SCIENCE AUTONOMY TRAINING BY VISUAL SIMULATION

Virtual Conference 19–23 October 2020

G. Paar[1], C. Traxler[2], O. Sidla[3], A. Bechtold[4], C. Koeberl[5]

[1]JOANNEUM RESEARCH, Graz, Austria, gerhard.paar@joanneum.at
[2]VRVis, Vienna, Austria, traxler@vrvis.at
[3]SLR Engineering, Graz, Austria, os@slr-engineering.at
[4]Natural History Museum, Vienna, Austria, a.bechtold@gmx.at
[5]University of Vienna, Austria, christian.koeberl@univie.ac.at

ABSTRACT

Robotic space missions contain optical instruments for various mission-related and science tasks, such as 2D and 3D mapping, geologic characterization, atmospheric investigations, or spectroscopy for exobiology, the characterization of scientific context, and the identification of scientific targets of interest. The considerable variability of appearance of such potential scientific targets calls for well-adapted yet flexible techniques, one of them being *Deep Learning* (DL). Our "*Mars-DL*" (Planetary Scientific Target Detection via Deep Learning) approach focuses on **training for visual DL by virtual placement of known targets in a true context environment**. The 3D context environment is taken from reconstructions using true Mars rover imagery. Scientifically interesting objects, such as impact-characteristic shatter cones (SCs) from several terrestrial impact structures, and/or meteorites, are captured and 3D reconstructed using photogrammetric techniques, gaining a 3D data base (high resolution mesh and albedo map) of objects to be randomly placed in the realistic scenes. Using a powerful image rendering tool, the assembled virtual scenes deliver thousands of training data sets, which are used for data augmentation for the following Deep Learning assets.

So far the simulation components have been assembled and tested. We report on the current status and first results of training and inference using the simulated data sets as well as prospects of the approach.

1. MARS-DL SCOPE

The remarkable success of deep learning (DL) for object and pattern recognition suggests its application to autonomic target selection of future planetary rover missions. DL can support planetary scientists and also the exploring robots by preselecting possibly interesting regions in imagery. Thereby DL will increase the potential for scientific discoveries, and speed-up the strategic decision-making.

Deep learning requires large amounts of training data to work reliably. Many different geologic features are to be detected and to be trained for in a DL-system. Past and ongoing missions such as the Mars Science Laboratory (MSL) do neither provide the necessary volume of training data, nor existing "ground truth". Therefore, realistic simulations are required. The Mars-DL project follows this idea and intends to demonstrate the validity of using such simulations for a real science assessment case as being relevant for present and future planetary on-site exploration activities.

2. MARS-DL COMPONENTS

2.1 Overview

Figure 1 gives an overview of the system layout. It consists of the following components:

- Scientifically interesting objects to be virtually placed in the scene are chosen to be 3D reconstructions of shatter cones (SCs) and/or meteorites as representative set
- 3D background for the training & validation scenes is taken from 3D reconstructions using true Mars rover imagery (e.g. from the MSL Mastcam instrument)
- Simulation of rover imagery for training is based on PRo3D, a viewer for the interactive exploration and geologic analysis of high-resolution planetary surface reconstructions
- Representative yet random positions for the SCs and fields-of-view for the training images are chosen to produce thousands of rendered training images
- Training is performed on a DL framework (neural network)
- The trained DL mechanisms are usable for automatic indexing of Planetary surface images.

