Investigating the Benefits of Foundation Models for Mars Science M. Purohit[1,2], S. Lu[2], S. Diniega[2], U. Rebbapragada[2], and H. Kerner[1]; [1]School of Computing and Augmented Intelligence, Arizona State University, Brickyard Engineering, 699 S Mill Ave BYENG, Tempe, AZ 85281, [2]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109; (mpurohi3@asu.edu, you.lu@jpl.nasa.gov, serina.diniega@jpl.nasa.gov, umaa.d.rebbapragada@jpl.nasa.gov, hkerner@asu.edu).

Introduction: Planetary science investigations often involve analyzing vast volumes of orbital data from Mars and other planets to map features and quantify other properties. However, analyzing large volumes of such data manually is time-consuming and laborious. To address this challenge, researchers have increasingly developed machine learning (ML) models to automate data analysis. Existing work has demonstrated the ability of ML models to accelerate Mars-based tasks, such as crater and cone mapping [1, 2], landmark classification [3, 4, 5] and many others [6, 7]. However, building efficient ML models for individual planetary science use cases often requires the creation of large labeled training datasets. Transfer learning offers a powerful solution, leveraging knowledge from one domain to facilitate learning in another. Despite the potential of models pre-trained on large datasets like ImageNet (14M images), prior work showed that fine-tuning still requires substantial labeled digital images ($\sim$70,000 data samples for HiRISENet).

Recent work in ML introduced "Foundation Models" which are neural networks pre-trained on large amounts of diverse datasets (labeled or unlabeled) and generalize efficiently to new tasks. Foundation models have been shown to perform remarkably well on a wide range of natural language processing and computer vision tasks [8, 9, 10]. This remarkable performance can be attributed to the pre-training since it learns a broad base of knowledge and improves the performance on downstream tasks via fine-tuning even with smaller, labeled datasets. Pre-training can be performed in different ways depending on the model's architecture and the type of data. For instance, supervised pre-training can be done when labels are available for training data, mapping input-output correlations. In contrast, self-supervised pre-training can be done when labels are not available and the model generates or predicts input data itself. Foundation models have gained significant attention in Earth Observation (known as Geospatial Foundation Models) for solving tasks across diverse categories, e.g., agriculture, natural disaster, and landmark classification [11, 12, 13]. While previous work explored foundation models for Mars rover images [14, 15, 16, 17, 18], foundation models have not been investigated for orbital Mars data.

In this research, we developed foundation models and explored their efficiency in doing Mars science tasks for orbital data. Specifically, we focus on two downstream tasks: HiRISE (High Resolution Imaging Science Experiment) Landmark Classification, and identifying Martian Frost in HiRISE images. To develop the foundation models, we employed various pre-training data strategies over the Inception and ViT models. Specifically, we developed baseline models using zero pre-training and supervised pre-training with ImageNet and DoMars16 data. We then leveraged the self-supervised pre-training strategy using CTX data and compared the performance on two downstream tasks with the baseline. Our preliminary results demonstrate that self-supervised pre-training is a promising approach for building a foundation model for a wide variety of Martian tasks.

Downstream tasks: We selected Martian Frost [19] and HiRISE Landmark classification [3] as our initial downstream tasks. Martian Frost is a binary classification task to classify between frost and background surface (non-frost). HiRISE Landmark Classification is a multi-class classification task across 8 classes: Bright Dune, Crater, Dark Dune, Impact Ejecta, Slope Streak, Spider, Swiss Cheese, and Other. It is currently in operation at the Planetary Data Service Imaging Node. For all the experiments, we used the published training, validation, and testing sets associated with these tasks.

Models: To develop the foundation model for Mars-based tasks, we pre-trained and fine-tuned the *Vision Transformer* (ViT) [20] model (specifically, ViT-Large) and compared its performance with the Inception-v3 baseline [21] (as benchmark results available in [15]).

Baseline Pre-training Data Strategies: As a baseline, we considered the following three pre-training data strategies:

1. Zero Pre-training: When the model is not pre-trained on any dataset and the downstream task is directly performed on randomly initialized weights.
2. ImageNet: ImageNet is a large-scale database with more than 14 million images across 1300 classes of daily-life objects (e.g., cat, dog, chair, etc) [22].
3. DoMars16: DoMars16 is a Martian surface landmark classification dataset with 16150 data samples across 15 classes [4]. This dataset was created from the Context Camera (CTX) [23] installed on the Mars Reconnaissance Orbiter (MRO) satellite. We used the original training and validation split to pre-train the model.

As discussed above, since labels are available for ImageNet and DoMars16 datasets, we employ a supervised pre-training strategy and directly pre-train the model (ViT and Inception) on these datasets.

Self-Supervised Pre-training: We conducted a preliminary investigation to evaluate the performance of a foundation model that is pre-trained using CTX data in a self-supervised manner. CTX has global coverage of Mars and this data is easily accessible through Murray lab [24]. The dataset is a seam-corrected global image mosaic of Mars rendered at ~ 5.0 meters/pixel and subtiles are $2^\circ \times 2^\circ$ wide. We randomly sampled 90 subtitles (resolution of 23710×23710) from each longitude (ranges -180° $\times$ 180° and with a step size of 4°). From each subtile, we created non-overlapping data samples of size 200×200, resulting in a total of 13,924 samples per subtile and 1,253,160 (1.2M) samples overall. Finally, we split these data samples at tile-level into 90%-10% of training and validation, i.e., we used data from 81 subtiles for training and the 9 subtiles for validation to avoid spatial autocorrelation between training and validation.

