OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Using State-of-the-Art Artificial Intelligence Detection for Retrieval of Sample Tubes on Mars

Leland Barbour
lelandbarbour1@gmail.com

ABSTRACT

The Mars Sample Return Mission, a collaborative effort by NASA and the European Space Agency (ESA), aims to return Martian soil and rock samples to Earth for the first time. A key aspect of the mission involves retrieving sample tubes dropped by the Perseverance rover on the Martian surface, which are critical for scientific analysis. However, the harsh and dynamic Martian environment presents challenges in locating these tubes, as they can be partially buried or obscured by dust and debris. This paper investigates the use of artificial intelligence (AI) to enhance the detection and retrieval of these sample tubes. Specifically, we explore the effectiveness of two AI models: the Segment Anything Model (SAM) and its language-augmented variant, LangSAM, for detecting the sample tubes in images. SAM demonstrated good performance on the Earth-based training data, but struggled in real Martian conditions. In contrast, LangSAM, augmented with natural language prompts, showed superior accuracy, especially in scenarios where the tubes were obscured or partially buried. Our results highlight the potential of LangSAM to overcome the challenges of small datasets and visually complex environments, making it a promising tool for future Mars exploration missions.

INTRODUCTION

The Mars Sample Return Mission is a joint effort by NASA and the European Space Agency (ESA) to bring Martian soil and rock samples back to Earth for the first time. The mission aims to search for signs of past life, study Mars' geology, and prepare for future human exploration [1]. The campaign began with NASA's Perseverance rover, which is currently collecting samples on Mars (see Figure 1). A future Sample Retrieval Lander will pick them up and launch them into orbit using a Mars Ascent Vehicle. ESA's Earth Return Orbiter will then capture the sample container and return it to Earth. Planned for the 2030s, this complex mission is expected to provide groundbreaking scientific insights and pave the way for human missions to Mars [2].

The Perseverance rover is currently exploring the Jezero Crater on Mars, searching for signs of ancient life and collecting rock samples. It has been dropping sealed sample tubes onto the surface of Mars as part of its mission. So far, ten canisters have been placed on the ground, out of a total of 43 tubes carried by

Perseverance [3]. These sample tubes are being left on the surface to create a backup collection in the event that Perseverance cannot deliver its onboard tubes. A future mission would pick up the tubes and return them to Earth. The total cost of the Perseverance mission is around $2.7 billion [4], which highlights the importance of this backup strategy.

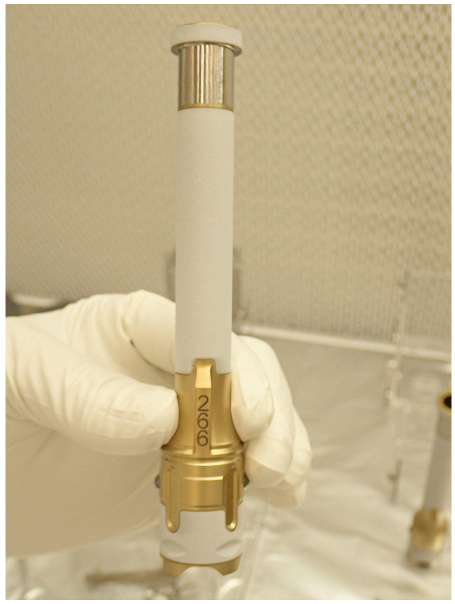

Figure 1. Mars tube before launch [3]

Figure 2. 3D printed tube used in this study

Although the tube coordinates are known, artificial intelligence would be especially helpful in finding the Martian sample tubes because the harsh and dusty Martian environment can partially bury or cover the tubes over time, making their precise location difficult to discern by the drive team via the robot camera. AI can process images from rover cameras and detect subtle shapes, shadows, or patterns that indicate a tube might be present, even if it is mostly hidden by sand or debris. By using machine learning trained on different visual scenarios, Artificial Intelligence increases the chances of identifying and locating the tubes quickly and accurately, making the recovery mission far more efficient. This paper aims to fill this gap.

AI OVERVIEW

Artificial intelligence faces significant challenges when working with small datasets. Past research has included developing tailored segmentation models for specific planetary science topics such as dust storms [5] and rocks [6]. Others have focused on making computational tools more efficient [7][8]. In this paper, two AI models were employed to identify sample tubes in Mars exploration missions: the Segment Anything Model (SAM), released by Meta, Facebook AI, in April 2023 [9], and LangSAM, an open-source model introduced later that year [10]. SAM was initially trained by its developers and fine-tuned for specific use cases, while LangSAM leverages zero-shot learning to perform tasks without additional user training. Both models aim to generate image masks that highlight relevant objects. These advanced transformer-based algorithms also integrate prompts alongside images to enhance performance. These approaches demonstrate the adaptability of modern AI in overcoming the limitations of small datasets.

