7th IEEE International Conference on Recent Advances and Innovations in Engineering
7th IEEE ICRAIE 2022

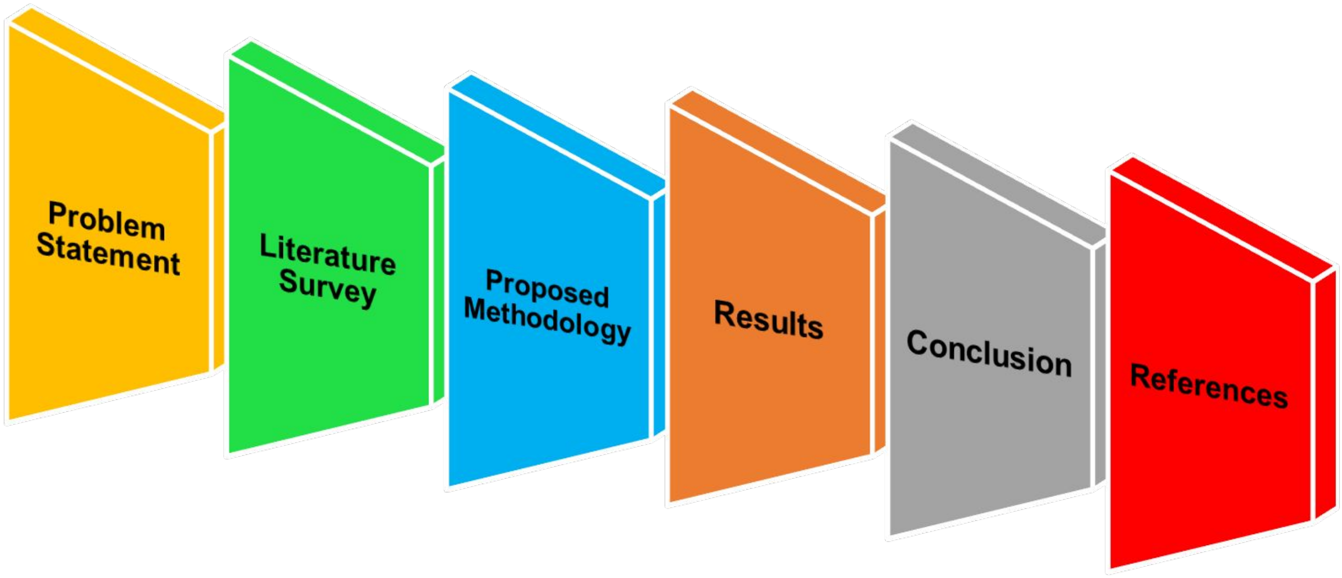
Semantic Segmentation on Martian Terrain for Navigation using Transformers

Gagan Goutham*, Hita Juneja*, Ankitha C*, V R Badri Prasad*

PES University, Bangalore

gagangoutham.n03@gmail.com hitajuneja@gmail.com ankithac45@gmail.com badriprasad@pes.edu

Agenda



gagangoutham.n03@gmail.com hitajuneja@gmail.com ankithac45@gmail.com badriprasad@pes.edu

Problem Statement

Sandy Slumps
Sharp Rocky
Surface

Many methodologies have been implemented.

No approaches using **Vision Transformers (ViT)**

An **Autonomous Navigation System** focusing on the **Martian terrain**.

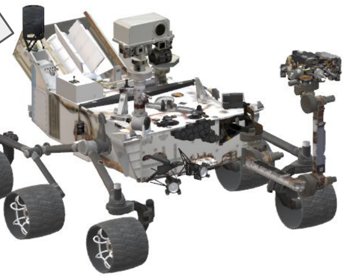
SOJOURNER

OPPORTUNITY

SPIRIT

CURIOSITY

PERSEVERANCE



“An alternative approach using the images captured by Rovers to Semantically Segment the Martian terrain for Navigation”

 Gagan Goutham, Hita Juneja, Ankitha C, V R Badri Prasad |
gagangoutham.n03@gmail.com hitajuneja@gmail.com ankithac45@gmail.com badriprasad@pes.edu

