LABELMARS.NET: CROWD-SOURCING AN EXTREMELY LARGE HIGH QUALITY MARTIAN IMAGE DATASET.
I. Wallace[1], S. P. Schwenzer[2], M. Woods[1] , N. Read[1], S. Wright[1], K. Waumsley[1] L. Joudrier[3] [1]SCISYS, 23 Clothier Road, Bristol BS4 5SS; iain.wallace@scisys.co.uk, [2]The Open University, Milton Keynes MK7 6AA, UK; susanne.schwenzer@open.ac.uk , [3]European Space Agency (ESA)

Introduction: The observation of landforms, outcrops and small features within a (Martian) landscape is key to the understanding of its geologic past as well as present environmental conditions. Studies of such features have – for example – revealed the nature of streambeds at Gale Crater [1], and allowed to study Aeolian bedforms as they were encountered by the Curiosity [2]–[4], and Spirit rovers [5], [6]. With two active rovers (Opportunity, Curiosity) currently on Mars, and two more to be launched in 2020 (ExoMars, Mars2020), the imaging data sets are a huge, growing resource, which need to be explored as best as possible.

LabelMars (www.labelmars.net) is a citizen science activity to collect geological annotations of Martian rover navigation camera images. As part of the ESA NOAH (Novelty Or Anomaly Hunter) project it will provide a large, high quality dataset to develop state-of-the-art machine vision algorithms for autonomous science detection, targeted at future rover missions.

.

The Value of Annotated Datasets: Large well-annotated datasets are of exceptional value to the computer vision community, and in this case the Mars exploration community. The goal of the LabelMars activity is to achieve five thousand annotated images – this will make it the largest geologically annotated Martian rover imagery database for vision research and development. In the wider image processing community there have been huge advances kick started by the availability of large datasets – most famously with ImageNet [7]. Of course we do not aim for the scale of ImageNet's million-plus images, the scale is smaller – but the ambition is grand within the restricted scope of autonomous planetary science.

Annotated and pre-screened scientific data sets are of large benefit to the scientific community too, because they allow research on data bases which cannot be created within the financial and personnel base of a single project. There are numerous examples, mostly under the envelope of 'citizen science' which have demonstrated the value of such data sets. For a set of examples from space science, see [8]. Two examples of such overall classification have returned especially prominent results. The 'Galaxy Zoo' project used classification of the currently available very large number of images for volunteer-based morphological classification of galaxies [e.g. [9]; this project even led to the identification of new objects (with the volunteers being acknowledged in the publications [10]). In a similar way, the 'stardust@home' project asks volunteers to find the tiny tracks particles made in the aerogel of the Stardust mission [11], which by 2016 had resulted in the identification of over 200 impacts in the aerogel tiles. This was made possible by over 30,000 volunteers, who carried out more than 10^8 searches [12], enabling the science that follows.

For rover based images of the Martian surface, labelled images would allow – similar to the Galaxy Zoo project – the identification of similar textures and objects at different landing sites and use statistical methods for comparison – as well as enabling topical work across the different current and future landing sites.

The Drive for Science Autonomy: Bandwidth or communication limitations may make real-time control of instruments for scientific discovery difficult or impossible. For planetary rovers there is a trade-off between detailed observation to ensure important targets are not missed, which requires slow traverses to downlink all the data, and maintaining sufficient progress to visit many science targets.

By adding an autonomous capabilities to detect novel or scientifically interesting phenomena in images we can enhance the scientific return of robotic exploration missions. The many images acquired for navigation during traverse operations that are normally discarded can be inspected, and those scoring highly selected for downlink or even further sensing of the target. Of course more powerful rover AI and autonomy will not replace ground-based scientists, but it offers the potential to to lower thye probability of missing proximal science targets. This has been demonstrated in practice in NASA's AEGIS [13] system deployed on MER and MSL, and in prior ESA and UKSA funded work investigating new methods in science autonomy such as the NOAH precursor MASTER [14] and CREST [15] projects.

Expected Scientific Outcomes: Creating such a large set of labelled data will have intrinsic scientific value. Labelled data sets allow the comparison of different sets of images – e.g., along a traverse of a single rover or between different landing sites. Questions such as frequency of float rocks, or the abundance of light-toned veins can be treated statistically over a large set of images. Environmental and formation parameters can thus be characterized, e.g., the intensity of fracturing at different rover positions/landing sites.