DATASETS

This research project involved the creation of a custom dataset to support the recognition of Martian sample collection tubes. Due to the limited availability of real images of NASA's sample tubes on Mars, a life-size model of the tube was 3D-printed (see Figure 2), and images were captured under Earth-based conditions that mimic the Martian surface (see Figure 3). The 3D printed tube used real dimensions and was painted to try and match the real tube [3]. The dataset was divided into two subsets: 20 training images featuring the homemade model and simulated terrain, and 24 test images consisting of authentic photographs taken on Mars taken by the rover (see Figure 4 [11]). The ground-truth masks for each image were made using a point prompt tool in the SAM Github. The images were analyzed to extract key visual features such as tube shape, orientation, and environmental context. This dataset enabled the development and evaluation of machine learning models for tube detection in scenarios with limited real-world data.

Figure 3. 3D printed tube in dirt on Earth

Figure 4. Mars tube on Martian surface

MODELS

To advance the segmentation capabilities of computer vision models on extraterrestrial imagery, we fine-tuned the Segment Anything Model (SAM), originally developed by Meta in 2023, using custom training data. Our objective was to adapt SAM and its language-augmented variant (LangSAM) for precise object identification and segmentation, particularly in unfamiliar domains like Mars imagery. The experimental design involved hardware-intensive model training, architectural adaptation using Grounding DINO for language prompts, and testing on visually complex datasets.

SAM is a transformer-based vision model capable of segmentation tasks. Our version was fine-tuned using the 20 training images. The fine-tuning process employed dual NVIDIA L40S GPUs over a period of approximately three days. During this process, we applied a series of geometric and photometric transformations to improve model generalization and robustness, particularly for use cases involving Mars rover imagery.

The SAM architecture was leveraged for its responsive segmentation capabilities, where segmentation masks can be predicted based on point, box, or text-based prompts. The tuning process maintained the original model backbone while adjusting the final segmentation heads and prompt encoders to better align with our specific dataset. Figure 5 shows an example of an image with a mask around the frog and snail [9].

Figure 5. Example image showing segmentation outline of the frog and snail [9].

To enhance SAM's capabilities with natural language prompts, developers incorporated Grounding DINO—a vision-language grounding model that can detect objects based on text input. This integration resulted in the public version known as LangSAM. Our implementation of LangSAM is derived directly from SAM, augmented by language understanding capabilities to enable flexible prompt-based segmentation.

The LangSAM model was provided with a language prompt to guide segmentation tasks. For example, we augmented the model with the single-word prompt "tube" (and nothing else) to test the language-guided response and assess segmentation precision. Figure 6 shows the training and testing of SAM on the left and the training of LangSAM on the right in flowchart format.

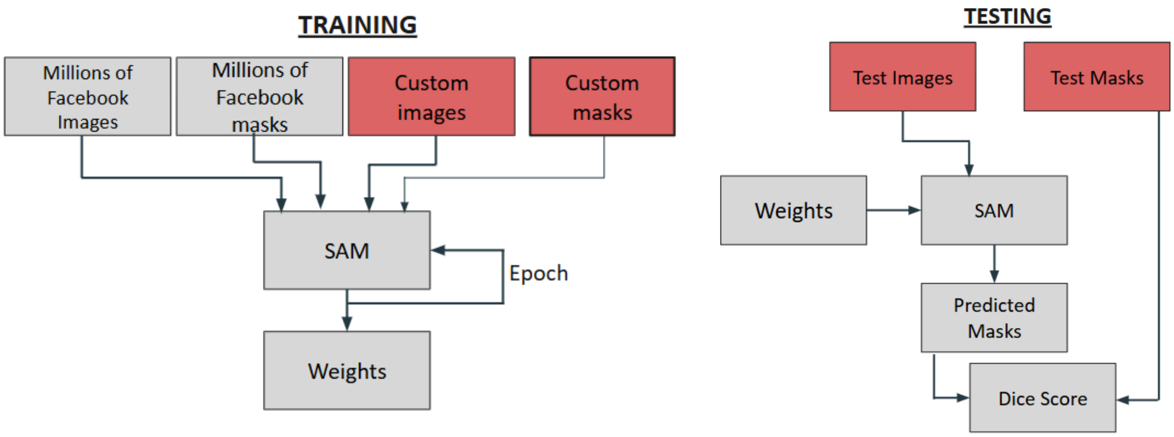

Figure 6. Simple guide showing how SAM works. Both training and testing run at the same time. The highlighted boxes were provided by the author.

