LABELMARS: CREATING AN EXTREMELY LARGE MARTIAN IMAGE DATASET THROUGH MACHINE LEARNING.

S. P. Schwenzer[1], M. Woods[2], S. Karachalios[2], N. Phan[2], L. Joudrier[3]. [1]The Open University, Milton Keynes MK7 6AA, UK; susanne.schwenzer@open.ac.uk, [2]SCISYS, 23 Clothier Road, Bristol BS4 5SS; mark.woods@scisys.co.uk, [3]European Space Agency (ESA).

Introduction: Four landers (Viking 1,2, Phoenix and Insight) and 4 rovers (Sojourner, Spirit, Opportunity, and Curiosity) have successfully operated on the Martian surface since 1976, with the combined operation time of landers exceeding 3680 sols and that of rovers exceeding 9660 sols at the time of writing of this abstract (December 2018)[1]. This has returned a large dataset of images from the cameras on board, with examples of recent image-based research on Curiosity images alone including, but not limited to, studies of aeolian, active bedforms [2,3], conglomerates and river beds [4], sedimentary structures [5,6,7], and erosional features [8], together for a reconstruction of the geology of the site [9,10,11]. This list is clearly not exhaustive but it demonstrates how images are key to understanding of the geologic environment at the site, and thus are basis for operational decisions and an invaluable science data resource. With two missions currently active (Curiosity, InSight) and two more scheduled to launch in 2020 (ESA ExoMars, NASA Mars2020) this data set is a 'big data' problem, and growing. In order to facilitate easier access, especially for researchers who do not have the luxury of following the mission on a daily basis, this research has developed an automated terrain labelling and classification system based on state of the art machine/deep learning which enables keyword based search.

The project: LabelMars achieved 5000 annotated images from the Spirit, Opportunity and Rover navigation camera data bases. The project was a part of a larger European Space Agency (ESA) project called Novelty or Anomaly Hunter (NOAH) which has a number of other tasks, including the labelling as a citizen science project [12], an AI algorithmic evaluation, and a prototype flight detector developed which ported some of the algorithms to flight C versions in order to enable future on-board operations of the technology. We started with the entire set of available navigation camera images from the MER and MSL rovers sourced from the Analyst's Notebook [13] and down-selected those by selecting continuous rows of sols from different terrains and subsequently deleting all unsuitable ones from this set of images (e.g., if they were too dark or an exact repeat of another one). This resulted in 5917 images {Spirit (2724), Opportunity (1173) and Curiosity (2020)}, which were further reduced to exactly 5000 by eliminating similar scenes and very dark images. The resulting images were manually labelled.

Table 1. Example of the category structure for the labelling.

Category	Sub-Category	Classifier 1
Artificial	Foreign object debris	
	Shadows from hardware	
	Spacecraft parts	
	Tracks	
Float Rock	Alteration	Concretions/Nodules
		Crystals
	Magmatic	Dark toned
		Light toned
	Meteorite	
	Sedimentary	Dark toned
		Light toned
Outcrop	Alteration	Bleaching
		Concretions/Nodules
		Veins
	Impact Related	Craters and Ejecta
		Rock Outcrops
	Magmatic	Dark toned
		Light toned
	Sedimentary	Dark toned
		Light toned
Unconsolidated	Drifts	
	Dunes	
	Gravel Beds	Homogeneous
		Structured
Sky		
Don't know		

We started by recruiting 'citizen scientists' targeting undergraduate students on geology modules and in the first 4 months had 185 labellers from this audience as well as "general users" with unknown to us background from internet advertisements. We found, consistent with scientific findings on larger citizen science projects (e.g., [14]), that only a small number of those volunteers engaged persistently, and 507 validated images were labelled, cross checked and added to the database. To finish the project on time, we changed the strategy and contracted 4725 images to 20 undergraduate students from University of Leicester. They had 4 weeks to complete their set of images. The students received a tutorial and labelled a test set of images. Each student's initial set of labelled images was checked and feedback was given as necessary. They proceeded to label all the images in their portfolio, with a subset of them being cross checked for consistency between labelers. Discussions and notices on a dedicated and monitored forum further enhanced the quality of the work. **Categories:** The categories were mainly based on the main morphological features. Table 1 lists the category outcrops with all its sub-categories. The other three main categories were float rocks (sub-categories sedimentary, magmatic, alteration and meteorite), unconsolidated material (sub-categories dunes, drifts and gravel beds), and artificial (sub-categories foreign object debris, spacecraft parts, shadows from hardware and tracks). Subcategories were divided further as needed, see Table 1.