Literature Survey

- With a focus on geological relevance rather than navigation, the majority of related research studies place an emphasis on categorising and classifying images of Martian topography from Mars imagery.
- On the rover images, experiments using Convolutional Neural Network, Support Vector Machines and K-means Clustering models have been conducted as seen in the past.
- AI4Mars data trained on DeepLabv3 model with labelling of images from Curiosity, Opportunity, and Spirit rovers, collected through crowdsourcing.
- SPOC employed Deep CNN on orbital and ground based imageries with terrain types of soil, rock and terrain features of scarp, ridges
- This work distinguishes itself from others in that it employs semantic segmentation using the **SegFormer** model.

gagangoutham.n03@gmail.com hitajuneja@gmail.com ankithac45@gmail.com badriprasad@pes.edu

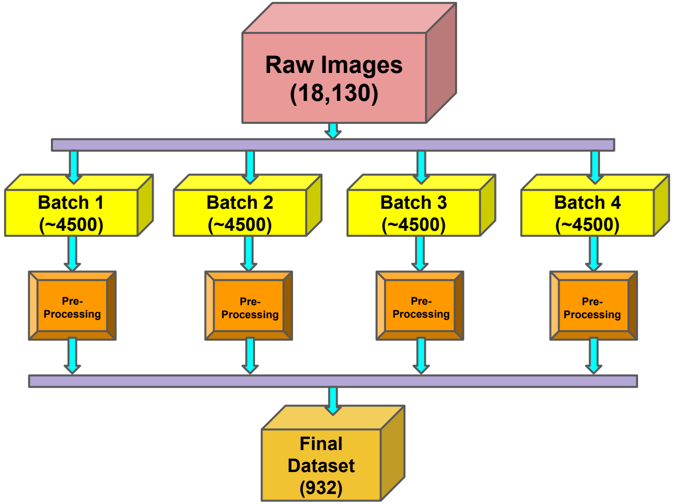
Proposed Methodology

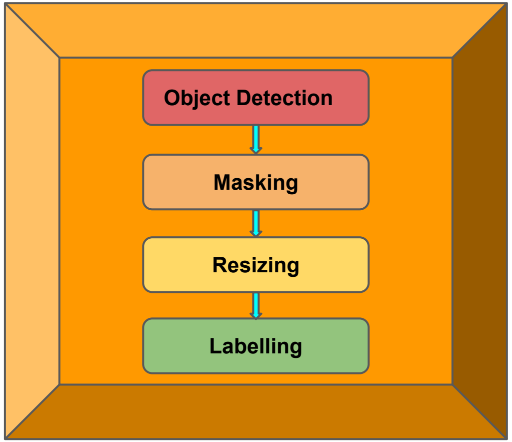
Dataset

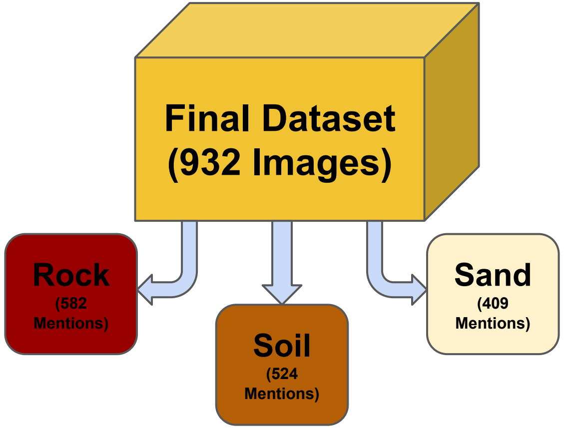
- The **AI4Mars** dataset includes majority of the existing high-resolution martian surface images.
- Comprises of ~35K MSL images sourced from:
  - Curiosity, Opportunity, and Spirit rovers.
  - Grayscale navigation camera (NAVCAM)
  - Color mast camera (Mastcam)
- Objective of AI4Mars is to train deep neural networks for sustainable self-driving on Mars.
  - Help future Martian rovers navigate the planet's surface.
- From MSL NAVCAM data, **18,130 raw images** are used for this problem statement.

