

Deep Learning-Driven Detection and Mapping of Rockfalls on Mars

Valentin Tertius Bickel¹, Susan J. Conway², Pierre-Antoine Tesson³, Andrea Manconi⁴,
Simon Loew⁵, and Urs Mall⁶

Abstract—The analysis of rockfall distribution and magnitude is a useful tool to study the past and current endogenic and exogenic activity of Mars. At the same time, tracks left by rockfalls provide insights into the mechanical properties of the Martian surface. While a wealth of high-resolution spaceborne image data are available, manual mapping of displaced boulders with tracks is inefficient and slow, resulting in: 1) a small total number of mapped features; 2) inadequate statistics; and 3) a suboptimal utilization of the available big data. This study implements a deep learning-driven approach to automatically detect and map Martian boulders with tracks in high resolution imaging science experiment (HiRISE) imagery. Six off-the-shelf neural networks have been trained either on Martian or lunar rockfall data, or a combination of both, and are able to achieve a maximum overall recall of up to 0.78 and a maximum overall precision of up to 1.0, with a mean average precision of 0.71. The fusion of training data from different planets and sensors results in an increased detection precision, highlighting the value of domain generalization and multidomain learning. Average processing time per HiRISE image is ~ 45 s using an NVIDIA Titan Xp, which is more than one order of magnitude faster than a human operator. The developed deep learning-driven infrastructure can be deployed to map Martian rockfalls on a global scale and within a realistic timeframe.

Index Terms—Mars, multidomain learning, object detection, rockfall.

Manuscript received February 17, 2020; revised April 1, 2020 and April 22, 2020; accepted April 27, 2020. Date of publication May 29, 2020; date of current version June 15, 2020. The work of V. T. Bickel was supported in part by the International Max Planck Research School, Max Planck Institute for Solar System Research at the University of Göttingen (IMPRS) and the Engineering Geology Group, Department of Earth Sciences, Swiss Federal Institute of Technology Zurich (ETHZ) and in part by the NVIDIA Academics Team under a GPU Grant. The work of S. J. Conway was supported by the French Space Agency CNES for the HiRISE-related work. The work of P.-A. Tesson was supported by the TEAM Program of the Foundation for Polish Science (TEAM/16-3/20) cofinanced by the European Union under the European Regional Development Fund. (*Corresponding author: Valentin Tertius Bickel.*)

Valentin Tertius Bickel is with the Planets and Comets Department, Max Planck Institute for Solar System Research, 37077 Göttingen, Germany, and also with the Engineering Geology Institute, Swiss Federal Institute of Technology, 8092 Zurich, Switzerland (e-mail: bickel@mps.mpg.de).

Susan J. Conway is with the Centre National de Recherche Scientifique, Laboratoire de Planétologie et Géodynamique, 44300 Nantes, France (e-mail: susan.conway@univ-nantes.fr).

Pierre-Antoine Tesson is with the Space Research Centre, Polish Academy of Science, 51622 Wrocław, Poland (e-mail: pt@cbk.pan.wroc.pl).

Andrea Manconi and Simon Loew are with the Engineering Geology Institute, Swiss Federal Institute of Technology, 8092 Zurich, Switzerland (e-mail: andrea.manconi@erdw.ethz.ch; simon.loew@erdw.ethz.ch).

Urs Mall is with the Planets and Comets Department, Max Planck Institute for Solar System Research, 37077 Göttingen, Germany (e-mail: mall@mps.mpg.de).

This article has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the authors.

Digital Object Identifier 10.1109/JSTARS.2020.2991588

I. INTRODUCTION

MARS is a dynamic planet and recent surface activity has been shown by a large number of studies using data returned from ground-borne and spaceborne missions, such as the Mars Science Laboratory (MSL or Curiosity) or the Mars Reconnaissance Orbiter (MRO). While Mars is host to a variety of active processes, mass wasting and aeolian processes are particularly relevant for this study. Dynamic aeolian features include dunes (e.g., [1], [2]) and dust devils (e.g., [3], [4]) that migrate across the Martian surface. Observed mass wasting processes include polar ice and dust avalanches (e.g., [5], [6]), rockfalls (e.g., [7]–[9]), slope streaks (e.g., [10]–[12]) and potentially recurring slope lineae (RSL) (e.g., [13]–[15]). As for RSL, other mass wasting phenomena are potentially related to the action of volatiles, such as flows in gullies (e.g., [13], [16], [17]) and flows over CO₂ ice (e.g., [18]).

The occurrence, frequency, and magnitude of mass wasting phenomena, and rockfalls in particular can be indicative of seismic activity of planetary bodies or moons in general, as several recent studies have shown ([7], [8] [19]–[21]). Thus, suspected regions of increased seismic activity inferred from rockfall frequency patterns could be priority targets for the deployment of future geophysical networks. The tracks created by extraterrestrial rockfalls, i.e., boulder tracks, are also a valuable tool to estimate the basic mechanical properties of the surface substrate present [22], [23].

However, the manual detection and mapping of Martian rockfalls (boulders with tracks, here also called feature of interest) in satellite imagery remains a challenging task. Since 2010, MRO's high resolution imaging science experiment (HiRISE) [24] has returned tens of thousands of high-resolution images of the Martian surface. While image data are abundant, traditional mapping of rockfalls is slow and inefficient—a human operator has to spend a substantial amount of time to scan through a single HiRISE image, which is a tiring and arduous process, particularly when mapping is performed over large regions using an increased number of images. In addition, the mapping of rockfalls on nonterrestrial surfaces is complicated by limitations of the used image data (resolution and coverage). However, a high number of rockfall detections is required to establish a robust statistical analysis, which in turn is the basis for scientifically robust conclusions. Previous studies on Martian rockfalls, such as those performed in [8] and [25], would greatly benefit from an increased number of detected and mapped boulders with tracks.