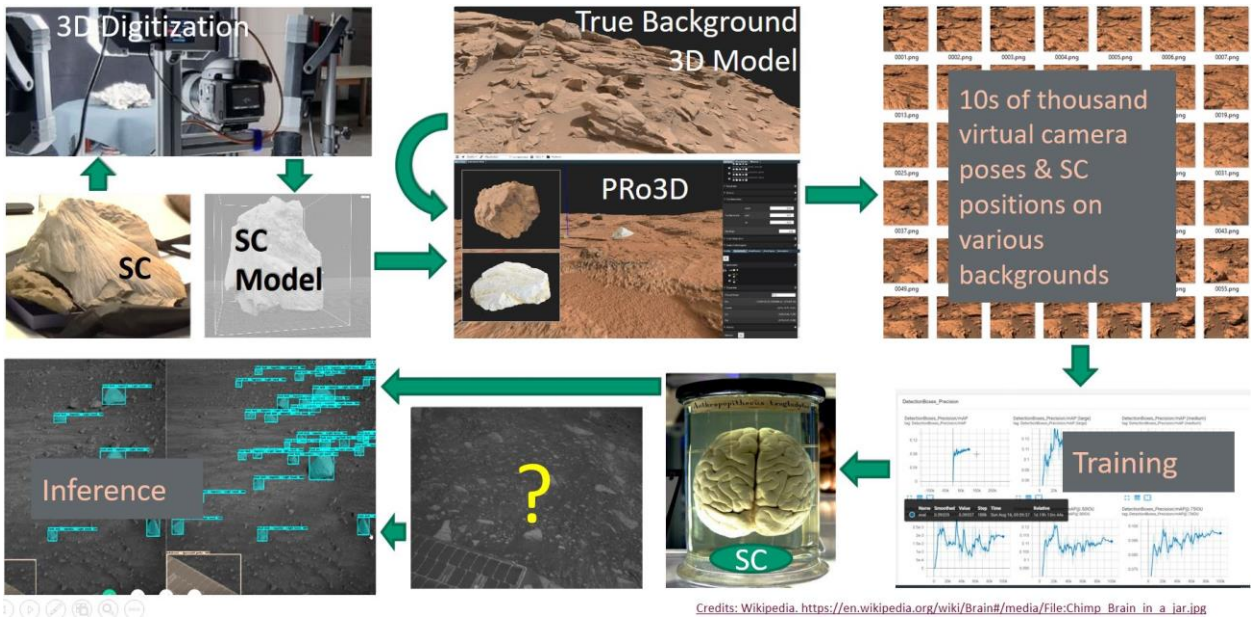


Figure 1: Mars-DL overview.

2.2 Objects of Interest

For Mars-DL it was decided to focus primarily on *shatter cones* (SC, Figure 2), which are the only macroscopic evidence of shock metamorphism and form during meteorite impact events [2], allowing to confirm impact structures.

Figure 2: Co-author Christian Koeberl presenting a shatter cone

Shatter cones take on a variety of sizes and shapes, depending mainly on the rock type, but all of them have distinctive fan-shaped "horsetail" structures that show striations, making them suitable objects to train a DL-system and assess its detection reliability.

As no shatter cones have yet been identified in real planetary imagery, no ground truth data are available to train a DL detection network. The creation of artificial visual samples based on real 3D models of shatter cones is, therefore, necessary. Our data augmentation algorithm places scanned 3D representations of textured and properly colored SC objects into the real context of a Martian digital terrain model. This process can be repeated thousands of times in order to create a realistic image sample database, which can be used to train a visual SC detector using a DL system.

In an image-capturing campaign at the Natural History Museum Vienna in May 2020, about 25 terrestrial objects (SCs and meteorites) were imaged (Figure 3), each with several hundred photos under controlled illumination conditions, keeping a scale and color target in the field-of-view.

Figure 3: SC capturing campaign

The resulting SC models have a sufficient geometric and texture resolution to preserve characteristic visual features in the simulation. To allow shading of SCs, the texture was captured with homogeneous lighting to obtain a pure albedo map without any prebaked lighting.

Many shatter cones from terrestrial sites have a contact surface, where they were broken or cut from the outcrop, and they often also bear a label. This section of the surface must not be visible in training images, and is marked as part of the meta-data of the 3D model to keep it invisible during automatic positioning.

Shatter cones usually occur within rock outcrops, not as isolated rocks, which is a challenge in producing training images.

The objects were seamlessly reconstructed using a COTS (Commercial-Off-The-Shelf) tool (*Reality Capture*) to gain high-resolution 3D models (wavefront .obj files, i.e., textured meshes). About 8 of them are being used for Mars-DL simulation sessions. See Figure 4 for an example.