gagangoutham.n03@gmail.com hitajuneja@gmail.com ankithac45@gmail.com badriprasad@pes.edu

Pre-Processing

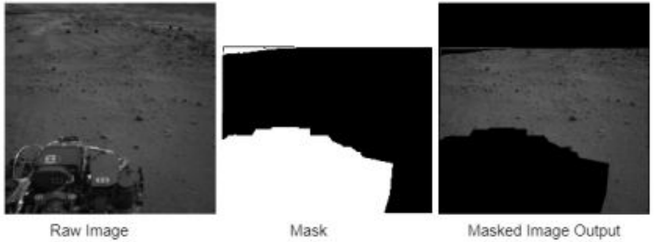






gagangoutham.n03@gmail.com hitajuneja@gmail.com ankithac45@gmail.com badriprasad@pes.edu

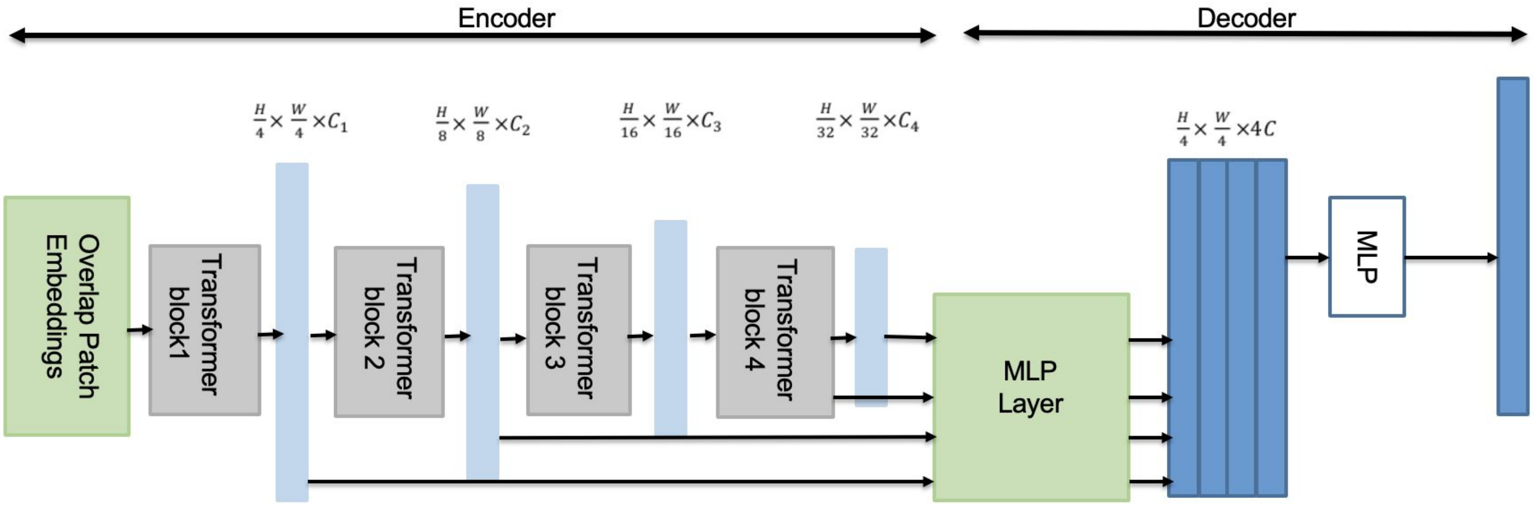
- **Object Detection:**
  - Area of the greatest bounding box formed around the detected rover > ¼(image size) -> **Image Rejected** from batch
  - Area of the greatest bounding box formed around the detected rover < ¼(image size) Or No rover found -> **Image retained** in the batch
  - **10,041 images** were retained from original 18,130 images
- **Masking:**
  - **Unwanted regions** were masked - sky / rover components
- **Resizing:**
  - Each image was downsized to **512x512** from original 1024x1024
- **Labeling:**
  - Each image was labeled into **3 classes - Rock, Soil, Sand**
  - Final Dataset with **932 Images**



gagangoutham.n03@gmail.com hitajuneja@gmail.com ankithac45@gmail.com badriprasad@pes.edu