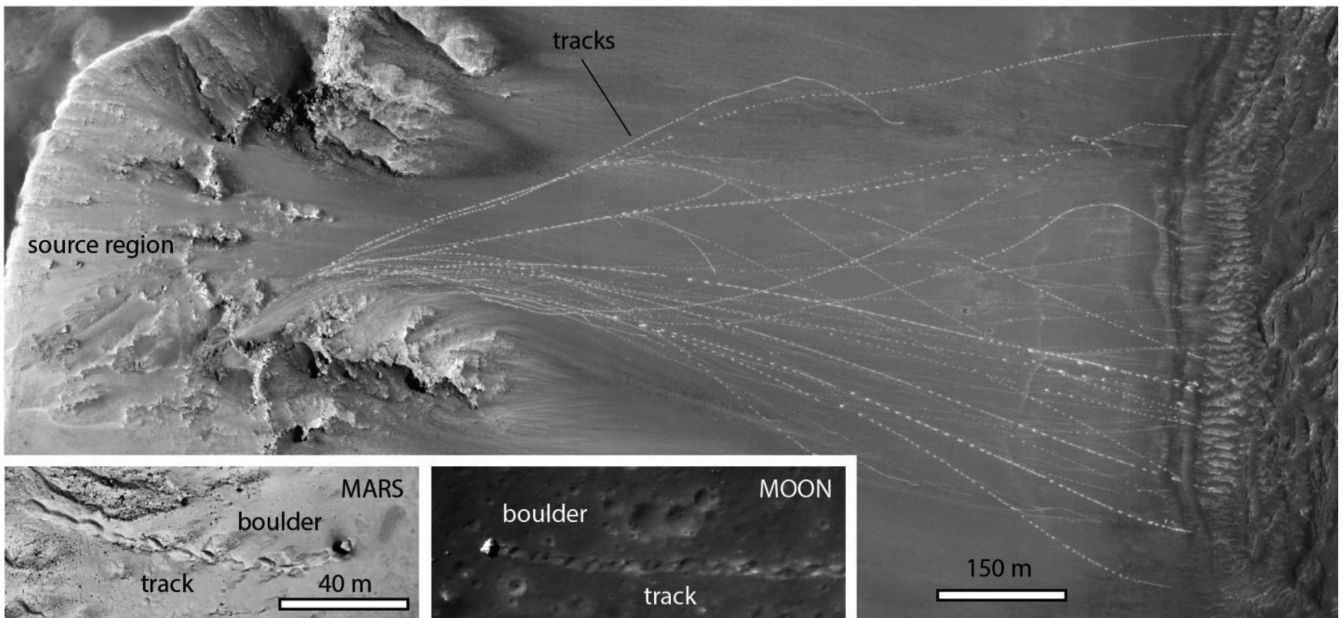


Fig. 1. Example of Martian and lunar rockfalls with their associated boulders and tracks (“feature of interest”). Overview image shows numerous rockfalls that detached from a bedrock outcrop near an impact crater rim on Mars. During their displacement, the boulders bounce and roll leaving tracks on the slope’s substrate. The downslope direction is towards the right. North is down. Detail of HiRISE image ESP_031280_1705. Bottom insets feature a close up of a Martian boulder with its track in HiRISE image ESP_039901_1665 and a lunar boulder with track in NAC image M138107059LC. Image credits: NASA/JPL/UofA/LROC/ASU.

Here, we describe how we implemented and applied a deep learning-based approach to automatically detect and map Martian rockfalls (boulders with tracks) in HiRISE imagery, to significantly enhance the statistical relevance of state-of-the-art geomorphological analyses on Mars. Deep learning-driven rockfall mapping in satellite imagery has first been demonstrated by [26] on the Moon, where other surface activity is limited. For the lunar environment, a convolutional neural network (CNN) was able to achieve near-human detection performance while reducing the required processing time (human versus machine) by more than one order of magnitude. On Mars, [27] developed a method to automatically detect ice falls in HiRISE images of the northern polar caps, using support vector machines. Due to its atmosphere and dynamic geologic environment, Mars generally poses additional challenges for the detection and mapping of rockfalls from orbit.

Combining experience from previous work with a novel multidomain learning approach, this study shows how deep learning and off-the-shelf CNNs can be used to extract new and valuable information from underexploited, but challenging Martian datasets, given the limitations of the available data. In addition, this study is relevant in light of spacecraft autonomy, where similar neural networks and infrastructures could be deployed to optimize mission procedures, e.g., by using AI for onboard selection of image data to improve utilization of the available data transfer bandwidth from remote regions of the solar system.

This article is organized as follows. Section I-A introduces Martian rockfalls and their physical appearance in high-resolution images taken by HiRISE. Next, the available and used datasets are described (Section I-B), followed by a description of

the applied deep learning approach (Section II). Finally, results are presented and discussed (Sections III and IV), with a focus on the limitations of the developed method and its value for the exploration of the red planet. Section V concludes this article.

A. Martian Rockfalls

Repeat coverage and high-resolution images (better than 0.5 m/pix—[24]) have revealed terrestrial-like rockfall (or boulder fall) present-day activity on Mars. Images show meter-scale blocks (here called boulders) which have detached from a bedrock cliff or boulder field, and have rolled/bounced downslope, leaving a visible track on the surface behind them (see Fig. 1). As aeolian processes are active on Mars (e.g., [1]), tracks left by rockfalls infill or erode over time, meaning, visible tracks are recent, forming during the Late Amazonian period (conservatively the last few tens of millions of years). They have been found on the walls of long, linear fissures (e.g., Cerberus Fossae—[7]), valleys (e.g., Valles Marineris—[8]), as well as impact craters [9].

Rockfalls can be produced by weathering (e.g., solar-induced thermal stress—[25]) followed by fracture propagation in the exposed rock (i.e., fatigue) at the top of a slope. In impact craters, loose blocks could also be formed by impact events (ejecta blocks) or could detach from ejecta blankets that have previously been emplaced on the crater walls [28]. Once formed, boulders move under the influence of gravity, while their displacement can also be driven by external factors such as nearby marsquakes or meteoritic impacts.

Recently, the spatial distribution of boulder tracks has been used to highlight the possibility of recent marsquakes in potentially active regions (e.g., [7]). As energy is increased closer to a seismic event, rockfall frequency and size should also increase [29]. On a larger scale, spatial distribution of rockfalls can be used to study active weathering processes responsible for rock breakdown on Mars (e.g., [25]), as well as to estimate weathering rates. Besides their implications for weathering processes and seismic activity of Mars, rockfalls and their tracks can be used to infer the mechanical properties of regolith, as demonstrated on the Moon by, e.g., [30] during the Apollo era and recently by [22] and [23], directly informing future surface exploration efforts. A deep learning-enabled global study of rockfall distribution on Mars could provide valuable information for a wide variety of relevant scientific and exploration-related applications.

B. Available Datasets

The surface of Mars has been imaged and is currently being imaged by numerous spacecrafts. While many instruments provide datasets with near-global coverage, such as ESA's Mars Express High Resolution Stereo Camera (HRSC, ~ 15 m/pix), and MRO's Context Camera (CTX, ~ 6 m/pix), their spatial resolution is insufficient to detect and map small geomorphological features such as boulders with tracks. The highest resolution is currently offered by MRO's HiRISE camera that takes subnadir images in the range from 0.25 to 0.75 m/pix with swath widths of about 6 km [24]. HiRISE is a pushbroom sensor comprised of 14 CCDs that provides data products with information in three different channels, covering the spectral range from 400 to 1000 nm (visual to near infrared wavelength). However, full multispectral coverage is only available for the central portion of each swath (roughly a 1.2-km wide footprint on the ground, i.e., approximately one-fifth of the entire image). The rest of the image is only available in a single channel, the red channel (550 to 850 nm). As the scope of this work is the detection and mapping of rockfalls on a large spatial scale; the utilization of imagery with maximum ground coverage has the highest priority. For this reason, network training and inference has been limited to the red channel, i.e., the full image swath. It has to be noted that despite its superior spatial resolution, HiRISE has not yet been able to cover a large portion of the Martian surface (around 3.56% as of February 14th 2020) due to its small footprint. Any analysis of rockfalls is therefore limited to regions that have been observed by the camera—for humans as for a CNN. Despite the reduced coverage, numerous targets of high scientific value across all regions and latitudes have been imaged, as illustrated by Fig. A1 in the Appendix. HiRISE's coverage might not be complete, but most geomorphological contexts are represented in the data stack.

We used nonmap projected, full-resolution .jp2 HiRISE images for this work. These data have undergone standard radiometric calibration routines and the data from the 9 or 10 red channel CCDs have been stitched to produce a single rectangular image. The perfect rectangular, upright orientation of nonmap projected HiRISE images helps to minimize the number of tiles when an image is cut into smaller pieces to fit into the RAM

of the used GPUs, directly optimizing processing speeds. They have the additional advantage that because no resampling of the data has been performed in order to map-project the data, these images represent the highest fidelity and resolution products. By using the available metadata for these images, in particular the coordinates of the image corners, the real-world coordinates of the detections in an image can be reconstructed using their respective image coordinates. Map-projected and, thus, rotated images result in more and potentially empty (NaN) tiles, larger file sizes, and consequently, longer processing times, which is not optimal for machine learning applications. The used nonmap projected .jp2s have been reduced to 8bit to further improve processing speeds and to enable the utilization of off-the-shelf machine learning toolboxes and libraries. One limitation of the nonmap projected products is that the stitching of these products is performed considering only the spacecraft pointing information and no postprocessing is done to improve the match between the CCDs in the overlapping area (as is done for the map-projected products). This means that there are visible seams caused by CCD mismatch on these products.

Besides image data, the laser altimeter (MOLA) onboard the Mars Global Surveyor (MGS) provides information about the Martian topography and relief. The United States Geological Survey (USGS) combined more than 600 million laser pulses in a global map product with a resolution of 463 m/pix [31]. Using MOLA and HRSC data, [32] blended a new DEM with a resolution of 200 m/pix. We include both products as potential auxiliary data for this study.