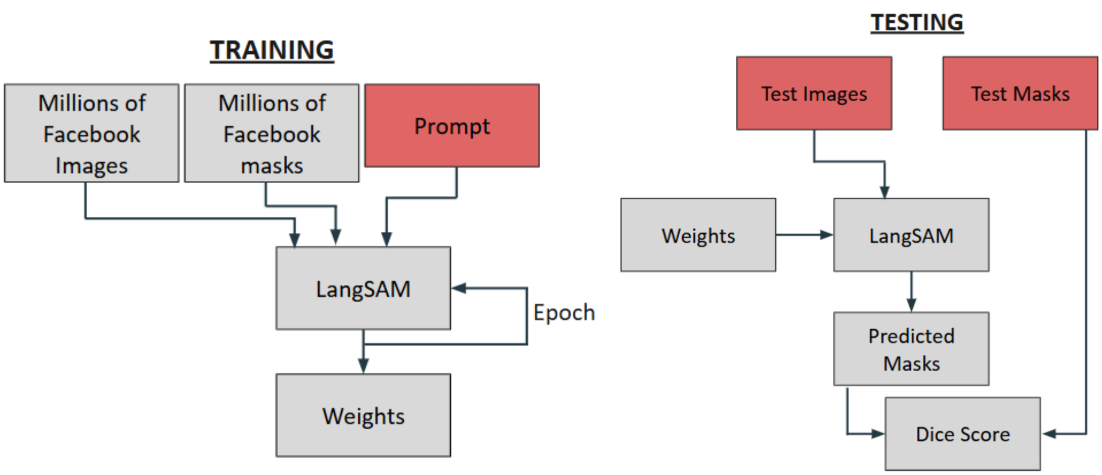

Figure 7. Simple guide showing how LangSAM works. Both training and testing run at the same time. The highlighted boxes were provided by the author, and LangSAM was fine-tuned by its creators.

RESULTS & DISCUSSION

Table 1 and Figure 8 show the dice scores for SAM versus LangSAM. We see that the percentage of dice scores over 0.5 for SAM was 62.5%, while for LangSAM that percentage was 75%. This study used dice scores to quantify algorithm performance. The dice score, D, is defined as the ratio of twice the number of common elements, A_0, to the sum of the number of elements in both sets, A_T.

$$D = \frac{2 \cdot A_0}{A_T}$$

A perfect match carries a dice score of 1.0 while a perfect mismatch carries a dice score of 0.0. Figure 8 shows the LangSAM model often outperformed the SAM model when tested on the 24 Mars surface images (Figure 8). SAM, while effective on the training data of 3D-printed tubes in simulated terrain, struggled with real Martian conditions—often confusing rocks or shadows for tubes, especially when tubes were partially buried or covered in dust. In contrast, LangSAM, using the text prompt "tube," demonstrated greater accuracy and reliability, successfully identifying more sample tubes and producing cleaner segmentation masks. Its ability to interpret visual data with the added context of language made it better suited for the unpredictable and visually complex Martian landscape. These results suggest that language-augmented AI models offer a promising solution for future missions where precise object recognition is critical under limited training data conditions.

Figures 9 through 11 show example images which illustrate different performance regimes. As shown in Figure 9, both models did very well due to the tube being exposed and isolated. In Figure 10, both SAM and LangSAM performed poorly due to the tube being too far. Lastly, Figure 11 shows a case where LangSAM performs well, but SAM does not because it mistook the rock at the bottom of the image for the tube.

Table 1. Dice scores for all 24 testing images and their mean and median.

Image Number	SAM Dice Score	LangSAM Dice Score
1	0.01	0.01
2	0.01	0.01
3	0.01	0.01
4	0.01	0.06
5	0.1	0.02
6	0.01	0.01
7	0.47	0.94
8	0.33	0.97
9	0.73	0.97
10	0.9	0.96
11	0.83	0.97
12	0.8	0.99
13	0.83	0.99
14	0.86	0.98

15	0.77	0.95
16	0.71	0.96
17	0.81	0.98
18	0.87	0.98
19	0.87	0.97
20	0.64	0.96
21	0.85	0.96
22	0.35	0.98
23	0.94	0.97
24	0.93	0.95
Mean	0.57	0.74
Median	0.75	0.96

