AN2DL - Second Homework Report
Vuoilaglio

Ricardo André Araújo De Matos, Melvin Curinier, Pedro Fonseca, Benjamin Mallefait

ricardomatos11076009, mcurinier, polimipfonseca, bmallefait

276544, 276717, 276962, 276918

December 14, 2024

1 Introduction

This project addresses *semantic segmentation* on 64x128 grayscale images of Martian terrain, aiming to classify each pixel into specific terrain classes. Unlike typical segmentation challenges, pretrained models are prohibited, necessitating training from scratch.

The primary objectives are to **preprocess data**, **design models** for accurate segmentation, and **evaluate solutions** to maximize *mean Intersection over Union* exluding the background class. This setup provides an opportunity to explore the impact of architecture design choices and data augmentation on model performance.

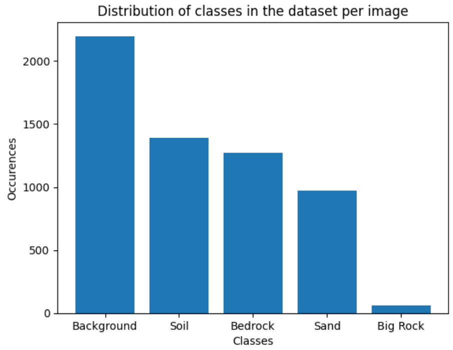

Figure 1: Class distribution per image

2 Problem Analysis

The dataset, `mars_for_students.npz`, includes training images, segmentation masks, and test images. The challenges include **data scarcity**, with only 2,615 labeled training images compared to 10,022 test images; **small dimensions**, where the 64x128 size limits terrain detail; and **imbalanced labels**, with the background dominating the dataset (Figure 1). Some corrupted data are also introduced so a cleaning is required.

3 Method

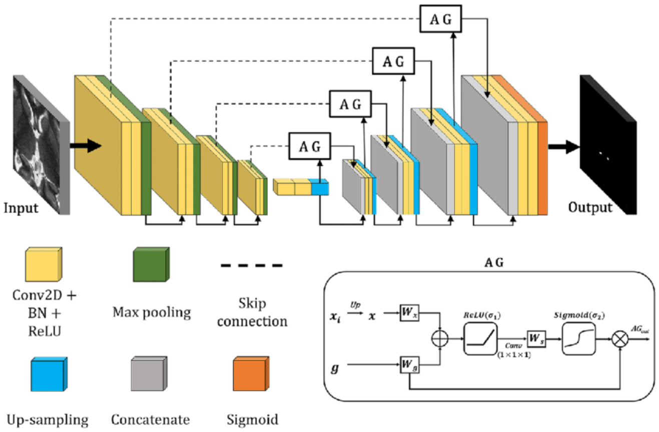

Figure 2: The architecture of the attention gate-based u-net [8]

3.1 Attention ResUNet

One of our best-performing models was the Attention ResUNet, achieving a MeanIoU of 61.666% on the Kaggle leaderboard. This architecture, a variant of the U-Net, is designed specifically for image segmentation tasks and incorporates **residual connections** and **attention mechanisms** [4].

Overall Structure The Attention ResUNet follows the general U-Net structure with an encoder (downsampling) path and a decoder (upsampling) path, enhanced by **residual connections** and **attention gates**. The overall architecture is displayed in Figure 2

Key Components

- **Residual Convolution Block:** Includes two convolutional layers with *batch normalization* and *ReLU* activation.
- **Attention Mechanism:** Computes attention coefficients between encoder features and decoder signals, refining encoder features for concatenation.
- **Final Layers:** A 1x1 convolution maps features to five terrain classes, followed by *batch normalization* and *sigmoid/softmax* activation.

3.2 MultiResUNet

Our second best-performing model, achieving 65.269% accuracy on the Kaggle leaderboard, utilized an improved UNet network. The MultiResUNet [6] enhances the U-Net by addressing limitations in feature extraction and spatial detail retention.

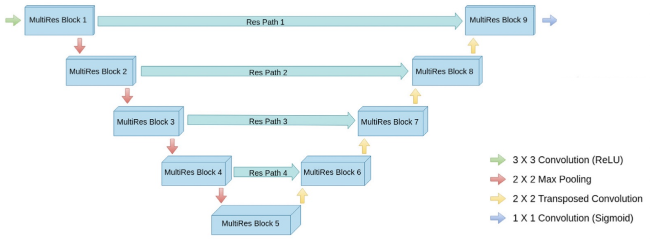

Figure 3: MultiResUNet architecture with *Res Path* and *MultiRes Block*

Overall Structure *MultiRes Blocks*, which use parallel convolutions of varying kernel sizes (e.g., 3x3, 5x5, 7x7) to capture multi-scale features, combined efficiently with pointwise convolutions (1x1), and *Residual Pathways*, inspired by ResNet, to ensure smooth gradient flow and better training stability. The overall architecture is displayed in Figure 3.

Advantages

- Combines features at multiple scales for robust segmentation.
- Residual pathways ensure efficient training and deeper networks.
- Offers improved accuracy over traditional U-Net on complex datasets.

3.3 MultiResUNet with Attention

The MultiResUNet with Attention model achieved 68.334% accuracy on the Kaggle leaderboard, improving the MultiResUNet by adding an attention mechanism.