II. METHODS

A. Deep Learning-Driven Object Detection

Recent advances in machine and deep learning have overcome some of the issues in computer and machine vision, such as the detection and classification of objects in images. An important milestone was the development of two-stage region-based CNNs (R-CNNs—[33]), and their subsequent improvements, i.e., fast and faster R-CNNs [33]–[35]. Recently, [36] introduced the so-called residual networks (ResNets) that solved the problem of vanishing gradients during training, enabling deeper and more powerful networks. Then, [37] established a much faster single-stage architecture (RetinaNet) while adding a new loss function (focal loss) that improves the training efficiency by minimizing the weight assigned to well-classified examples. RetinaNet consists of two main networks, a ResNet for deep feature extraction and a feature pyramid network (FPN [38]) for the construction of rich, multiscale convolutional feature pyramids, as well as two subnetworks, an anchor classification network and an anchor regression network [37]. RetinaNet is outperforming other off-the-shelf state-of-the-art detectors in terms of classification accuracy and speed [37], is robust, and well established [39]–[41]. The combination of simplicity, reliability, recognition, speed, and performance is the reason why we chose RetinaNet for this study. We run RetinaNet with both ResNet50 and ResNet101 backbones [36], where the number refers to the layers of the network, i.e., its depth. In theory, a deeper neural network would yield better results, see, e.g., [42]. We

used RetinaNet—with open-source software only—in Python3, using Keras (v2.2.0, RetinaNet implementation by Fizr—[43]) and TensorFlow (v1.7.0). For the training and inference, we used two GPU-enabled desktop computers (NVIDIA GeForce GTX 1080 Ti and GeForce Titan Xp).

B. RetinaNet Labeling, Training, and Testing

Labels for training have been created by three experienced human operators (with domain knowledge) following the workflow of [26], using 136 full-swath HiRISE images with varying spatial resolutions and solar incidence angles, taken over different Martian regions (see Table A1 and Fig. A2 in the Appendix). Each label represents a classification and localization of a rockfall by a human operator that can be used by RetinaNet for training, validation, and testing. We placed the labels in a way to train the CNNs on the combination of boulder and track, to avoid detections of tracks without boulders and boulders without tracks (see Fig. A3 for details). The problem of how to label such irregular features, such as rockfalls and tracks, is interesting and is further discussed in Section IV. The used images are from a wide range of longitudes and latitudes and represent a variety of geomorphological contexts, i.e., crater slopes, fissures (graben), and valleys (see Fig. A2). We trained and tested six different CNNs. Their names indicate the origin of the main training data (Ma for Mars and M for Moon), the level of complexity of the training data (basic or complex), and the ResNet backbone depth (either $_{50}$ or $_{101}$). Four CNNs are exclusively trained with Martian image data while two CNNs are trained with a combination of Martian and lunar rockfall labels, where lunar labels were taken from [26], using NAC images [narrow angle camera, ptif (pyramid tif) single band 8bit, see [26]]. The level of complexity loosely describes the “difficulty” of the used training labels: “Basic” labels contain very clear examples of rockfalls with distinct tracks in a noncomplex environment (bare regolith); “Complex” labels feature more challenging environments such as debris fields. Examples of both levels are displayed in Fig. A4. All “Complex” networks utilize the labels from “Basic,” but further add complex examples, as described above. In addition, we included a series of “negative” examples (1024 labels) in the training data to highlight morphologic features that mimic the feature of interest, such as sand ripples, etc., but that must not be classified as rockfalls.

The six trained CNNs are: (I) Ma Basic $_{50}$, (II) Ma Basic $_{101}$, (III) Ma Complex $_{50}$, (IV) Ma Complex $_{101}$, (V) Ma/M Complex $_{50}$, and (VI) Ma/M Complex $_{101}$. We trained (I) and (II) with 2340 labels over 74 and 48 epochs; (III) and (IV) with 4969 labels over 85 and 66 epochs; (V) and (VI) with 13964 labels over 75 and 111 epochs. Here, (V) and (VI) have been trained with 4969 Martian and 8995 lunar labels, making (V) and (VI) the first cross-planet multicamera CNNs for rockfall detection. The term epoch describes one complete pass of all labels and images through the network. Training has been performed with RetinaNet’s focal loss function and a default 80–20 split for validation. At each epoch, the entire training data stack has been augmented using various affine transformations, such as image flipping and rotation, as well as up- and down-sampling

(see Fig. A3). Due to the augmentation, the effective number of training labels was (I) 173 160, (II) 112 320, (III) 422 365, (IV) 327 954, (V) 1 047 300, and (VI) 1 550 004. Training was stopped as soon as the regression of the loss stagnated after multiple learning rate reductions, i.e., as soon as the validation accuracy converged.

We tested the performance of all six CNNs with a test set of four full-swath HiRISE images—a total of 5048 image patches—taken over different geographic and geomorphic contexts: Zunil crater, a graben near Grjota Valles, a slope of a large crater in Terra Meridiani, and a crater in Syrtis Major (Table A2). The testing data have been labeled by an experienced human operator and testing data have not been used for training. The patches contain a total of 274 rockfalls. All testing images used have spatial resolutions ranging from 25 to 50 cm/pixel and solar incidence angles ranging from 51° to 59°. The vast majority of the used patches do not feature any rockfalls, representing a realistic testing scenario. A selection of the used patches is displayed in the Appendix (see Fig. A5) and a general example of rockfall detections is shown in Fig. 2.

For assessing the performance, we applied four different standard metrics: 1) recall (r); 2) precision (p); 3) average precision (AP); and 4) the F1 score. These metrics are calculated based on three numbers measured during testing: 1) true positives (TPs, correct detections); 2) false negatives (FNs, missed targets); and 3) false positives (FPs, wrong detections). All metrics describe different portions or aspects of CNN performance and can be used to quantify the general CNN detection behavior. Recall represents the percentage of rockfalls detected (e.g., 25/32 rockfalls have been detected). In the best case, recall would be 1, meaning that a CNN has found all rockfalls in the test set. Precision describes the percentage of rockfalls that have been detected and classified correctly (8/10 detections are correct). Precision is therefore a measure for the reliability of a CNN. AP is a metric that describes the stability of the relation between recall and precision along varying network confidence levels (here, the average of the precision at recall levels of 0 to 0.8, in 0.1 steps) and is therefore, a measure for the overall robustness of a CNN. The AP applied here is described by

$$AP = \frac{[p(r_0) + p(r_{0.1}) + \dots + p(r_m)]}{n}$$

where m is the maximum applicable/reached recall level (limited by the assessed networks’ performance) and n is the number of the applied steps or $p(r)$ values. The F1 score is the harmonic mean of the recall and precision and is another tool to assess the relation between the two inputs, particularly in the presence of high numbers of FNs.

