National Aeronautics and Space Administration

FY24 Strategic University Research Partnership (SURP)

Lowering the Barriers to Planetary Science Studies with a Large Mars Model

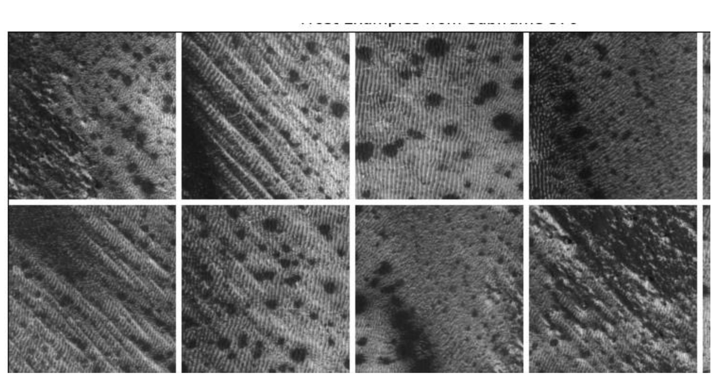
Principal Investigator: Umaa Rebbapragada (398)
JPL Co-Investigators: You Lu (398), Serina Diniega (322)
Arizona State University: Hannah Kerner (Assistant Professor), Mirali Purohit (Ph.D. Student)

Objective, Background and Approach

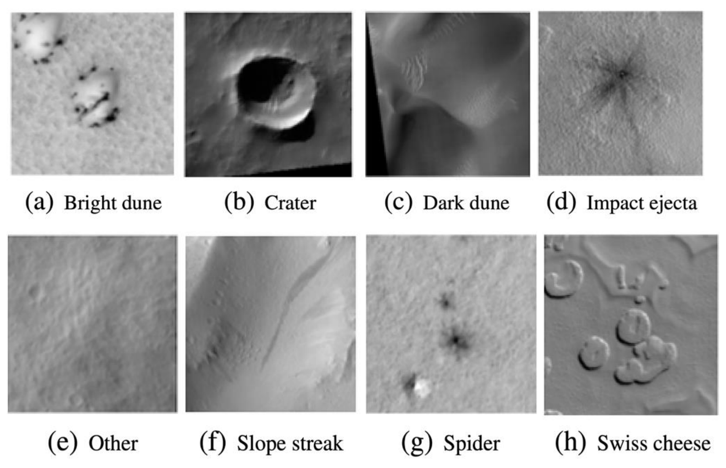
Our technical objective is to build and evaluate a first-of-its-kind foundation model for planetary science using use cases already benchmarked by JPL.

Machine learning (ML) is increasingly used to perform tasks within Martian datasets, such as:

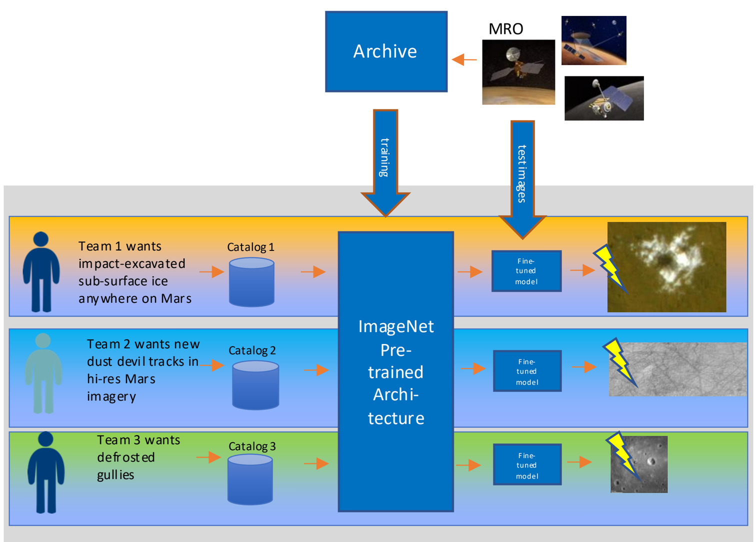
Martian Frost
Detection of global Martian frost in HiRISE images. (Doran et. al, 2024)



HiRISE Landmarks
Detection of surface features. (Wagstaff et. al., 2021)



Both projects used this paradigm for training their models:

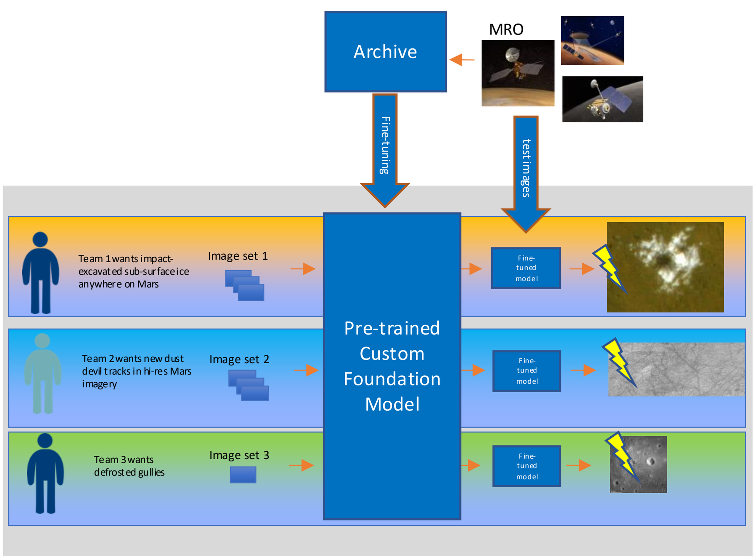


- ImageNet architecture pre-trained with *out-of-domain* digital photography images.
- Individual task models are fine-tuned with ~10K images
- High cost and effort to annotate image

What are Foundation Models?

