AN2DL - Second Homework Report
NeuralDropouts

Pinar Erbil, Sergio Pardo, Angela Remolina, Matteo Sissa

perbil, sergiopardo, angelaremolina, matteosissa

244638, 243066, 242814, 247064

December 14, 2024

1 Introduction

Computer vision has become a mainstream technology used by a lot of systems in our world, from security surveillance, to medical image analysis. Computer vision systems are also implemented on robots exploring other planets, such as Mars. This aligns with the goal of this project, which is to assign the correct class label to each image pixel, selecting among four classes, each representing a particular type of terrain on Mars.

2 Problem Analysis

Classifying terrain images is a challenging problem, as the differences between terrains can be very subtle. The analysis started from the provided dataset, which consisted of a training set of 2615 64x128 gray scale images, and, 10022 images for testing. Each image in the training set had a corresponding mask, with each pixel labeled with a class.

3 Method

3.1 Data wrangling

The dataset was inspected using **Principal Component Analysis (PCA)** [3] to identify outliers. In this analysis, first evident noise samples were removed and then others located further from the distribution. These samples include the masks resembling alien spaceships and those where the majority of the pixels were assigned to the background class.

After the initial cleaning, DBSCAN [2] was used. However, only 20 samples were classified as outliers, which are the images where the robot or its shadow obstructed the view. Despite this, some of these images contained valuable information, so a decision was made to retain them and preserve the first distribution.

3.2 Data Augmentation

After the cleaning, this was the initial distribution of the classes:

Class 0	Class 1	Class 2	Class 3	Class 4
20.52%	35.60%	24.44%	19.30%	0.14%

Evidently, the dataset was highly unbalanced especially for class 4. To mitigate this problem, various data augmentation techniques were tried out, targeting specifically class 4 samples.

3.2.1 Augmenting only class 4 - Version 1

The process began by identifying a bounding box around the class 4 pixels in all the images where class 4 appeared, ensuring it met a minimum size. If the patch had sufficient coverage of class 4, it was cropped from the image and mask. The extracted region was then resized using interpolation to fit the target dimensions, maintaining the class distribution. The newly augmented class 4 samples were added to the dataset.

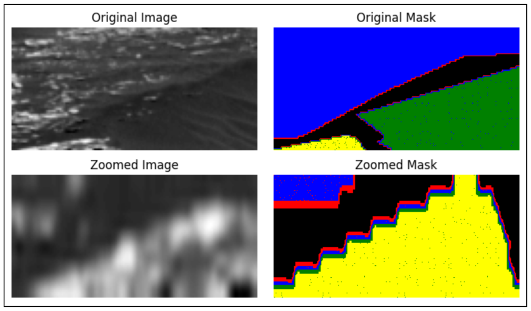

Figure 1: Cropping and zooming augmentation

3.2.2 Augmenting only class 4 - Version 2

This version of augmentation for class 4 starts with resizing and cropping the images and masks, just as in Version 1 3.2.1. This time the extracted patches are tiled together to create larger images that match the original size.

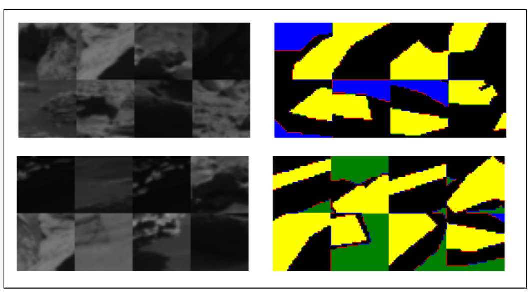

Figure 2: Tiling augmentation

3.2.3 Augmenting only class 4 - Version 3

Since the samples extracted with the previous techniques seemed insufficient, light geometric and intensity transformations were applied to the images extracted for class 4 to increase diversity and representation coverage.

3.3 Model Design

Three model architectures were employed throughout this project, each one inspired from some state-of-the-art model for semantic segmentation problems. The first trials were made with a **U-Net** architecture [6], which seemed the most reasonable choice to start experimenting. When the unbalanced nature of the dataset was revealed as the most important challenge to face in this project, other architectures were devised. First, **Attention U-Net** [5] seemed to be a good solution as it is an extension of the original U-Net model, integrating an attention mechanism to focus on the most relevant regions of the input images during the skip connections. Second, **Dense U-Net** [1] was tested. The dense connectivity of this model and its improved feature propagation capabilities seemed more applicable to the domain of the problem.

