Creating Microscopic Digital Elevation Models of Planetary Regolith Using Monocular Images and Convolutional Neural Networks. F. Diotte[1], M. Lemelin[1] and E. A. Cloutis[2], [1]Département de géomatique appliquée, Université de Sherbrooke, QC, Canada J1K 2R1 (Frederic.Diotte@usherbrooke.ca), [2]Department of Geography, University of Winnipeg, MB, Canada R3B 2E9.

Introduction: The accurate estimation of the surface properties of regoliths is essential for understanding the structure and texture of a planet's surface. Soil properties such as grain size, roughness, and porosity can provide insights into the processes that have shaped a planet's surface over time. For example, remotely sensed images of lunar swirls and permanently shadowed regions suggest that these regions have geotechnical properties that differ from those of surrounding soils. High forward scattering of lunar swirls is consistent with lower millimeter scale roughness [1], while space weathering trends in lunar swirls may indicate a higher content of fine-grained feldspathic material [2]. In permanently shadowed regions, low far-UV albedo suggests a higher porosity compared to sunlit soils [3].

To verify these hypotheses, in situ data such as high spatial resolution camera images are needed. Future missions such as Lunar Vertex, VIPER, and the Canadian lunar rover include multispectral microscopes. These instruments will provide images that can be used to generate digital elevation models (DEMs), which in turn can be used to calculate the microscopic topographic roughness.

Although stereophotogrammetry is a robust approach, it cannot be applied to monocular images, and the quality of its results is highly dependent on the baseline distance of the stereo pair. Deep learning approaches for mineralogy, texture, grain size, porosity, and microtopography estimation (which we term "depth" here), on the other hand, now offer robust, accurate, and rapidly improving performance.

In this study, we propose using deep learning techniques to estimate depth from monocular images of Martian soil acquired by the Microscopic Imager (MI) instrument, in preparation for upcoming lunar data. Our approach leverages the rich information contained in these images, including texture, color, and shape, to estimate the distance of each pixel to the camera. We demonstrate the ability of our approach to consistently predict depth for submillimeter scale soils, although fine-tuning of the model will be required for acceptable accuracy.

Dataset: We used a dataset of stereo pairs of images and corresponding DEMs produced by [4] of 1359 Martian targets acquired using the Microscopic Imager on the Spirit and Opportunity missions. The input images are 8-bit focal section merges with a pixel size of 35 µm. The DEMs also have a radiometric resolution of 8 bits, a horizontal resolution of 460 µm, and a depth repeatability of approximately 600 µm [4]. To create a larger dataset, we divided each image into nine 224 x 224 pixel tiles, resulting in a total of 12,168 images. The data was then divided into training, validation, and testing sets in an 80/20/20% proportion. To increase the diversity of the dataset, we applied data augmentation techniques such as flipping, adjusting contrast, hue, saturation, and brightness. The inputs were standardized using the mean and standard deviation of the training set. The target values were normalized between 0 and 1.

Figure 1 provides an overview of the MI dataset. The targets have variable textures, from fine dust to consolidated surfaces. Illumination conditions vary greatly, and some surfaces are entirely covered by shadow. The focusing is not optimal at all pixels, with some blurred regions in the images. Albedo variations caused by mineralogy and alterations are also visible.

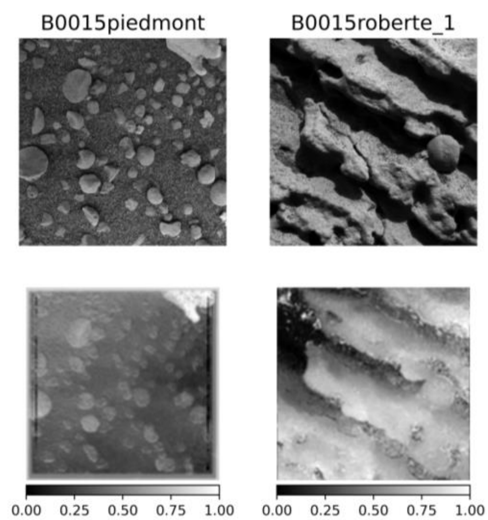

Figure 1. Images (row 1) and DEMs (row 2) of targets B0015piedmont and B0015roberte_1 from the Microscopic Imager on the Spirit and Opportunity missions. Depth values are normalized between 0 and 1 and are unitless.

Model: The architecture of our model is largely inspired by that of [5], but we adopt a supervised training approach instead of their self-supervised approach, using focal merge sections as inputs and DEMs as targets. We also use a simple MSE loss function and employ a pre-trained Resnet-18 encoder with a matching number of upsampling layers, all of which use ELU activation functions and have skip connections. The Adam optimizer is used with a base learning rate of 1e-4 and a batch size of 8. The weights of the network are adjusted by iterating through the training dataset over 67 epochs. The validation dataset is used to select the optimal version of the model based on performance metrics.