We performed the entire assessment for varying network confidence thresholds (CT), i.e., the confidence of the CNN in a detection (the posterior probability output of the CNN). Usually, the application of lower CTs will result in more detections, increasing recall and decreasing precision. In turn, the application of higher confidences would result in lower recall and higher precisions. For this study, CT values ranging from 0.2 to 0.8 have been investigated. The applied intersection over union value ($\text{IoU}_{\text{testing}}$) for testing has been defined as

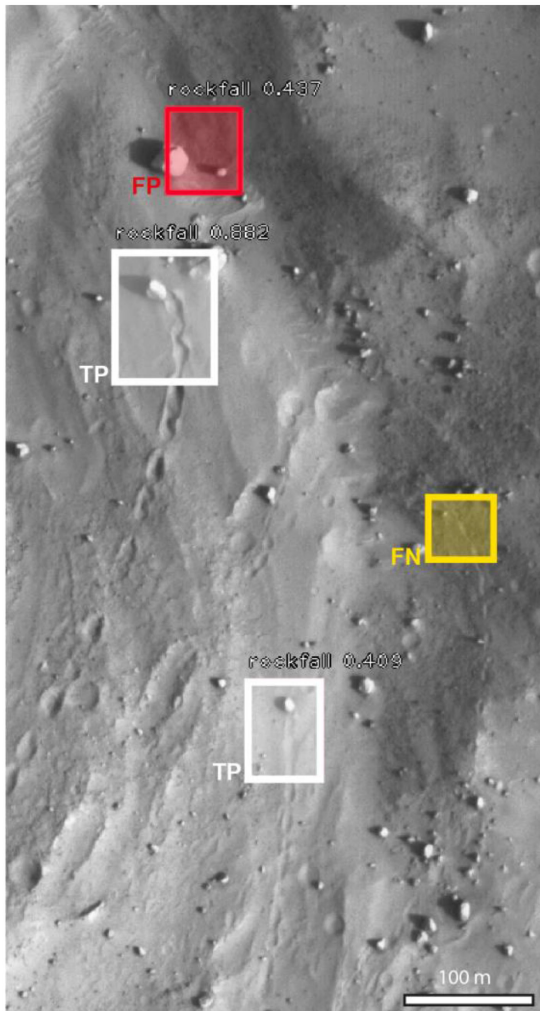


Fig. 2. Rockfall detections using Ma/M Complex₅₀: One high (88.2%) and one low (40.9%) confidence TP (white boxes), one FN (yellow box), and one low confidence FP (43.7%) (red box). Note that the general potential for FPs is substantial due to the amount of morphology that mimics the feature of interest (boulder with track), like sand ripples, other aeolian features, boulders without tracks, craters, and shadows. The downslope direction is towards the top. Please refer to the web version of this article to view color figures. Detail of HiRISE image ESP_036624_1720. Image credits: NASA/JPL/UofA.

0.5. $\text{IoU}_{\text{testing}}$ describes the required overlap between a CNN predicted bounding box and the labeled ground truth to count a detection as a TP during testing. Where required, a standard nonmaximum suppression filter (NMS) adapted from [44] has been applied to remove double detections. We used this NMS with $\text{IoU}_{\text{inference}} = 0.3$ during inference. Depending on the application, a user might be interested in optimizing a CNN for either maximum recall (missing fewer targets, while having many FPs), for maximum precision (few mistakes, even if many targets are missed), or for a compromise between the two extremes.

C. Domain Generalization and Multidomain Learning

One of the main limitations of supervised learning is the access to large quantities of high-quality labels that adequately

describe the feature of interest (rockfall). As lunar and Martian rockfalls share a common overall physical appearance (see Fig. 1) and as there are large numbers of labels (8995) available from a previous study of rockfall detection on the Moon [26], we included them for the training of two CNNs, (V) and (VI). The subtle differences in the appearance of rockfalls as well as in the radiometric properties of the images (HiRISE and NAC) could be expected to help the resulting CNNs to better generalize, potentially minimizing any present bias, e.g., related to a geomorphic setting or to label background. In addition, a better domain generalization might enable a hybrid-CNN to achieve a more robust performance (increased AP). We investigated whether the result of such a multidomain learning approach is beneficial for the detection performance of a CNN or rather reduces it.

III. RESULTS

The performance of one of the CNNs (Ma/M Complex₅₀) is visualized for one exemplary image (see Fig. 2), showing two TPs, one FN, and one FP. More examples of detections in more complex environments, such as dune fields, debris fields, and slopes with cross-cutting rockfall tracks, are shown in Figs. A6 and A7 in the Appendix. Average processing time per HiRISE nonmap projected image (i.e., ~ 950 patches) is around 45 s using a Titan Xp. Performance of all investigated neural networks for all used confidence thresholds is summarized in Table I and visualized in Fig. 3. Lower CTs generally produce higher recall values but lower precisions, while higher CTs generally produce lower recall values but higher precisions (as expected). The “Basic” CNNs (I) and (II) achieve very high average precisions (0.91 and 0.97) but to the cost of a significantly reduced recall. Fig. 3 suggests that deeper networks (ResNet 101) generally produce a lower number of FPs, but to the cost of a slightly reduced recall. Interestingly, multidomain learning with Martian and lunar data appears to potentially help the CNNs (V) and (VI) to better generalize, improving the detection precision, particularly for higher CT values. However, the potentially improved generalization does not appear to improve the recall of the respective CNNs. The performance of all tested CNNs is showcased for one test patch in Fig. 4. (bounding boxes), while Figs. 5 and A8 report the results of the testing on full-swath HiRISE image for a graben in the Grjota Valles area and a slope in Terra Meridiani, where TPs and FPs are displayed as points. To place the performance of CNN (VI) in context, Fig. 5 additionally shows all detections that were made by an experienced human operator.

IV. DISCUSSION

Testing of all trained networks demonstrates the potential of automated CNN-based mapping of Martian rockfalls (boulders with tracks) in full HiRISE swaths and on large spatial scales. The CNNs’ processing speed is about 60–100 times faster than that of a human operator (for the tested images). The performance of the trained and implemented CNNs in combination with the associated processing times allow for the utilization of the entire HiRISE image archive to rapidly and effectively scan

TABLE I
PERFORMANCE RESULTS FOR ALL TRAINED AND TESTED CNNs: MA BASIC₅₀ AND MA BASIC₁₀₁ (MARS), MA COMPLEX₅₀ AND MA COMPLEX₁₀₁ (MARS), AS WELL AS MA/M COMPLEX₅₀ AND MA/M COMPLEX₁₀₁ (MARS AND MOON); IOU IS 0.5 (SHOWN IN THE HEADER AS IOU₅₀)

CT	Ma Basic ₅₀				Ma Basic ₁₀₁			
	AP ₅₀ (n=2)	Recall ₅₀	Precision ₅₀	F1	AP ₅₀ (n=2)	Recall ₅₀	Precision ₅₀	F1
0.2	0.905	0.11	0.79	0.19	0.97	0.14	0.93	0.24
0.3		0.11	0.79	0.19		0.11	0.91	0.20
0.4		0.08	0.88	0.14		0.10	0.93	0.18
0.5		0.07	0.90	0.13		0.08	0.92	0.15
0.6		0.06	0.89	0.12		0.07	0.95	0.13
0.7		0.01	1.00	0.03		0.05	0.94	0.10
0.8		0.00	1.00	0.01		0.02	1.00	0.04

CT	Ma Complex ₅₀				Ma Complex ₁₀₁			
	AP ₅₀ (n=9)	Recall ₅₀	Precision ₅₀	F1	AP ₅₀ (n=7)	Recall ₅₀	Precision ₅₀	F1
0.2	0.51	0.78	0.29	0.42	0.65	0.60	0.36	0.45
0.3		0.45	0.37	0.41		0.41	0.52	0.46
0.4		0.31	0.47	0.37		0.31	0.62	0.41
0.5		0.20	0.62	0.31		0.22	0.75	0.34
0.6		0.14	0.74	0.24		0.13	0.78	0.22
0.7		0.07	0.83	0.13		0.06	0.76	0.11
0.8		0.03	1.00	0.05		0.01	1.00	0.02

CT	Ma/M Complex ₅₀				Ma/M Complex ₁₀₁			
	AP ₅₀ (n=8)	Recall ₅₀	Precision ₅₀	F1	AP ₅₀ (n=7)	Recall ₅₀	Precision ₅₀	F1
0.2	0.555	0.70	0.21	0.32	0.68	0.61	0.39	0.48
0.3		0.46	0.35	0.40		0.41	0.52	0.45
0.4		0.30	0.51	0.38		0.31	0.60	0.41
0.5		0.20	0.72	0.32		0.24	0.73	0.36
0.6		0.13	0.95	0.23		0.19	0.84	0.31
0.7		0.07	0.95	0.13		0.12	0.89	0.21
0.8		0.01	1.00	0.02		0.07	0.95	0.14

