *ReduceLROnPlateau* to decrease the learning rate on performance plateaus, and *EarlyStopping* to prevent overfitting by halting training when no improvement is seen.
- **Loss Function:** Used *SparseCategoricalFocalLoss* [3] for better handling of class imbalance:
$$L = -\alpha_t (1 - \hat{y}_i)^\gamma \log(\hat{y}_i) \tag{1}$$
Class weights were used to account for class imbalances in the training dataset.
- **Optimizer:** Employed *Adam* for adaptive learning rate adjustments.
- **Reference Metric:** Evaluated model performance using *Mean Intersection over Union (Mean IoU)*:
$$\text{Mean IoU} = \frac{1}{C} \sum_{i=1}^{C} \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i + \text{FN}_i} \tag{2}$$

4 Experiments

We conducted a series of experiments focusing on different *network architectures*, *preprocessing steps*, and *loss functions*.

4.1 Network Architectures

We explored several configurations of the U-Net architecture:

- **Custom U-Net**: We experimented with different numbers of blocks, specifically using 4 and 5 blocks.
- **Library U-Net**: We exploited the **segmentation-models** [2] library to experiment with different backbone encoders: *ResNet Encoder*, *MobileNet Encoder* and *EfficientNetB0 Encoder*.
- **Custom U-net++**: we replicated the implementation of Z. Zhou et al [5]

4.2 Preprocessing Steps

We experimented with different configurations of data augmentation:

- **Albumentation-only**: we experimented with different transformations and ordering of them, always performing online-augmentation.
- **Dual-phase augmentation**: we tried to augment the dataset size before performing training using a custom function, combining it with the preexisting Albumentation pipeline.

4.3 Loss Functions

Multiple loss functions were evaluated to find the optimal combination. We experimented with combinations of different losses, including *Dice Loss*, *Categorical Crossentropy*, *Boundary Loss*, *Sparse Categorical Cross Entropy*, and *Sparse Categorical Focal Loss*.

Table 1: The metrics obtained on the local validation dataset

Loss-function	test MIoU	val MIoU
SparseCategoricalFocalLoss	**0.6923**	**0.6884**
Custom loss combination	0.4821	0.4771

5 Results

Our experiments led to several findings:

- The custom U-Net with 5 blocks demonstrated thebest performing, with a Mean IoU significantly improving upon other configurations.
- The dual-phase augmentation strategy yielded the best performance in terms of enhancing model generalization.
- Employing Sparse Categorical Focal Loss alone was most effective, leading the enhancement in the model MIoU score.

6 Discussion

The model's superior performance is driven by:

- **Network Depth:** The 5-block U-Net effectively captures complex terrain features.
- **Augmentation Techniques:** Enhanced data augmentation prevents overfitting and adapts the model for diverse Martian environments.
- **Loss Function:** Sparse Categorical Focal Loss addresses class imbalance, improving segmentation accuracy across varied terrain.

7 Conclusions

In conclusion, this study demonstrates that tailored U-Net architectures and sophisticated preprocessing techniques are highly effective for segmenting Mars terrain images. Our deployment of a custom 5-block U-Net, strategic data augmentation, and a focused loss function optimization sets a robust foundation for future advancements in planetary surface analysis using machine learning. This approach underscores the potential for further academic exploration and practical applications in interplanetary studies.

References

[1] E. K. V. I. I. A. Buslaev, A. Parinov and A. A. Kalinin. Albumentations: fast and flexible image augmentations. *ArXiv e-prints*, 2018.

[2] P. Iakubovskii. Segmentation models. `https://github.com/qubvel/segmentation_models`, 2019.

[3] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár. Focal loss for dense object detection, 2018.

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

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