Feature Extraction

- Hierarchical **Transformer encoder** to extract coarse and fine features
- **Lightweight MLP decoder** to directly fuse these multi-level features and predict the semantic segmentation mask.



gagangoutham.n03@gmail.com hitajuneja@gmail.com ankithac45@gmail.com badriprasad@pes.edu

Model

SegFormer:
Transformer Encoder +
Multilayer Perceptron Decoder

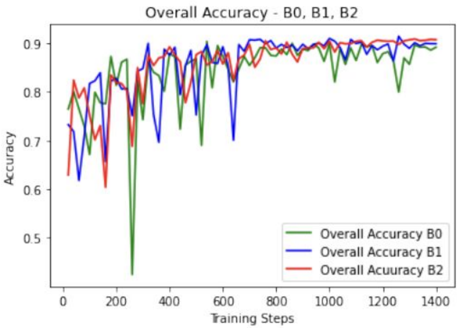
Training Set - 746 images
Testing Set - 186 images

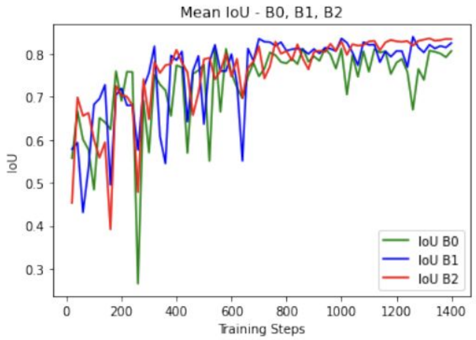
Batch Size - 8
Learning Rate - 0.0006
Epochs - 15

Model variant	Depths	Hidden sizes	Decoder hidden size	Params (M)	ImageNet-1k Top 1
MiT-b0	[2, 2, 2, 2]	[32, 64, 160, 256]	256	3.7	70.5
MiT-b1	[2, 2, 2, 2]	[64, 128, 320, 512]	256	14.0	78.7
MiT-b2	[3, 4, 6, 3]	[64, 128, 320, 512]	768	25.4	81.6

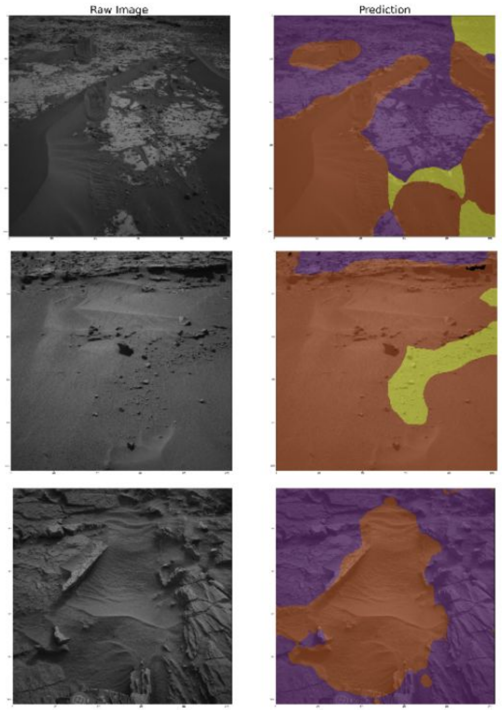
gagangoutham.n03@gmail.com hitajuneja@gmail.com ankithac45@gmail.com badriprasad@pes.edu

Results





Model	Mean IoU	Mean Accuracy	Overall Accuracy
SegFormer - B0	81.15%	88.61%	89.92%
SegFormer - B1	83.02%	90.33%	90.66%
SegFormer - B2	83.55%	90.75%	90.86%



Ground truth vs Prediction

gagangoutham.n03@gmail.com hitajuneja@gmail.com ankithac45@gmail.com badriprasad@pes.edu