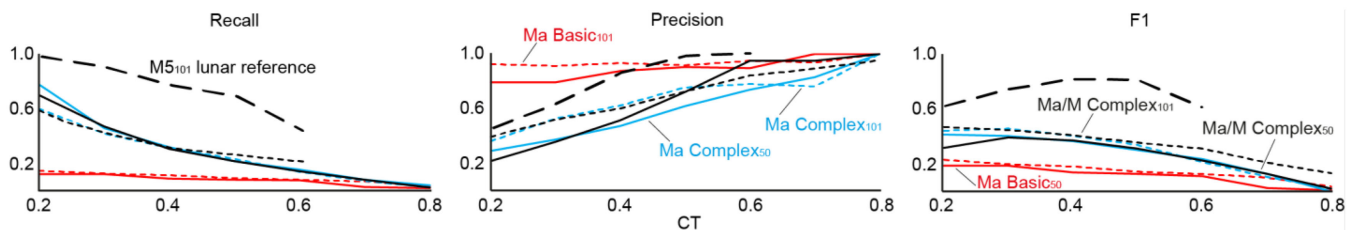


Fig. 3. Graphical visualization of the networks' performance, from left to right: Recall, precision, and F1 score. The colors indicate the different CNNs and the dashed line represents a cross-planet reference derived with M5₁₀₁ (a lunar network) on lunar NAC data, see [26]. The x-axis represents the confidence threshold (CT). Please refer to the web version of this article to view color figures.

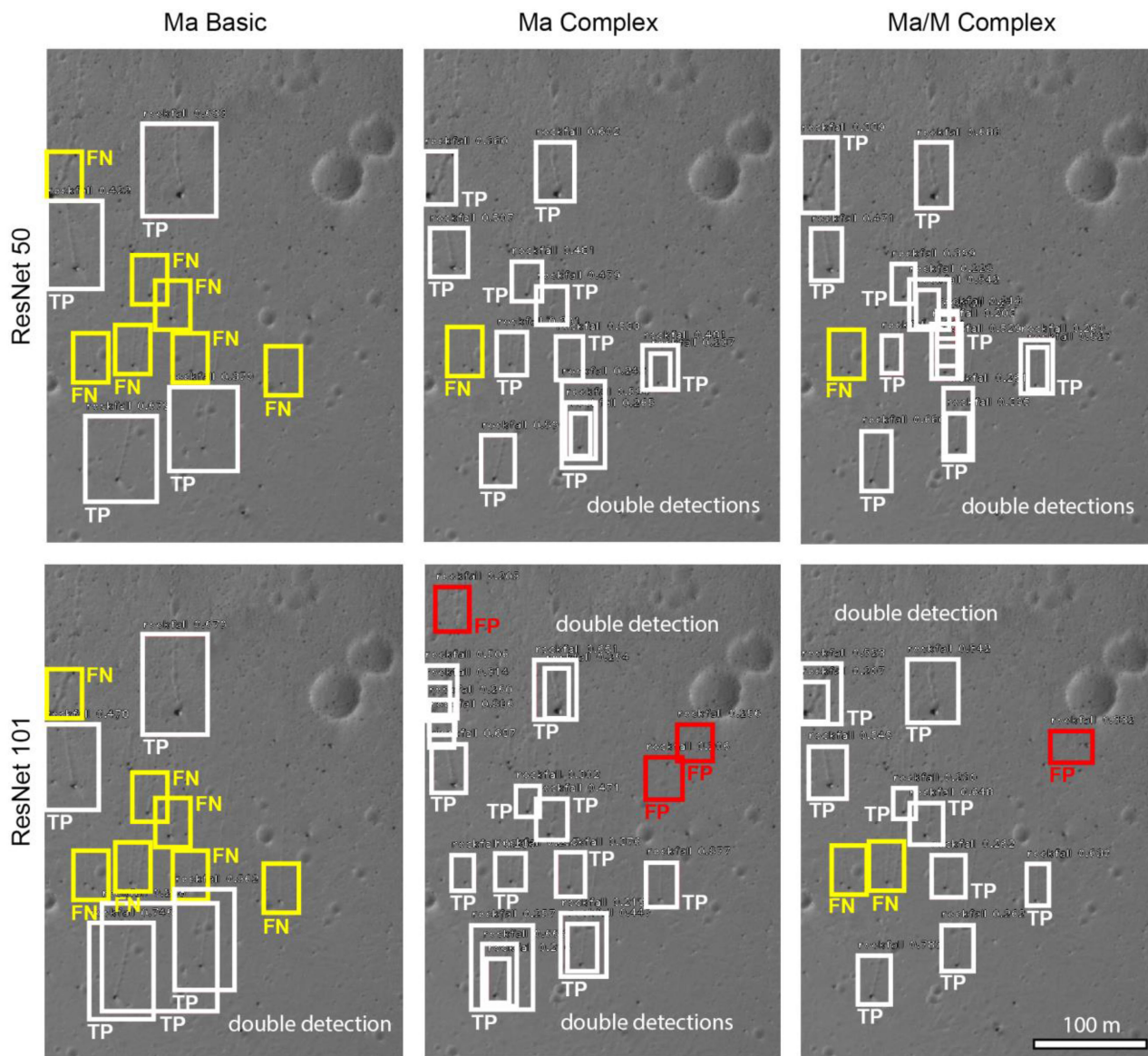


Fig. 4. Visual, exemplary comparison of CNN performance, showing detections with CTs > 0.2. The Ma Basic networks draw larger bounding boxes and have a significantly lower recall. Please refer to the web version of this article to view color figures. Detail of HiRISE image ESP_03770_1710. Image credits: NASA/JPL/UoA.

for rockfalls on a global scale—and through newly acquired data, as MRO’s mission continues.

Interestingly, the utilization of “Basic” and more “Complex” rockfall labels is clearly represented in the performance (e.g., Fig. 3). In general, the CNNs trained with more diverse—and potentially more problematic or less clear—labels achieve significantly higher recall values, but appear to be more easily confused by features that mimic rockfalls. In turn, the newly implemented utilization of labels from two different planetary bodies (domains) appears to enable the respective networks (V) and (VI) to counter the drawback of a reduced precision, as encountered by the “Complex” single-domain networks (III) and (IV). Besides the training on multidomain labels, the generally increased number of labels could have a positive impact on the

overall performance as well. We will continue to systematically investigate these observations in future work to explore and exploit the potential of multiplanetary and multisensor (rockfall) detectors (multidomain learning).

The placement of feature labels, i.e., their position and size, for training and testing is a challenging problem, as rockfalls have a highly variable appearance and as the background within each bounding box is part of the label itself, although it might not represent the actual feature of interest. In order to avoid confusion with static boulders and tracks without boulders, we draw relatively large boxes to capture the entire rockfall (boulder with associated track). It has to be noted that these relatively large boxes also result in an increased background content (as shown in Fig. A3), such as small boulders, sand ripples, and