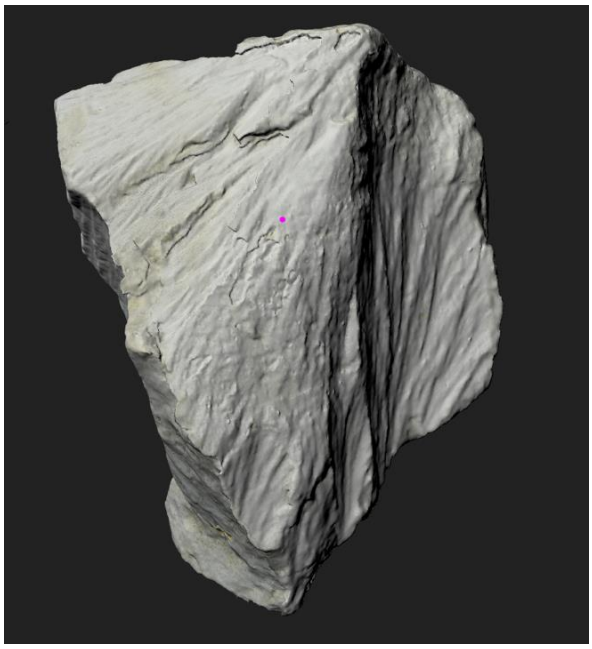


Figure 4: Rendered shatter cone 3D model.

2.3 Background: Virtual Martian 3D Representations

Realistic simulations are based on accurate 3D reconstructions of the Martian surface, which are obtained by the photogrammetric processing pipeline PRoVIP [1]. The resulting 3D terrain models can be virtually viewed from different angles and hence can be used to obtain large volumes of training data.

For Mars-DL simulations, about 10 stereo reconstructions from multi-stereo models from MSL Mastcam data are used (Figure 5), emphasizing close-range scenes to match the scale of the scanned SCs.

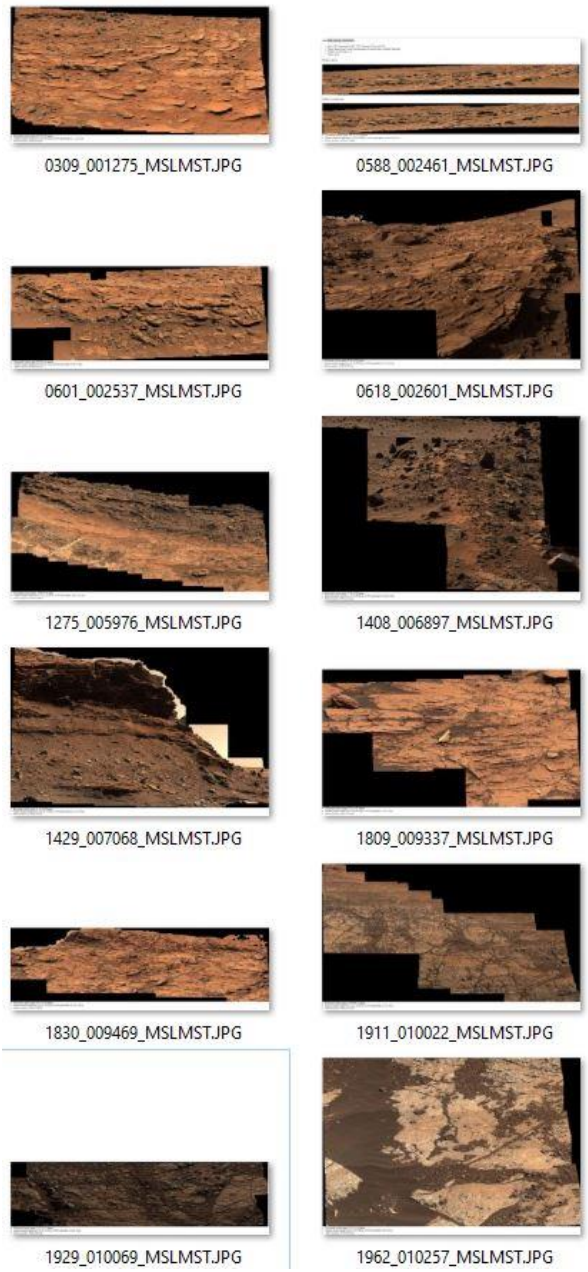


Figure 5: Overview of background scenes as 3D-reconstructed from MSL Mastcam stereo data