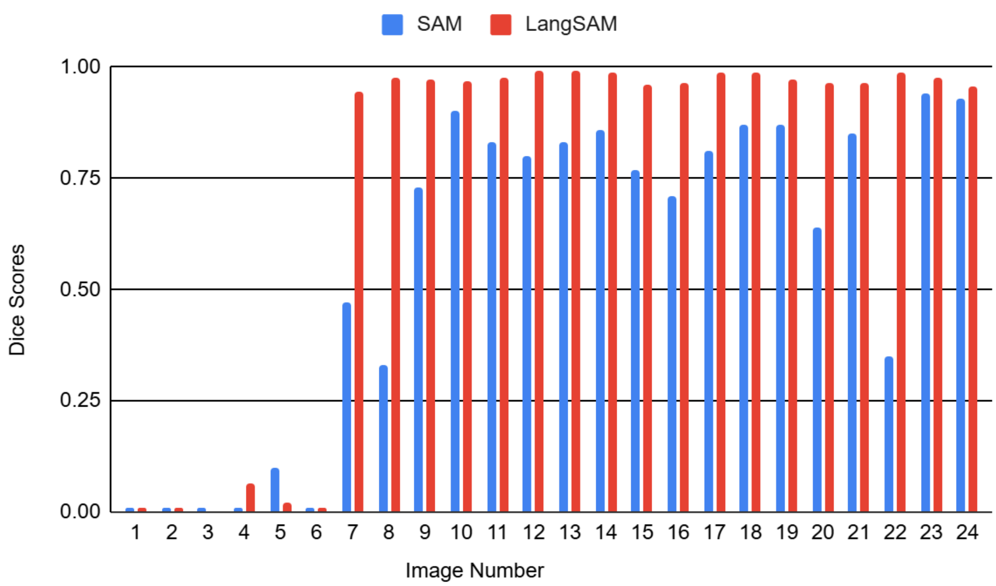

Figure 8. Dice score for all 24 test images using SAM and LangSAM.

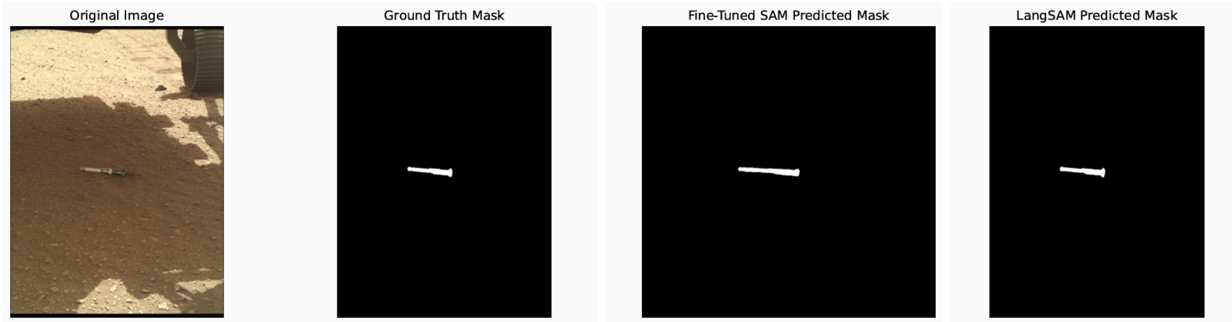

Figure 9. Typical image for high dice scores for both SAM and LangSAM (image 23).

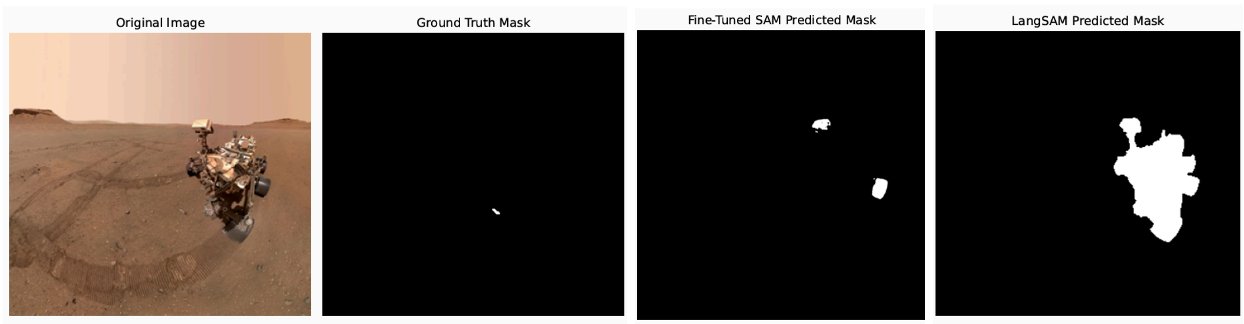

Figure 10. Typical image for low dice scores for both SAM and LangSAM (image 2).