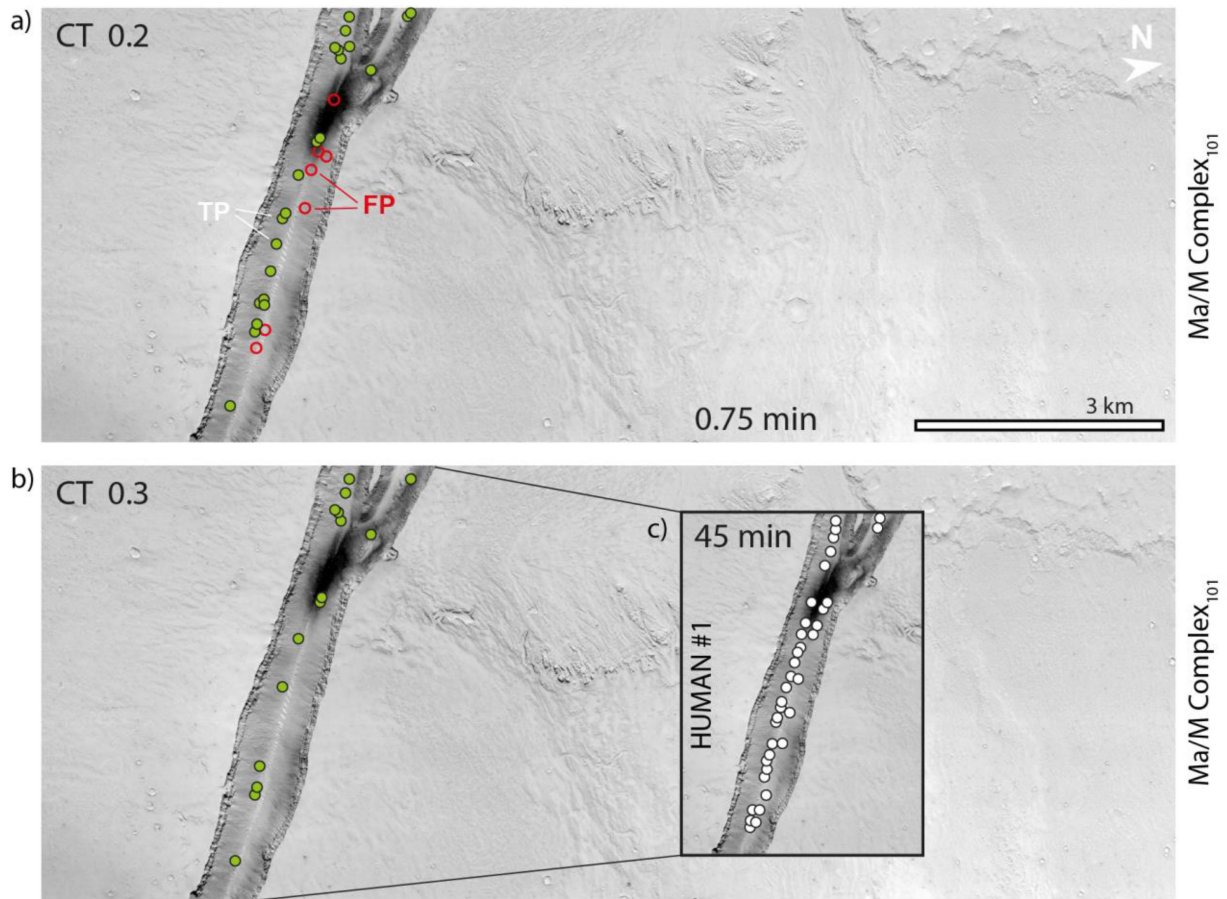


Fig. 5. CNN-driven rockfall mapping with a full-scene HiRISE image in a graben (Grjota Valles area) by Ma/M Complex₁₀₁ with (a) CT 0.2 and (b) 0.3. Rockfalls mapped by a human operator are shown in the inset in (c) (white dots). Green dots are TPs, red indicates FPs. The increase of the CT by 0.1 effectively removes all FPs from the output, while slightly reducing the recall. The majority of FPs is related to aeolian deposits. Processing times are ~ 45 min (human) versus 45 s (CNN). Note that the human processing time would be much longer if not only the graben itself would have been searched. Full HiRISE image ESP_025156_1965. Image credits: NASA/JPL/UofA. Please refer to the web version of this article to view color figures.

other nonsignal features. In small datasets or in datasets with limited variation, this background can become dominant and can negatively influence the filter weights during training, making it harder for a CNN to identify the actual rockfall during inference (see “Basic” versus “Complex”). The potential negative influence of the increased background is very hard to quantify and remains an interesting subject for future studies with relevant applications for computer vision in general.

In contrast to rockfall mapping on the Moon, automated mapping on Mars is affected by the more dynamic Martian environment, particularly by its atmosphere. This is obvious from the difference in recall and precision for the lunar reference [26] and Martian detectors (this study), highlighted by Fig. 3. Aeolian processes deposit sand ripples and scour, occasionally in the lee of static boulders. As sand ripples tend to be rather linear features, the combination of boulder and ripple can mimic the appearance of a rockfall with track, as shown in Fig. A9. During training, we took care to include a significant number of negative “boulder and ripple” examples to sensitize the CNNs for the subtle differences in their physical appearance. The vast majority of aeolian FPs have very low confidence scores, usually lower than 0.3 (see Fig. A9), which facilitates their removal (e.g.,

in Fig. 5). Occasionally, some CNNs assign higher confidence scores (~ 0.4 to ~ 0.6) to aeolian FPs by mistake, which is reflected in the nonperfect (i.e., < 1.0) precision scores of some CNNs in the respective CT range (~ 0.4 to ~ 0.6). An additional limitation is created by the spatial resolution of the used HiRISE images and by the size of the target features, respectively: When GSD (ground sampling distance) and target size converge, detection and classification ambiguities are the result—affecting CNNs in a similar way to experienced human operators. These ambiguities are a fundamental problem of automated as well as human-based mapping and will always remain (although being “shifted in scale”), even if imagery with better spatial resolution becomes available in the future.

In order to enhance the overall mapping performance and reliability of the trained detectors, an additional postprocessing step could be implemented using auxiliary MOLA DEM-derived slope angle maps (e.g., based on [31] and [32]). As rockfalls are features that are usually connected to topography gradients, the slope maps could be used to discard detections in flat terrain, which are likely FPs caused by aeolian processes. However, MOLA-guided postfiltering would need to consider: 1) the spatial resolution of the underlying DEMs (200 and 463 m/pix) that

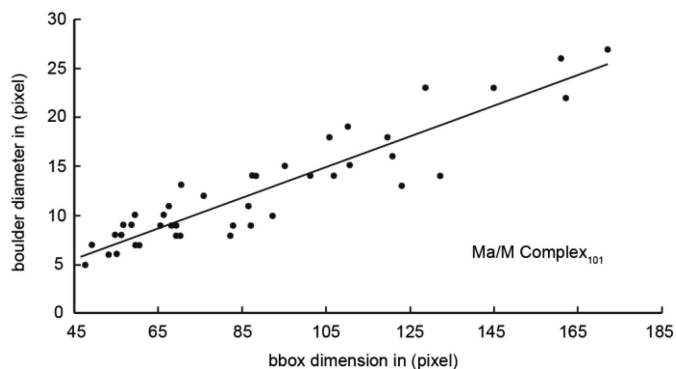


Fig. 6. Relation of the CNN-predicted bounding box diameter in pixel (x -axis) and the respective boulder diameter in pixel (y -axis). The linear relation is highlighted with a least square fit (R^2 is 0.851). This first-order approximation can be applied to estimate rockfall diameters in (meters), using the bounding box diameter in combination with the spatial resolution of the respective HiRISE image.

could misrepresent slope conditions and angles of very small topographic features; 2) local inconsistencies in the DEMs; and 3) the user-defined cutoff slope angle (rockfall area versus nonrockfall area).

Fully automated mapping of Martian boulders with tracks, either as stand-alone process or in combination with MOLA-based postfiltering, is feasible and will produce reliable map products, depending on the selected network and CT criteria. It has to be kept in mind that the produced rockfall distribution maps might not be representative of the entire Martian rockfall population due to the limited recall (<1.0) of all CNNs. Still, their output can serve as big data foundation for statistical analyses and can directly inform and focus future mapping efforts that can perform follow up investigations in greater detail. For example, human operators could focus their attention on areas highlighted by the CNN, not scour the whole image. As the CNN-driven mapping proceeds, the neural networks will collect more and more rockfall examples (labels), which will help us to train new generations of CNNs with enhanced capabilities, increasing the value of the produced output as time passes.