Unlike task-specific deep neural network models, a foundation model is trained on large corpuses of unlabeled data to serve as a foundation for many different machine learning tasks. They promise:

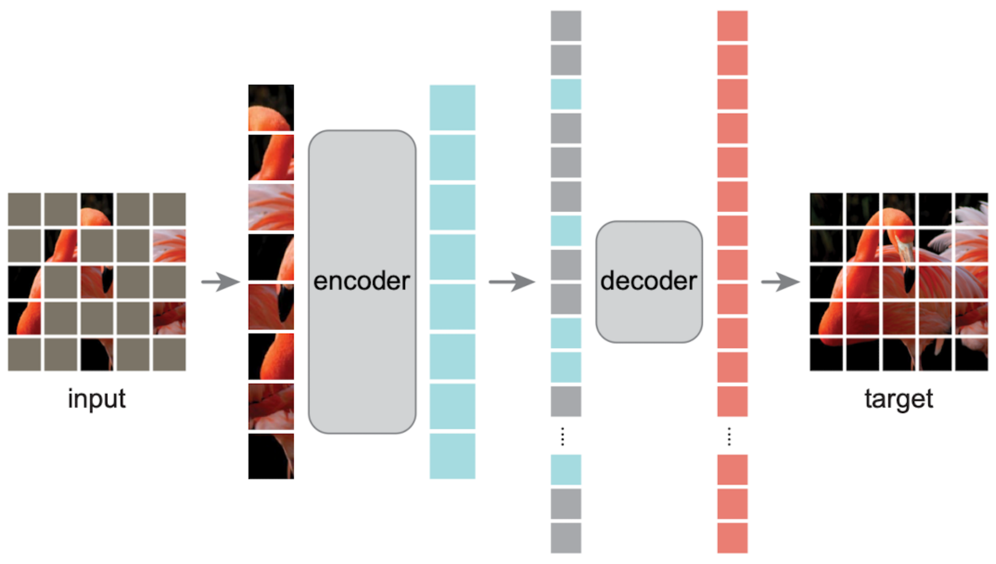
- Better performance
- Large reduction in number of training images needed for fine-tuning
- Increased number of "downstream" tasks supported



How are Foundation Models trained?

The training objective is to learn meaningful representations of the data without relying on human-provided labels. Thus, it can use a full archive of unlabeled satellite or rover data. Below is a depiction of the Masked AutoEncoder (MAE) approach (He et al., 2021).

- Learns to reconstruct images through random deletion of patches
- Uses a vision transformer (ViT) backbone (Dosovitskiy et. al., 2021)



Approach and Preliminary Results

We upgraded benchmarks to use state-of-the-art ViT models and compared the following:

- ViT with randomized weights
- ViT pre-trained with ImageNet (Deng et al., 2009)
- ViT pre-trained with in-domain labeled data from DoMars16 (Wilhelm et al. 2020)
- ViT pre-trained with in-domain unlabeled data from CTX using the MAE (He et al., 2021)

Model	Pre-training Data	Pre-training Data Size	Martian Frost	HiRISE Landmark
			AUC ↑	Accuracy ↑
ViT	Random Initialized Model	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

Prelminary Result: ImageNet still outperforms in-domain pre-training, but…

For both tasks, the CTX model approaches ImageNet performance with a fraction of the pre-training data

Future Work

- Adding more downstream science tasks
- Adding more CTX pre-training data
- Vary pre-training data sampling techniques
- Adding pre-training data from HiRISE, THEMIS, etc.
- Long-term goals: expand torover-based tasks
- Deploying model on PDS or JMARS

National Aeronautics and Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

www.nasa.gov

RPD-000 Clearance Number: CL#00-0000

References:
- Doran et. al., Evaluating Terrain-Dependent Performance for Martian Frost Detection in Visible Satellite Observations, arXiv (2024).
- Wagstaff et. al., Mars Image Content Classification: Three Years of NASA Deployment and Recent Advances, AAAI Workshop (2021).
- Dosovitskiy et. al., An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, ICLR (2021).
- Deng et. al., ImageNet: A Large-Scale Hierarchical Image Database, CVPR (2009).
- Wilhelm et. al., DoMars16k: A Diverse Dataset for Weakly Supervised Geomorphologic Analysis on Mars, Remote Sensing (2020).
- He et. al., Masked Autoencoders Are Scalable Vision Learners, CVPR (2021).