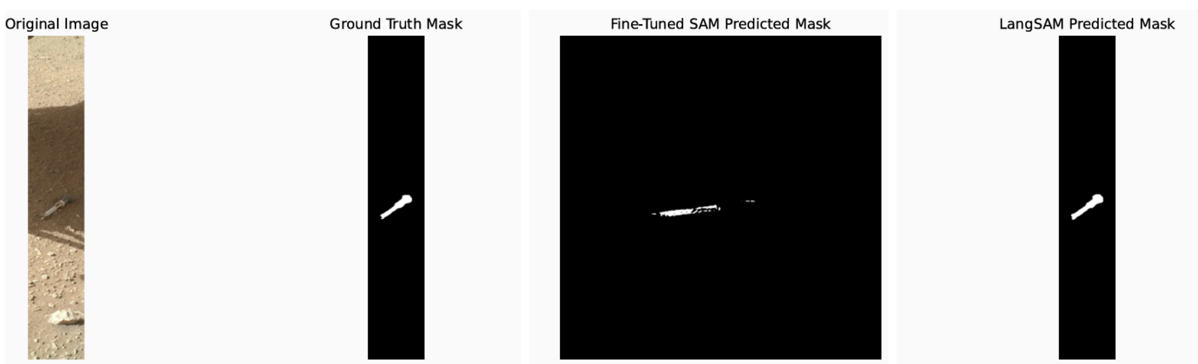

Figure 11. Typical image for low dice score (SAM) and high score (LangSAM) (image 7).

CONCLUSIONS

The LangSAM model demonstrated superior performance over SAM, particularly excelling in scenarios involving small datasets. This capability highlights LangSAM's potential effectiveness in resource-constrained environments such as planetary exploration. Given its strengths, incorporating advanced AI models like LangSAM into NASA's Mars rover missions could enhance the detection of features such as the Martian tubes dropped by Perseverance. Future research should continue to explore a range of segmentation models and deepen the understanding of how specific image features influence detection outcomes, with the goal of optimizing accuracy and adaptability across varying conditions.

ACKNOWLEDGEMENTS

The author would like to thank Anthony Cuturrufo for providing his guidance and suggestions during this project.

REFERENCES

1. NASA. (2025, April 3). Mars Sample Return - NASA Science. Science.nasa.gov. https://science.nasa.gov/mission/mars-sample-return/
2. Mars sample return. (2020). Esa.int. https://www.esa.int/Science_Exploration/Human_and_Robotic_Exploration/Exploration/Mars_sample_return
3. Perseverance Sample Tube 266 - NASA Science. (2021, September 6). Nasa.gov. https://science.nasa.gov/resource/Perseverance-sample-tube-266/
4. The Planetary Society. (2024). Cost of Perseverance. The Planetary Society. https://www.planetary.org/space-policy/cost-of-perseverance
5. Shukla, B. P., Kumar, A. B., & Pandya, M. R. (2026). Automated identification of Martian dust storms using AI-based approaches: A preparatory study for India's Mars Lander Mission (MLM). Journal of Atmospheric and Solar-Terrestrial Physics, 282, 106798. https://doi.org/10.1016/j.jastp.2026.106798
6. Dunlop, H., Thompson, D. R., & Wettergreen, D. (2007). Multi-scale features for detection and segmentation of rocks in Mars images. In 2007 IEEE Conference on Computer Vision and Pattern Recognition (pp. 1–7). IEEE. https://doi.org/10.1109/CVPR.2007.383257
7. Xiong, Y., Xiao, X., Yao, M., Cui, H., & Fu, Y. (2024). Light4Mars: A lightweight transformer model for semantic segmentation on unstructured environment like Mars. ISPRS Journal of Photogrammetry and Remote Sensing, 214, 167–178. https://doi.org/10.1016/j.isprsjprs.2024.06.008
8. Li, J., Chen, K., Tian, G., Li, L., & Shi, Z. (2025). MarsSeg: Mars surface semantic segmentation with multilevel extractor and connector. IEEE Transactions on Geoscience and Remote Sensing, 63, 1–12. https://doi.org/10.1109/TGRS.2025.3526630
9. Meta. (n.d.). Segment Anything | Meta AI. Segment-Anything.com. https://segment-anything.com/
10. Medeiros, L. (2024, March 3). luca-medeiros/lang-segment-anything. GitHub. https://github.com/luca-medeiros/lang-segment-anything
11. mars.nasa.gov. (n.d.). Images from the Mars Perseverance Rover. Mars.nasa.gov. https://mars.nasa.gov/mars2020/multimedia/raw-images/