3.4 Loss Function Design

Three different types of loss functions were tested, to assess the response of the model and cope with the imbalanced representation of class 4. The first and most straightforward approach was to apply **weighted categorical cross entropy**, which aims to focus more attention on the underrepresented classes by means of weights associated to each class. Then **dice loss**[4] was implemented. It is derived from the Dice coefficient, which measures the overlap between predicted and ground truth region with an emphasis on small and underrepresented regions. Lastly **focal loss**[7] is used. It ends up being the best fit for our problem because it is designed to address class imbalance.

4 Experiments

Early experiments: The experimentation started with the simplest model design and no modifications on the dataset to assess the complexity of the task and the baseline. A U-Net model was trained on the original dataset with sparse categorial cross-entropy and Dice loss functions. This baseline turned out to be quite good, with almost 52% accuracy on the test set.

Mid-stage experiments: After cleaning the data set, data augmentation was tried out on the entire dataset, but there was no big improvement. Therefore, a more in-depth data inspection process started. This analysis showed that class 4 was severely underrepresented and was hindering the learning process of the model drastically. It also showed that the background was predominant in the dataset, leading to confusion in the learning process.

To cope with these, augmentation techniques were designed specifically for class 4. Additionally, the weight associated to the background class was set to 0, to force the model to stop learning background segmentation. The results coming from these changes in the approach were promising, leading to a mean IoU of 67% over the test set. This was achieved with Dense U-Net, and the dataset enhanced for class 4 representation balancing.

Late experiments: At this point minor adjustments were done to every strategy adopted throughout the process; removing some extra outliers or data samples for balancing the dataset; trying different types of augmentation for class 4 and combining them; playing with the weights associated to the classes; and trying different combinations of models and loss functions. Not every combination of these techniques achieved great results, but throughout this period of trials multiple attempts scored high, between 68% and 70% mean IoU (over the test set), leading to the highest results for the project.

5 Results

Model	Augmentation	Loss Function	Local Mean IOU	Kaggle Result
U-Net	Unclean data No augmentation	Dice Loss	0.5094	0.5191
Attention U-Net	Clean Data Only Class 4 Version1	Weighted Categorical Crossentropy	0.5560	0.6654
Dense U-Net	Clean Data Only Class 4 Version2	Focal Loss	0.663	**0.7055**
Dense U-Net	Clean Only Class 4 Version3	Weighted Categorical Crossentropy	**0.6525**	0.6854

Table 1: Some of the results. All the models, except the first one, sets background weight zero.

The most significant results are presented in Table 1. All the detailed results are in the Appendix 9 Table 2. The best result is achieved by Dense U-Net using only class 4 augmentation (version 2), focal loss and setting background weights to 0 with the Mean Intersection over Union (Mean IoU) score of **70.55%** on Kaggle test data. Overall, considering the other augmentation versions, **Dense U-Net** performs better than Attention U-Net. Furthermore, the models perform better with **focal loss** compared to dice loss and weighted categorical crossentropy. Focal loss is mathematically designed to discourage the model to keep learning on classes that already have high accuracy. As a consequence of this, it managed to drive the model's attention to classify correctly class 4 pixels, even if they were the minority of the original dataset.

6 Discussion

The final results obtained showed that it is not always necessary to have a really large and correctly labeled dataset in order to achieve high scores and metrics for a semantic segmentation problem. This is possible thanks to fully convolutional neural networks such as U-Net. Despite all of this, a well-designed dataset, with a larger amount of samples and correctly labeled masks would have most likely improved the prediction power of the model.

7 Conclusions

In this assignment, we worked on a segmentation problem with images from the Mars terrain. Having a highly unbalanced dataset, most of the effort was put into augmentation. Three types of augmentation, three different model architectures, and three different types of loss functions were implemented. Out of all these combinations, the best results were achieved with the augmentation techniques applied for class 4 and the Dense U-Net architecture with focal loss, scoring a mean IoU of 70.5% over the test set.