Conclusion

- Highlighting the importance of Transformers instead of employing Convolutional Neural Networks, autonomous rovers can navigate by using MultiLabel Image Segmentation for the landscape of Mars.
- It can be concluded that the segmentation process was accomplished successfully by all the three models, with SegFormer - B2 achieving the best accuracy of 90.86%.
- An upcoming scope of this project is to apply image segmentation to film captured by the rover.
- Overall, this research emphasizes the significance of autonomous navigation for rovers and presents a new method to implement for the challenging martian terrain.

gagangoutham.n03@gmail.com hitajuneja@gmail.com ankithac45@gmail.com badriprasad@pes.edu

References

[1] "AI4MARS: A dataset for terrain-aware autonomous driving on Mars." ,Swan, R. Michael, Deegan Ath et.al., In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1982-1991. 2021.
[2] "MRSCAtt: A spatio-channel attention-guided network for mars rover image classification.", Chakravarthy, Anirudh S et.al., In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1961-1970. 2021.
[3] "Combining support vector machines and information gain ranking for classification of mars mcmurdo panorama images.", Shang, Changjing et.al., In 2010 IEEE International Conference on Image Processing, pp. 1061-1064. IEEE, 2010.
[4] "Spoc: Deep learning-based terrain classification for mars rover missions.",Rothrock, Brandon et.al., AIAA SPACE 2016 (2016): 5539.
[5] "A simple framework for contrastive learning of visual representations.", Chen, Ting et.al., In International conference on machine learning, pp. 1597-1607. PMLR, 2020.
[6] DeepLabV3+/Efficientnet Hybrid Network-Based Scene Area Judgment for the Mars Unmanned Vehicle System.", Hu, Shuang et.al., Sensors 21, no. 23 (2021): 8136.
[7] "An image is worth 16x16 words: Transformers for image recognition at scale.", Dosovitskiy, Alexey, Lucas Beyer et.al., arXiv preprint arXiv:2010.11929 (2020).
[8] "Deep Mars: CNN classification of mars imagery for the PDS imaging atlas." , Wagstaff, Kiri et.al., In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, no. 1. 2018.
[9] "Fuzzy-rough feature selection aided support vector machines for mars image classification.", Shang, Changjing et.al., Computer Vision and Image Understanding 117, no. 3 (2013): 202-213.
[10] "SemiSupervised Learning for Mars Imagery Classification.", Wang, Wenjing et.al., In 2021 IEEE International Conference on Image Processing (ICIP), pp. 499-503. IEEE, 2021.
[11] "Self-Supervised Learning to Guide Scientifically Relevant Categorization of Martian Terrain Images." , Panambur, Tejas et.al., In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1322-1332. 2022.
[12] "Classification of martian terrains via deep clustering of mastcam images.", Parente, Mario, and Tejas Panambur, In IGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium, pp. 1054- 1057. IEEE, 2020.
[13] "Improved deep clustering of mastcam images using metric learning.", Panambur, Tejas, and Mario Parente, In 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 2859-2862. IEEE, 2021.
[14] "Mars Terrain Segmentation with Less Labels.", Goh, Edwin, Jingdao Chen, and Brian Wilson, arXiv preprint arXiv:2202.00791 (2022).
[15] "Visualizing and understanding convolutional networks." , Zeiler, Matthew D., and Rob Fergus, In European conference on computer vision, pp. 818-833. Springer, Cham, 2014.
[16] "Deep residual learning for image recognition.", He, Kaiming et.al., In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770-778. 2016.
[17] "Rich feature hierarchies for accurate object detection and semantic segmentation.", Girshick, Ross et.al., In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587. 2014.
[18] "Convex hulls in image processing: a scoping review.", Jayaram, M. A., and Hasan Fleyeh, American Journal of Intelligent Systems 6, no. 2 (2016): 48-58.
[19] "SegFormer: Simple and efficient design for semantic segmentation with transformers.", Xie, Enze et.al., Advances in Neural Information Processing Systems 34 (2021): 12077-12090.
[20]"Satellite image classification and segmentation with transfer learning.", Giorgiani do Nascimento, Renato, and Felipe Viana., *In* AIAA Scitech 2020 Forum, p. 1864. 2020.

gagangoutham.n03@gmail.com hitajuneja@gmail.com ankithac45@gmail.com badriprasad@pes.edu