Overall Structure The architecture follows MultiResUNet but incorporates attention gates that refine features from the encoder and decoder, focusing on important image regions.

3.4 Loss Function

Sparse Categorical Cross-Entropy that does not take into account the class 0. Another loss functions were tested, such as *Dice Loss*, but the former yield the best results.

3.5 Training

The training uses the Adam optimizer with an exponential decay learning rate, early stopping to prevent overfitting, and a custom Mean Intersection Over Union (MIoU) metric for segmentation accuracy.

3.6 Data Preprocessing

To prepare the data for our model, we performed the following preprocessing steps:

3.6.1 Data Cleaning & Splitting

After removing the fake alien images, we separated the training data into:

- **Train Dataset:** 2505 images/masks used for training.
- **Validation Dataset:** 300 images/masks for validation.

3.6.2 Data Augmentation

To tackle data scarcity and enhance generalization, we employed the `Albumentations` library to apply various augmentations, including *random flips*, *brightness* and *contrast adjustments*, *geometric transformations* (shift, scale, rotate), *distortions* (optical and grid), *grid shuffling*, *blurs* (Gaussian and motion), *and coarse dropout.*

Additionally, **CutMix** was applied to both input images and masks, focusing on underrepresented classes like ”big rock” to have a more balanced dataset, as shown in Figure 4.

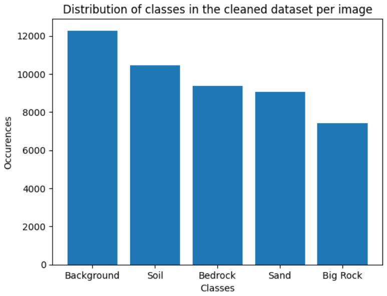

Figure 4: Class distribution per image after augmentation

4 Experiments

4.1 Performance on Kaggle

MultiResUNet with attention achieved **68.344%** accuracy, outperforming other models.

Table 1: Model Performance Comparison

Model	MIoU Kaggle	MIoU Local
SegMarsViT [3]	57.03	57.57
Attention ResUNet	61.666	63.06
MultiResUNet	**65.269**	61.84
Att.-MultiResUNet	**68.344**	69.12

4.2 Ablation Study

An ablation study highlighted:

- **Attention Gates:** Boosted MeanIoU by ∼3%.
- **CutMix Augmentation:** Enhanced class-wise IoU for *big rock.*
- **Background Loss:** By removing the background from the loss function calculations, IoU improved drastically by ∼10%

4.3 Generalization and Robustness

Testing with noisy and augmented datasets showed *MultiResUNet with attention* maintained high MeanIoU, with minimal performance degradation. The model also demonstrated robustness to minor label inconsistencies, ensuring better generalization.

5 Discussion

MultiResUNet with Attention proved effective, though challenges remain:

- **Class Imbalance:** Augmentation improved but did not fully resolve issues.
- **Image Resolution:** Higher resolution could enhance detail capture.

Future work could explore super-resolution techniques and advanced augmentation techniques such as *Test-Time augmentation.* Additionally, incorporating transformer-based architectures could further refine segmentation performance.

6 Conclusions

The Attention ResUNet achieved 61.666%, while the MultiResUNet reached 65.269%. Adding Attention to the latter improved our score to 68.344%. Incorporating attention mechanisms boosted segmentation performance. Further optimization could address class imbalance and image resolution limitations. The insights from this project highlight the potential of combining advanced architectures with robust preprocessing techniques.

Everyone contributed equally.

References

[1] Oleg Belkovskiy. Enhancing unet: Tailoring superior segmentation models through transfer learning, 2023. Available at: https://medium.com/@oleg.belkovskiy/enhancing-unet-tailoring-superior-segmentation-models-through-transfer-learning-8323a519b877.

[2] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *Google Research*, December 2017. Contact: lcchen, gpapan, fschroff, hadam@google.com.

[3] Yuqi Dai, Tie Zheng, Changbin Xue, and Li Zhou. Segmarsvit: Lightweight mars terrain segmentation network for autonomous driving in planetary exploration. *Remote Sensing*, 2022.

[4] DigitalSreeni. Attention resunet, 2024. Available at: https://www.youtube.com/watch?si=cbcrG_NP640vIj8-&v=KOF38xAvo8I&feature=youtu.be.

[5] Edwin Goh, Jingdao Chen, and Brian Wilson. Mars terrain segmentation with less labels. In *Proceedings of Jet Propulsion Laboratory*, Jet Propulsion Laboratory, USA, February 2022. California Institute of Technology. Contact: edwin.y.goh@jpl.nasa.gov, chenjingdao@cse.msstate.edu, bdwilson@jpl.nasa.gov.

[6] Nabil Ibtehaz and M. Sohel Rahman. Multiresunet: Rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural Networks*, 2019. Available at: https://www.sciencedirect.com/science/article/abs/pii/S0893608019302503.

[7] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[8] Sang-Heon Lim, Jihyun Yoon, Young Jae Kim, Chang-Ki Kang, Seo-Eun Cho, Kwanggi Kim, and Seung-Gul Kang. Reproducibility of automated habenula segmentation via deep learning in major depressive disorder and normal controls with 7 tesla mri. *Scientific Reports*, 11:13445, 06 2021.