As observed in a previous study [26], the diameter of the CNN-placed bounding boxes is related to the physical diameter of the detected boulders (see Fig. A4). This relation has been quantified with the testing dataset and is plotted in Fig. 6. This additional information further enriches the value of the produced data, as it enables, e.g., rockfall magnitude and spatial frequency analyses, which are useful to investigate the relation between rockfall abundance and proximity to tectonic features, among others. It has to be noted that this relation strongly depends on the quality and consistency of the labels placed by the human operators and only represents a first-order approximation.

V. CONCLUSION

This study demonstrates the feasibility of using off-the-shelf CNNs in combination with a novel multidomain learning approach to efficiently extract new and yet unexplored information from the growing HiRISE big data stack. The trained CNNs are

capable of mapping Martian rockfalls orders of magnitude faster than a human operator, and achieve inference performances that can be deployed to produce statistically relevant products.

Testing of six different CNNs showed that training data complexity and network depth directly affect the detection performance, where “Complex” CNNs achieve significantly higher recall, but slightly reduced precision values, potentially caused by the more complex label background. The application of deeper networks, such as a ResNet 101, has proven to generally benefit the average precision of all tested CNNs. The novel application of multidomain learning principles, using a combination of rockfall labels from the Moon and Mars, resulted in an apparent, further improvement of detection precision values, particularly for higher confidence scores. In summary, deep multidomain networks trained on more complex data achieved the peak overall performance in this study. Future work will elaborate on the beneficial impact of multidomain training and domain generalization, as well as on the influence of label size, shape, and background.

MOLA slope map-based postfiltering could be used to effectively remove FPs and to further improve the reliability of the CNN output for subsequent applications. The trained and tested CNNs represent a capable tool that will support the scientific community in characterizing Mars’ past and current endogenic and exogenic activity, among other highly relevant research questions.

APPENDIX

Supplementary information is shown in a separate file.

ACKNOWLEDGMENT

The authors would like to thank E. Law, N. Gallegos, and the Solar System Treks team at NASA JPL and NASA Ames for their support and the offer to implement this tool in Mars Trek. The authors would like to thank S. Kumar for sharing a list of HiRISE images that contain rockfalls. Data can be made available upon request to the authors. The workflow is entirely based on open source software. The developed tool will eventually be accessible online on the Mars Trek platform that is being developed by NASA JPL at Caltech, USA.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

- [1] M. C. Bourke, K. S. Edgett, and B. A. Cantor, “Recent aeolian dune change on Mars,” *Geomorphology*, vol. 94, pp. 247–255, 2008.
- [2] N. T. Bridges *et al.*, “Earth-like sand fluxes on Mars,” *Nature*, vol. 485, no. 7398, pp. 339–342, 2012.
- [3] M. Balme and R. Greeley, “Dust devils on Earth and Mars,” *Rev. Geophys.*, vol. 44, no. 3, 2006.
- [4] R. D. Lorenz and D. Reiss, “Solar panel clearing events, dust devil tracks, and in-situ vortex detections on Mars,” *Icarus*, vol. 248, pp. 162–164, 2005.
- [5] P. Russell *et al.*, “Seasonally active frost-dust avalanches on a north polar scarp of Mars captured by HiRISE,” *Geophys. Res. Lett.*, vol. 35, no. 23, 2008, Art. no. L23204.
- [6] N. Schorghofer, O. Aharonson, and S. Khatiwala, “Slope streaks on Mars: Correlations with surface properties and the potential role of water,” *Geophys. Res. Lett.*, vol. 29, pp. 41-1–41-4, 2002.

- [7] G. P. Roberts, B. Matthews, C. Bristow, L. Guerrieri, and J. Vetterlein, "Possible evidence of paleomarsquakes from fallen boulder populations, Cerberus Fossae, Mars," *J. Geophys. Res., Planets*, vol. 117, 2012, Art. no. E02009.
- [8] P. S. Kumar *et al.*, "Recent seismicity in Valles Marineris, Mars: Insights from young faults, landslides, boulder falls and possible mud volcanoes," *Earth Planet. Sci. Lett.*, vol. 505, pp. 51–64, 2019.
- [9] T. de Haas, S. J. Conway, and M. Krautblatter, "Recent (Late Amazonian) enhanced backweathering rates on Mars: Paracratering evidence from gully alcoves," *J. Geophys. Res., Planets*, vol. 120, no. 12, pp. 2169–2189, 2015.
- [10] O. Aharonson, N. Schorghofer, and M. F. Gerstell, "Slope streak formation and dust deposition rates on Mars," *J. Geophys. Res., Planets*, vol. 108, no. E12, 2003, Art. no. 5138.
- [11] N. Schorghofer, O. Aharonson, M. F. Gerstell, and L. Tatsumi, "Three decades of slope streak activity on Mars," *Icarus*, vol. 191, no. 1, pp. 132–140, 2007.
- [12] E. S. Brusnikin *et al.*, "Topographic measurements of slope streaks on Mars," *Icarus*, vol. 278, pp. 52–61, 2016.
- [13] C. M. Dundas *et al.*, "Granular flows at recurring slope lineae on Mars indicate a limited role for liquid water," *Nature Geosci.*, vol. 10, no. 12, pp. 903–907, 2017.
- [14] F. Schmidt, F. Andrieu, F. Costard, M. Kocifaj, and A. G. Meresescu, "Formation of recurring slope lineae on Mars by rarefied gas-triggered granular flows," *Nature Geosci.*, vol. 10, no. 4, pp. 270–273, 2017.
- [15] D. E. Stillman *et al.*, "Evaluation of wet and dry recurring slope lineae (RSL) formation mechanisms based on quantitative mapping of RSL in Garni Crater, Valles Marineris, Mars," *Icarus*, vol. 335, 2019, Art. no. 113420.
- [16] K. Pasquon, J. Gargani, M. Massé, and S. J. Conway, "Present-day formation and seasonal evolution of linear dune gullies on Mars," *Icarus*, vol. 274, pp. 195–210, 2016.
- [17] J. Raack, D. Reiss, T. Appéré, M. Vincendon, O. Ruesch, and H. Hiesinger, "Present-day seasonal gully activity in a south polar pit (Sisyphi Cavi) on Mars," *Icarus*, vol. 251, pp. 226–243, 2015.
- [18] E. Gardin, P. Allemand, C. Quantin, and P. Thollot, "Defrosting, dark flow features, and dune activity on Mars: Example in Russell crater," *J. Geophys. Res., Planets*, vol. 115, no. E6, 2010, Art. no. E06016.
- [19] J. Taylor, N. A. Teanby, and J. Wookey, "Estimates of seismic activity in the Cerberus Fossae region of Mars," *J. Geophys. Res., Planets*, vol. 118, pp. 2570–2581, 2013.
- [20] P. S. Kumar *et al.*, "Recent shallow moonquake and impact-triggered boulder falls on the Moon: New insights from the Schrödinger basin," *J. Geophys. Res., Planets*, vol. 121, pp. 147–179, 2016.
- [21] P. S. Kumar *et al.*, "The seismically active lobate scarps and coseismic lunar boulder avalanches triggered by 3 Jan. 1975 (MW 4.1) shallow moonquake," *Geophys. Res. Lett.*, vol. 46, no. 14, pp. 7972–7981, 2019.
- [22] V. T. Bickel *et al.*, "Analysis of lunar boulder tracks: Implications for tractability of pyroclastic deposits," *J. Geophys. Res., Planets*, vol. 124, no. 5, pp. 1296–1314, 2019.
- [23] H. S. Sargeant *et al.*, "Using boulder tracks as a tool to understand the bearing capacity of permanently shadowed regions of the moon," *J. Geophys. Res., Planets*, vol. 125, no. 2, 2020, Art. no. e2019JE006157.
- [24] A. S. McEwen *et al.*, "Mars Reconnaissance Orbiter's high resolution imaging science experiment (HiRISE)," *J. Geophys. Res., Planets*, vol. 112, 2007, Art. no. E05S02.
- [25] P.-A. Tesson *et al.*, "Evidence for thermal-stress-induced rockfalls on Mars impact crater slopes," *Icarus*, vol. 342, 2020, Art. no. 113503.
- [26] V. T. Bickel, C. Lanaras, A. Manconi, S. Loew, and U. Mall, "Automated detection of lunar rockfalls using a convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3501–3511, Jun. 2018.
- [27] L. Fanara, K. Gwinner, E. Hauber, and J. Oberst, "Automated detection of block falls in the north polar region of Mars," *Planet. Space Sci.*, vol. 180, Jan. 2020, Art. no. 104733.
- [28] N. Krishna and P. S. Kumar, "Impact spallation processes on the Moon: A case study from the size and shape analysis of ejecta boulders and secondary craters of Censorinus crater," *Icarus*, vol. 264, pp. 274–299, 2016.
- [29] D. K. Keefer, "Landslides caused by earthquakes," *Geol. Soc. Amer. Bull.*, vol. 95, pp. 406–421, 1984.
- [30] H. J. Moore, "Estimates of the mechanical properties of lunar surface using tracks and secondary impact craters produced by blocks and boulders," Nat. Aeronaut. Space. Admin., Washington, DC, USA, NASA Interagency Rep. Astrogeology 22, 1970.
- [31] MOLA Team, "Mars MGS MOLA DEM 463m v2," Astrogeol. PDS Annex, U.S. Geol. Surv., Reston, VA, USA, 2014. [Online]. Available: https://astrogeology.usgs.gov/search/map/Mars/GlobalSurveyor/MOLA/Mars_MGS_MOLA_DEM_mosaic_global_463m
- [32] R. L. Ferguson, T. M. Hare, and J. Laura, "HRSC and MOLA blended digital elevation model at 200m v2," Astrogeol. PDS Annex, U.S. Geol. Surv., Reston, VA, USA, 2017. [Online]. Available: http://bit.ly/HRSC_MOLA_Blend_v0
- [33] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 142–158, Jan. 2016.
- [34] R. Girshick, "Fast R-CNN," 2015, *arXiv:1504.08083*.
- [35] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," 2016, *arXiv:1506.01497*.
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.
- [37] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," 2017, *arXiv:1708.02002*.
- [38] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 936–944.
- [39] A. Douillard, "Object detection with deep learning on aerial imagery," *Medium*, 2018. [Online]. Available: <https://medium.com/data-from-the-trenches/object-detection-with-deep-learning-on-aerial-imagery-2465078db8a9>
- [40] K. R. Singh, "ESRI object detection challenge 2019," 2019. [Online]. Available: https://github.com/kunwar31/ESRI_Object_Detection
- [41] B. G. Weinstein, S. Marconi, S. Bohlmann, A. Zare, and E. White, "Individual tree-crown detection in RGB imagery using self-supervised deep learning neural networks," 2019. [Online]. Available: <https://www.mdpi.com/2072-4292/11/11/1309>
- [42] D. Eigen, J. Rolfe, R. Fergus, and Y. LeCun, "Understanding deep architectures using a recursive convolutional network," 2014, *arXiv:1312.1847*.
- [43] Fyzir, "Keras RetinaNet," 2020. [Online]. Available: <https://github.com/fyzir/keras-retinanet>
- [44] T. Malisiewicz, "Blazing fast nms.m," 2011. [Online]. Available: <http://www.computervisionblog.com/2011/08/blazing-fast-nmsm-from-exemplar-svm.html>
- [45] P. R. Christensen *et al.*, "JMARS— A planetary GIS," 2009. [Online]. Available: <http://adsabs.harvard.edu/abs/2009AGUFMIN22A..06C>