In future work, the given dataset maybe analyzed further, since the correctness of the given masks was constantly questioned and was a recurrent problem throughout the process. It is also possible to further investigate possible ensemble models and hybrid loss functions.

8 Team contributions

Sergio focused on data wrangling and implementing initial data analysis distribution. **Pinar** worked on the building the base U-Net, dense U-Net and ensemble models. **Angela** contributed by building the attention U-Net approach and class balancing functions. The augmentation techniques for class 4 for class balancing were created by **Matteo**, who also analyzed the performance of different loss functions.

References

[1] S. Cai, Y. Tian, H. Lui, H. Zeng, Y. Wu, and G. Chen. Dense-unet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. *Quantitative imaging in medicine and surgery*, 10(6):1275, 2020.

[2] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231, 1996.

[3] I. T. Jolliffe and J. Cadima. *Principal Component Analysis*. Springer, 2016.

[4] X. Li, X. Sun, Y. Meng, J. Liang, F. Wu, and J. Li. Dice loss for data-imbalanced nlp tasks. *arXiv preprint arXiv:1911.02855*, 2019.

[5] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, et al. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*, 2018.

[6] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[7] 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.

9 Apendix

Model	Data	Augmentation	Loss Function	Class Weights	LMIOU	KR
U-Net	Unclean	No	Sparse Categorical Crossentropy	N/A	0.4176	0.4299
Dense U-Net	Unclean	No	Dice Loss	N/A	0.4829	0.47964
U-Net	Unclean	No	Dice Loss	N/A	0.5094	**0.5191**
U-Net	Clean	No	Dice Loss	BG weight = 0	0.5609	0.4899
U-Net	Clean	No	Focal Loss	BG weight = 0	**0.5166**	0.4834
AU-Net	Clean	All classes	Dice Loss	N/A	0.5086	0.5002
AU-Net	Clean	All classes	Dice Loss	BG weight = 0	0.4270	0.4166
AU-Net	Clean	Only Class 4 (V1)	WCC	BG weight = 0	0.5560	0.6654
AU-Net	Clean	Only Class 4 (V1)	WCC	Normalized BG weight = 0	0.5405	0.6190
Dense U-Net	Clean	Only Class 4 (V1)	WCC	BG weight = 0	**0.57**	**0.6753**
U-Net	Clean	Only Class 4 (V2)	WCC	BG weight = 0	0.5868	0.5690
AU-Net	Clean*	Only Class 4 (V2)	WCC	All balanced BG weight = 0	0.5163	0.5811
AU-Net	Clean	Only Class 4 (V2)	WCC	BG weight = 0	0.5943	0.6019
Dense U-Net	Clean	Only Class 4 (V2)	WCC	BG weight = 0	0.6602	0.6876
Dense U-Net	Clean	Only Class 4 (V2)	WCC	Normal weights	**0.7821**	0.5224
Ensemble: BG AU-Net Dense U-Net	Clean	Only Class 4 (V2)	Binary Crossentropy WCC	Binary weights BG weight = 0	0.515 0.6602	0.6315
Dense U-Net	Clean	Only Class 4 (V2)	Dice Loss	N/A	0.768	0.5354
Dense U-Net	Clean	Only Class 4 (V2)	Focal Loss	BG weight = 0	0.663	**0.7055**
Dense U-Net	Clean	Only Class 4 (V2)	Focal Loss	BG weight = 0.1	0.6944	0.5580
Dense U-Net	Clean	Only Class 4 (V3)	WCC	BG weight = 0	0.6525	**0.6854**
AU-Net	Clean	Only Class 4 (V3)	WCC	BG weight = 0	0.6374	0.6221
Dense U-Net	Clean**	Only Class 4 (V3)	WCC	Normal weights	**0.8264**	0.5253
Dense U-Net	Clean	Only Class 4 (V3)	Focal Loss	BG weight = 0	0.6535	0.6902

Table 2: Results of all the models. Red ones indicate a possible overfit and green is the best result(**LMIOU**: Local Mean IOU, **KR**: Kaggle Result, **BG**: Background, **AU-Net:** Attention U-Net, **WCC**: Weighted Categorical Crossentropy, ***:** Removed some data for balancing, ****:** Remove Outliers)