The Annotation Process: Registered users will be presented with Navcam images to annotate, which they will do with a combination of drawing polygons to indicate areas of interest and confirming or labelling polygons drawn by other users. The Martian rover imagery is ideal to provide a narrative appeal combined with a visual indicator of progress by offering participants images from sols in sequence (or within a sol) with the idea that they are "seeing where the rover goes next". Participants can then visualize their progress by seeing pins on a map of Mars. This also provides a global progress indicator to show all users the overall annotation progress. Additionally there will be a financial prize awarded for the completion of high quality annotations, as defined by agreement between labels produced by multiple participants.

Get Involved: Participants with geological expertise are invited to take part in this activity via the website, where they will get a unique chance to explore the surface of Mars through the eyes of the NASA rovers. By annotating Martian images with objects of scientific interest participants will get to deepen their understanding of the Martian terrain whilst helping shape the future of on-board autonomous science

The LabelMars project is open for registration at www.labelmars.net. Some understanding of geology is required, but there is a simple online guide which allows potential participants to gauge their capacity to contribute. Participants have the chance to influence state-of-the-art space autonomous system development, and contribute their effort to a lasting impact on future planetary exploration.

References:

[1] R. M. E. Williams *et al.*, "Martian Fluvial Conglomerates at Gale Crater," *Science*, vol. 340, no. 6136, pp. 1068–1072, 2013.

[2] M. G. A. Lapotre *et al.*, "Large wind ripples on Mars: A record of atmospheric evolution," *Science*, vol. 353, no. 6294, pp. 55–58, 2016.

[3] M. Day and G. Kocurek, "Observations of an aeolian landscape: From surface to orbit in Gale Crater," *Icarus*, vol. 280, pp. 37 – 71, 2016.

[4] R. E. Arvidson *et al.*, "Mars Science Laboratory Curiosity Rover Megaripple Crossings up to Sol 710 in Gale Crater," *J. Field Robot.*, 2016.

[5] R. Greeley *et al.*, "Columbia Hills, Mars: Aeolian features seen from the ground and orbit," *J. Geophys. Res. Planets*, vol. 113, no. E6, 2008.

[6] "Spirit at Gusev Crater: Plates," *Science*, vol. 305, no. 5685, pp. 811–818, 2004.

[7] O. Russakovsky *et al.*, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[8] B. J. H. Méndez, "SpaceScience@Home: Authentic Research Projects that Use Citizen Scientists," in *EPO and a Changing World: Creating Linkages and Expanding Partnerships*, 2008, vol. 389, p. 219.

[9] C. Lintott *et al.*, "Galaxy Zoo 1: data release of morphological classifications for nearly 900 000 galaxies," *\mnras*, vol. 410, pp. 166–178, Jan. 2011.

[10] D. Clery, "Galaxy Zoo Volunteers Share Pain and Glory of Research," *Science*, vol. 333, no. 6039, pp. 173–175, Jul. 2011.

[11] A. J. Westphal *et al.*, "Stardust@home: Virtual Microscope Validation and First Results," in *37th Annual Lunar and Planetary Science Conference*, 2006, vol. 37.

[12] A. Westphal *et al.*, "A Massively Distributed Search for Impacts in Aluminum Foil on the Stardust Interstellar Collector," in *Lunar and Planetary Science Conference*, 2016, vol. 47, p. 2275.

[13] T. A. Estlin *et al.*, "AEGIS Automated Science Targeting for the MER Opportunity Rover," *ACM Trans. Intell. Syst. Technol. TIST*, vol. 3, no. 3, p. 50, 2012.

[14] I. Wallace and M. Woods, "MASTER: A Mobile Autonomous Scientist For Terrestrial and Extra-Terrestrial Research," *13th Symp. Adv. Space Technol. Robot. Autom. ASTRA*, no. 3, 2015.

[15] M. Woods, A. Shaw, D. Barnes, D. Price, D. Long, and D. Pullan, "Autonomous science for an ExoMars Rover–like mission," *J. Field Robot.*, vol. 26, no. 4, pp. 358–390, 2009.

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

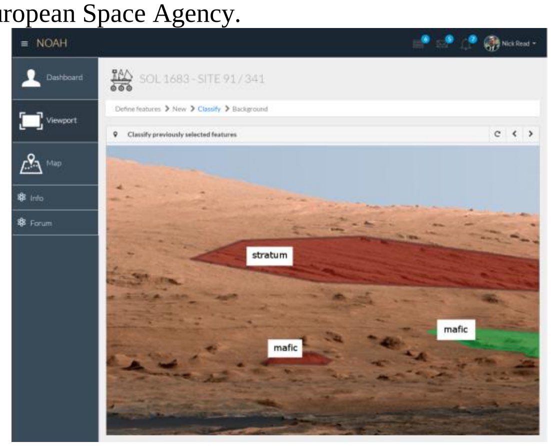

Figure 1 - Illustration of an example web labelling interface.