Valentin Tertius Bickel was born in 1990. He received the B.S. degree in geosciences from the Technical University in Munich and the Ludwig-Maximilians University Munich, Munich, Germany, performing his thesis related research at the University Centre in Svalbard, Longyearbyen, Norway, and the M.S. degree in engineering geology from the Swiss Federal Institute of Technology Zurich (ETH Zurich), Zurich, Switzerland. He has been working toward the Ph.D. degree with the Max Planck Institute for Solar System Research, Göttingen, Germany, and ETH Zurich, since 2017.

His research interests include neural networks for object and change detection, as well as planetary geomorphology and terramechanics.

Susan J. Conway was born in the U.K. She received the M.Sc. degree from the Earth Sciences Department, Cambridge University, Cambridge, U.K., and the Ph.D. degree in planetary science from the Open University, Milton Keynes, U.K.

She is currently with the French National Research Institute, the CNRS based at the Laboratoire de Planétologie et Géodynamique, University of Nantes, Nantes, France. She is a Guest Investigator on the European Space Agency Trace Gas Orbiter mission to Mars and Co-I on NASA's Mars Reconnaissance Orbiter's HiRISE instrument. Her research interests include planetary geomorphology, focusing on topographic and image analysis to understand the role of volatiles in planetary surface processes with a particular focus on the planet Mars.

Pierre-Antoine Tesson was born in 1993. He received the B.S. degree in geosciences from the University of Nantes, Nantes, France. He pursued studies in geomorphology and planetary geosciences with the University of Tromsø, Tromsø, Norway, and the University of Nantes. He received the M.S. degree in 2017, performing a thesis about rockfalls on Mars. He has been working toward the Ph.D. degree with the Space Research Centre PAS, Warsaw, Poland, since March 2018.

His research interests include planetary geomorphology and deformation and lava flow chronology in volcanic provinces on Mars.

Andrea Manconi was born in 1978. He received the M.S. degree in environmental engineering from the University of Cagliari, Cagliari, Italy, in 2005, and the Ph.D. degree (*Magna cum Laude*) in geophysics from the University of Potsdam, Potsdam, Germany, and the GeoForschungsZentrum Potsdam, Potsdam, Germany in 2009.

Between February 2009 and July 2015, he was a Researcher with the National Research Council, Italy, first with IREA-CNR and second, with IRPI-CNR. Since August 2015, he has been with ETH Zurich, Zurich, Switzerland, as a Research Assistant and Lecturer. Since May 2019, he has also been a Project Scientist with GAMMA Remote Sensing AG, Switzerland. His research interests include active volcanoes, seismogenetic areas, and unstable slope dynamics, investigated mainly through remote sensing techniques.

Simon Loew was born in 1956. He studied geology with the University of Basel, Basel, Switzerland, completing his dissertation in 1985.

In 1985, he joined an international consulting company, working in the area of energy and environmental technology, including large interdisciplinary projects related to the geological disposal of nuclear and toxic waste, large traffic tunnels, and natural hazards. In 1996, he became a Full Professor and Chair of Engineering Geology with ETH Zurich, Zurich, Switzerland. He directs the Major in Engineering Geology and the CAS in Applied Earth Sciences. His Chair has received much international attention for research and outreach related to alpine tunneling, landslides, deep geothermal energy, and nuclear waste disposal. His research interests include all hydromechanical aspects of fractured rocks at project relevant scales.

Urs Mall was born in Switzerland. He studied physics, mathematics, and astronomy with the University of Basel, Basel, Switzerland, completing his dissertation in 1989 in experimental particle physics.

In 1989, he joined the University of Bern, Bern, Switzerland, to work on mass spectrometry and solar wind research. He was a Research Associate with the University of Michigan, Ann Arbor, MI, USA, and at the University of Maryland, College Park, MD, USA, where he worked on solar-terrestrial missions. In 1998, he became a member of the Scientific Staff of the Max Planck Institute for Aeronomy, Lindau, Germany, developing instrumentation for magnetospheric missions. He is a PI and Co-I of numerous space missions. His current research with the Max-Planck Institute for Solar System Research, Göttingen, Germany, includes lunar and cometary research.