2.4 Simulation via PRo3D

PRo3D is an interactive viewer dedicated to planetary scientists for the virtual exploration and geologic interpretation of planetary surface reconstructions [3]. Fluent navigation through a detailed geospatial context

allows a visual experience close to field investigations. This is enabled by high-resolution geometries and textures (Figure 6). Therefore, PRo3D offers much of the required functionality to generate large volumes of training images with a high degree of realism.

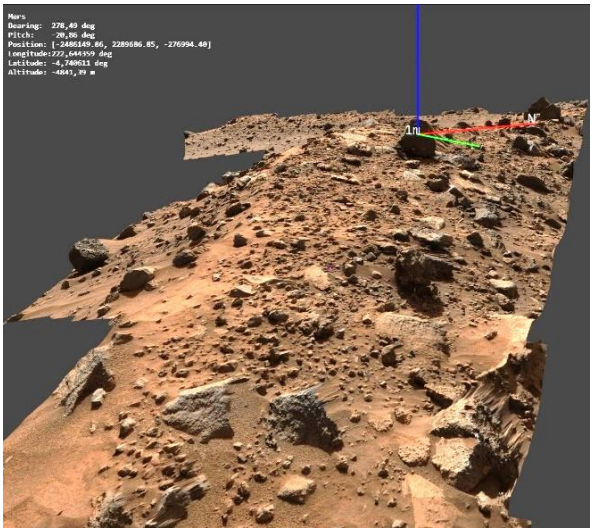


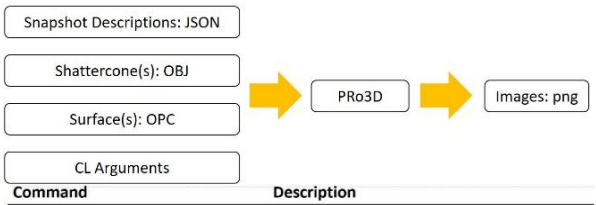
Figure 6: PRo3D visualization of Mastcam stereo reconstruction of MSL Sol 1408. Data credits: NASA/JPL/CalTech/ASU.

PRo3D allows for batch rendering to enable mass-production of training data. The simulation is controlled by commands in a JSON file, which defines a background scene, a viewpoint, SC models and various other parameters such as allowed ranges for random variations. These parameters can be set interactively in PRo3D and then exported into a JSON file (see Figure 7 for data flow, commands and parameters).

SCs are placed using a Halton sequence[1] within the given viewpoint region, which guarantees an even spatial distribution (Figure 8). Intersections between them are detected by checking their bounding boxes for overlapping regions and are avoided. One of two intersecting SCs is removed. An example is given by the object marked in red in Figure 8. On the other hand, intersections with the surrounding surface are required. For this, an intrusion depth interval is specified. Sizes and rotations of SCs are randomly varied within the specified ranges, making sure that contact surfaces are not visible (see section 2.2).

Realism must be enhanced in order to achieve suitable images for training, which should resemble true rover imagery as closely as possible. High resolution digital terrain models (DTM) of the Martian surface usually have an image texture derived from rover instrument imagery. Their prebaked lighting effects, such as shading, shadows, and specular highlights, are determined by the sun's direction at capturing time.

For a shatter cone to perfectly blend into the scene, it first has to be color adjusted to fit to the material properties of the surrounding landscape. Then it is shaded from the same illumination direction as the background scene, which is obtained from SPICE [10]. Shadows casted by the shatter cones from this direction are also calculated and contribute to the realism of the final rendering. Figure 9 shows the two stages to enhance realism, and Figure 10 the final result.