Results: Examples of predictions are shown in Figure 2. The model is most successful at understanding well defined, contrasting objects with simple shapes, such as larger grains scattered on a finer dust. Despite the model's ability to recognize shapes, the absolute range of depths is sometimes significantly off, as shown by performance metrics described below. The model is also sensitive to shadows and will sometimes infer higher depth in large regions that are simply obscured (see fourth row in Figure 2).

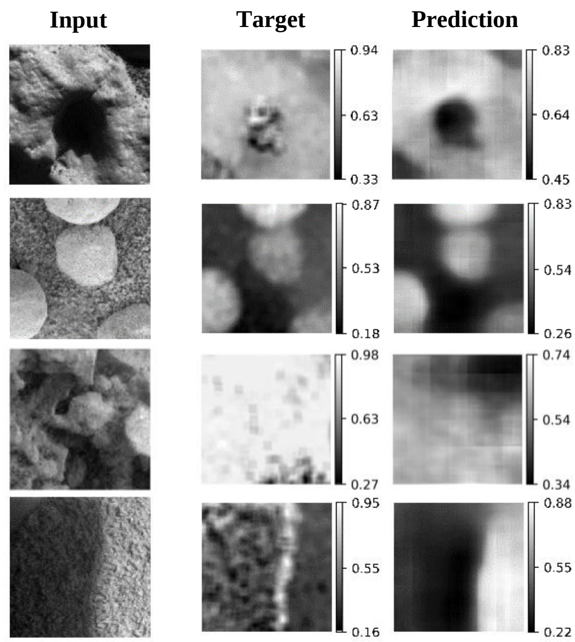

Figure 2. Tiles of focal section merges (input), MNEs (target) and predictions for four MI targets. From top to bottom: B0015roberte_1, B0046rubel2, B2944Amboy_R9, B2485Luis_de_Torres_1.

It appears that the quality of the DEMs from the MI dataset is a significant bottleneck for the model's performance. Many DEMs contain square artifacts, and depth variations in the images do not always align with the objects depicted (third row in Figure 2). The coarse horizontal and vertical resolutions of the DEMs also affect the model's predictions. As a result, the model is unable to predict depth for grains with a diameter of a few pixels.

Evolution of loss with training epochs is shown in Figure 3. The model reaches the minimum validation loss at 20 epochs, but the validation and training loss quickly diverge after a few (< 5) epochs, indicating overfitting of the data and poor generalization ability of the model. To evaluate the model's performance, we calculated the relative absolute error (3,52), the relative squared error (2,42), RMSE (0,5117), and $\delta < 1.25$ threshold accuracy percentage (9,62 %). The errors are high, considering the range (0 – 1) of values used. The low percentage of the predicted/reference ratio being less than 1.25 indicates that only a small fraction of the estimated depths is accurate.

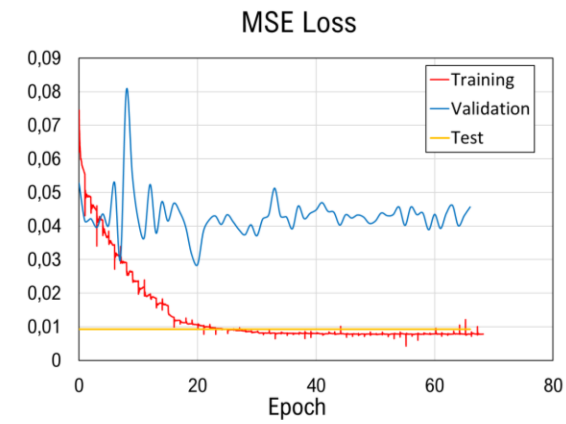

Figure 3. Evolution of loss with training epochs.

Future work: There are two ways in which we propose to improve this approach. (1) We will render numerically modelled soils with high variance in target, illumination, and observation conditions. Numerical modelling will greatly increase the size of the dataset and should reduce overfitting. It will also help to overcome the restriction in spatial resolution imposed by stereophotogrammetry. (2) We will add a fractal statistics constraint to the loss function. Using Apollo Lunar Stereo Camera images, [6] demonstrated that the scale dependence of lunar's surface roughness follows fractal statistics at submillimeter to subcentimeter scales. We will therefore calculate closeness of the model's predictions to fractal statistics to adjust the loss function, with the hope of improving the model's performance when applied to lunar soils.

References:[1] Schultz and Srnka (1980) *Nature*, 284, 22-26. [2] Garrick-Bethell *et al.* (2014) *J. Geophys. Res. Planets*, 119, 5. [3] Gladstone *et al.* (2012) *J. Geophys. Res. Planets*, 117, E00H04. [4] Herkenhoff *et al.* (2006) *J. Geophys. Res. Planets*, 111, E2. [5] Peng *et al.* (2021) *CVPR* 2021. [6] Helfenstein and Shepard (1999) *Icarus*, 141, 107-131.

Acknowledgments: This project is undertaken with the financial support of the Canada Research Chair in Northern and Planetary Geological Remote Sensing, NSERC, and the Canadian Space Agency [22EXPVIPER], held by prof. Lemelin.