AI Based Terrain Classification: NOAH was conceived to investigate the suitability of state of the art machine learning techniques for classification of in-situ Mars Rover images. It follows on a from a series of work in the area of science autonomy [15,16]. A precursor project [17], used what at the time was the best in class classification techniques but the field of machine learning has accelerated dramatically in recent years. In particular Deep Learning has set the benchmark for image classification, object detection and semantic segmentation. In 2015 such techniques were shown to outperform human labelers in benchmark tests [18].

NOAH therefore sought to update MASTER by investigating appropriate Deep Learning techniques in addition to those already in use. A key feature of these approaches is that they rely heavily on large volumes of datasets to learn how to approximate a mapping between image pixels and class probabilities. To address this challenge NOAH included LabelMars which sought to create a labelled dataset of exiting Mars Rover images. To enable this a dedicated web-based tool called the Data Annotation Labelling Tool (DAT) was developed. The DAT was successfully used to facilitate detailed labelling of rover images. Examples are outlined in Figure 2 and 2.

The dataset gathered during the two stage labelling activity has been used to train and evaluate a range of algorithms. In particular we used the FASTER R-CNN [19] architecture to realise sub-elements of the MASTER pipeline approach including saliency detection and region of interest classification. This work is now being completed and results will be reported shortly.

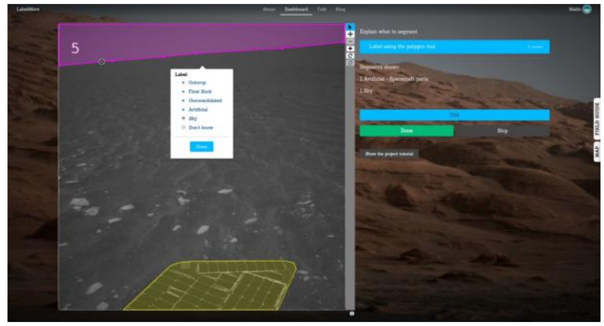

Figure 1: *Example of the DAT being used to label NavCam images using the ontology outlined in Tab. 1.*

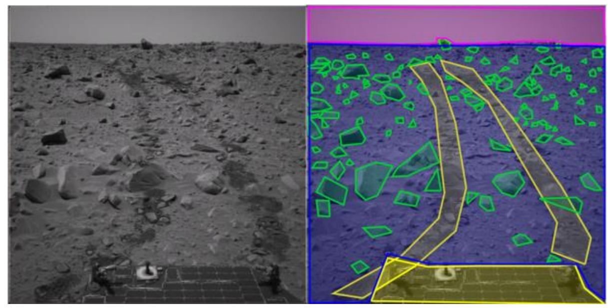

Figure 2: *Example shows labelling of float rocks, sky, rover tracks, ground and rover parts.*

Conclusions: The LabelMars activity has successfully created a large dataset which has been used to train and evaluate deep learning algorithms for automatic terrain classification of Mars rover images. The computer vision and machine learning field is developing rapidly and this dataset and the NOAH evaluation framework will allow new algorithms to be evaluated and deployed as they emerge ensuring that scientists have ready access to important mission science data.

References: [1] NASA mission websites of the named missions at www.nasa.gov. [2] Baker, M. M. et al. (2018) *GRL, 45*, 8853–8863, [3] Weitz, C. M. et al. (2018) *GRL, 45*, 9471-9479. [4] Williams, R. M. E. et al. (2013) *Science, 340*, 1068–1072. [5] Stack, K. M. et al. (2018) Sedimentology, doi: 10.1111/sed.12558. [6] Williams, R: M. E. et al. (2018) Icarus, 309, 84–104. [7] Banham et al. (2018) *Sedimentology, 65*, 993-1042. [8] Bridges, N. T. et al. (2014) *JGR Planets, 119*, 1374-1389. [9] Grotzinger, J. P. et al. (2014) *Science*, DOI: 10.1126/science.1242777. [10] Grotzinger, J. P. et al. (2015) *Science*, DOI: 10.1126/science.aac7575. [11] Kah et al. (2015) *Elements, 11*, 27-32. [12] Wallace et al. (2017) *LPSC*, abstr. #1170. [13] https://an.rsl.wustl.edu/. [14] Ponciano, L., and Basiliero, F., (2014) *Human Computation (2014) 1*, 245-264. [15] Woods, M. et al. (2009) JFR, doi: 10.1002/rob.20289. [16] Paar, G. et al. (2012), *ICRA XXIX*, 83010A-83010A-14, [17] Wallace, I. et al (2015), ASTRA 2015, [18] Russakovsky, O. et al. (2015), *IJCV, 2015*, [19] Shaoqing, R. et al. (2015), *NIPS' 15*, 1, pp 91-99.

Acknowledgements: The NOAH Project is carried out under a R&D program of and funded by the European Space Agency.