Command	Description
--help	show help
--obj [path];[path];[path];[...]	path to one or more OBJ files
--opc [path];[path];[path];[...]	path to one or more OPC folders
--asnap [path\snapshot.json]	path to a snapshot file containing surface transformations as well as camera views
--out [path]	specify a folder where the generated images will be saved
--renderDepth	Render additional image with depth
--excentre	show exploration centre
--refsystem	show reference system
--exitOnFinish	quit PRo3D once all screenshots have been saved
--verbose	use verbose mode
--noMagFilter	turn off linear texture magnification filtering
--snap	path to a snapshot file containing camera views

Parameter	Description
name	name of the shatter cone obj file
count	maximal number of this shatter cone in the snapshot
color	color adaption filter: color for the shatter cone (optional)
contrast	color adaption filter: default value:0, min:-255, max:255, stepsize:1 (optional)
brightness	color adaption filter: default value:0, min:-255, max:255, stepsize:1 (optional)
gamma	color adaption filter: default value:1, min:0.01, max:10.0, stepsize:0.01 (optional)
scale [min,max]	we randomize a scale factor in the min-max range (optional)
xRotation [min,max]	we randomize a rotation in degree in the min-max range (optional)
yRotation [min,max]	we randomize a rotation in degree in the min-max range (optional)
zRotation [min,max]	we randomize a rotation in degree in the min-max range (optional)

Parameter		Comment
fieldOfView	required	
resolution	required	resolution of the output image
snapshots	required	each entry in this list will result in one output image
filename	required	name of the output file
view	required	camera parameters
forward	required	forward vector of the camera
location	required	location of the camera
up	required	up vector of the camera
surfaceUpdates	optional	
opcname	required	name of the surface to be transformed
trafo	optional	transformation to be applied to the surface
visible	optional	visibility of the surface
mask	optional	use mask for all obj files in the scene

Figure 7: PRo3D batch data flow and commands.

[1] https://en.wikipedia.org/wiki/Halton_sequence

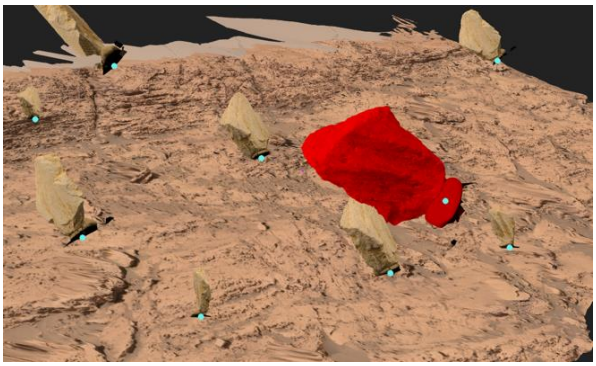


Figure 8: Intersecting ("colliding") shatter cones (the one marked with red will be removed). Cyan dots show positions of the Halton sequence. Background data credits: NASA/JPL/CalTech/ASU.