Self-supervised pre-training for CTX data has been done using the Masked AutoEncoder (MAE) model, where the encoder (backbone) of the MAE [8] model is ViT. During pre-training, 75% of the portion of the CTX data samples were masked and the task of the MAE model is to predict the masked portion of the CTX images. Once the pre-training of MAE is completed, the backbone (ViT model) is used in the downstream tasks.

Experiment Results: The Martian Frost task was evaluated using the Area Under Curve (AUC) score, which quantifies how well the model differentiates between positive and negative classes; and the HiRISE Landmark classification task was evaluated using Accuracy. Results are shown in Table 1. Results for both tasks show that all the pre-trained models showed better performance compared to Zero-pretraining for both models (ViT and Inception). In addition, for both tasks, the ViT model pre-trained on ImageNet outperformed all other pre-training and Inception baselines. Interestingly, despite ImageNet being a cross-domain dataset for Mars-based tasks, it outperformed the DoMars16 and CTX data pre-training for Martian Frost and showed a comparable performance for HiRISE Landmark Classification. These results demonstrate that cross-domain pre-training can still be beneficial to improve the performance for Martian tasks. We can observe that the number of pre-training data samples in ImageNet is significantly higher compared to DoMars16/CTX. This suggests that the model's ability to learn general feature extraction capabilities from a huge amount of data can be advantageous. A similar observation is shown by Atha *et. al.* for Martian terrain segmentation task in [14]. For CTX data, we can see that despite having a smaller number of samples (1.2M) compared to ImageNet data (14M), the model pre-trained on CTX data shows comparable performance with the model pre-trained on ImageNet. This suggests that a smaller amount of in-domain data (CTX) can be just as effective for pre-training compared to a much larger, but out-of-domain dataset (ImageNet).

Model	Pre-training Data	Pre-training Data Size	Martian Frost	HiRISE Landmark
			AUC ↑	Accuracy ↑
ViT	Zero Pre-training	0	0.83	0.46
	ImageNet	14M	**0.99**	**0.88**
	DoMars16	16K	0.95	0.56
	CTX data	1.2M	0.93	0.86
Inception	Zero Pre-training	0	0.95	0.44
	ImageNet	14M	0.96	0.68

Table 1: Result for Martian Frost and HiRISE Landmark Classification (Here, K and M indicate thousands and millions, respectively). ↑ indicates higher the result, better the performance.

Conclusions and Future Work: In this research, we built and investigated the capability of foundation models for two Martian tasks. Our experimental results show that the foundation models improve performance compared to supervised training. Furthermore, our proposed self-supervised pre-training strategy based on CTX data shows promising results, suggesting value for exploring curated data preparation methods for self-supervised learning. Our future work involves, developing a more curated foundation model for Martian tasks and including a wide variety of downstream tasks, e.g., crater counting and mapping, finding the existence of water bodies, etc.

Acknowledgment: Part of this research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References: [1] Malvi S. *et al.* (2023) *ACM SIGSPATIAL*, 6, 110-120. [2] Purohit M. *et al.* (2024) *IEEE/CVF WACV*, 6026-6035. [3] Wagstaff K. *et al.* (2021) *AAAI*, 35, 17, 15204-15213. [4] Wilhelm T. *et al.* (2020) *Remote Sensing*, 12 (23), 3981. [5] Wagstaff K. *et al.* (2018) *AAAI*, 32, 1. [6] Swan r. M. *et al.* (2021) *IEEE/CVF CVPR*, 1982-1991. [7] kerner H. *et al.* (2017) *AAAI*, 33, 01, 9484-9491. [8] Brown T. *et al.* (2020) *NeurIPS*, 34, 1877-1901. [9] Radford A. *et al.* (2021) *ICML*, 8748-8763. [10] He K. *et al.* (2022) *IEEE/CVF CVPR*, 16000-16009. [11] Tseng G. *et al.* (2023) *arXiv*, 2304.14065. [12] Klemmer K. *et al.* (2023) *arXiv*, 2311.17179. [13] Manas O. *et al.* (2021) *ICCV*, 9414-9423. [14] Atha D. *et al.* (2021) *IEEE AERO*, 1-13. [15] Goh E. *et al.* (2022) *IEEE AERO*, 1-10. [16] Vincent G. *et al.* (2022) *ECCV*, 96-111. [17] Ward, I. R. *et al.* (2023) *ECCV*, 170-185. [18] Vincent, G. M. *et al.* (2023) *Journal of Spacecraft and Rockets*, 1-13. [19] Doran G. *et al.* (2024) *arXiv*, 2403.12080. [20] Dosovitskiy A. *et al.* (2020) *ICLR*, 8. [21] Szegedy C. *et al.* (2016) *IEEE/CVF CVPR*, 2818-2826. [22] Russakovsky O. *et al.* (2009) *IJCV*, 115, 3, 211-252. [23] Malin M. C. *et al.* (2007) *Journal of Geophysical Research: Planets*, 112, E5. [24] Dickson J. L. *et al.* (2018) *LPSC*, 49, 1-2.