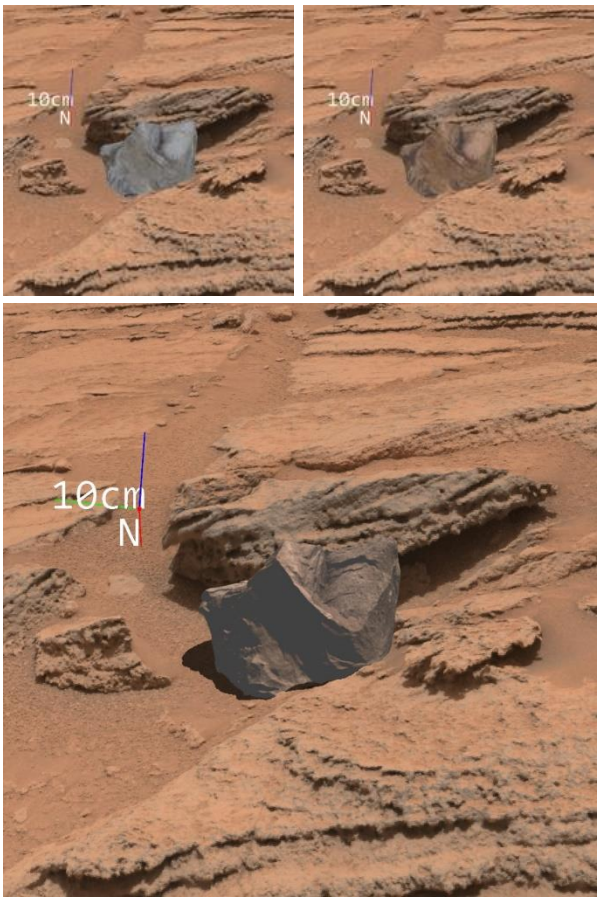


Figure 9: The original appearance of a shatter cone (top left), after color adjustment (top right), and shaded (bottom).

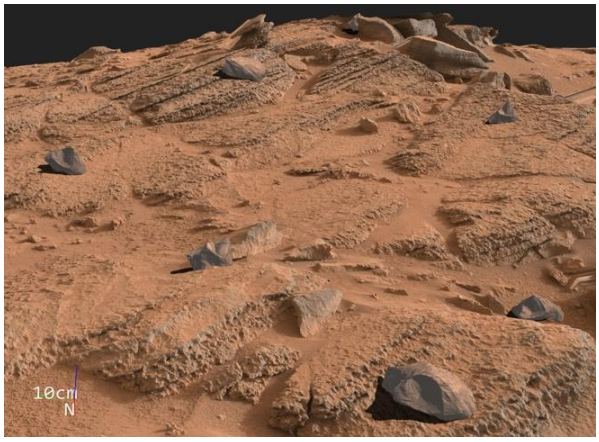


Figure 10: Shatter cones embedded in a Martian 3D reconstruction. Background data credits: NASA/JPL/CalTech/ASU/MSSS.

The training of DL systems also requires masks of the shatter cones and depth images of the entire scene. For both, special GPU shaders were implemented (*Figure 11*).

Figure 11: Cues for the training interface. Left: Image with realistic texture and shading. Middle: Depth map of entire scene. Right: Ground-truth mask.

2.5 Training and Inference

The visual detection of objects using Deep Learning Methods requires the training of neural networks with typically thousands of sample images with known ground truth. Using the data creation and subsequent augmentation process described in this work we are able to train a Deep Convolutional Network (DCNN), which should be able to detect objects which have actually not yet been spotted in Mars Rover images. The following subsections describe our approach in detail.

2.5.1 CNN Core Module

The heart of the STC CNN core module is a deep neural network used for object detection and instance segmentation. For the Mars DL prototype, we trained a Mask R-CNN [4] model, using the Inception Resnet V2 [5] backbone. The latter fully-convolutional network extracts high-level features from the input image. A region proposal network (RPN) uses those features to select region of interest candidates as coarse bounding

boxes. The final prediction stage generates object proposals (class, bounding box and mask) combining the features extracted by the backbone and the areas of interest generated by the RPN.

The next subsections first describe the training details, and following that the details of the inference system. An overview of the neural network structure is given in Figure 12.

2.5.2 Training

The object detection network is trained with Tensorflow's Object Detection API [6]. At the end of training the network weights are "frozen" and the subset of the network used to perform the detection on new images is extracted and later used in the inference engine.

We initialize our model with parameters pre-trained on the COCO detection dataset [7]. It is then fine-tuned using *Stochastic Gradient Descent with momentum* [8] for 100k steps and a starting learning rate of 2e-4, which decays to 2e-5 after 60k steps. The batch size is set to 4. We augment the training dataset by applying a random horizontal flip, the input images are then resized to 800 x 800 pixels.

In scenarios were depth information is available, as it is likely for Mars Rover Datasets, we add the depth map as a further image channel. An RGB image thus becomes an RGBD image (see also Figure 11).

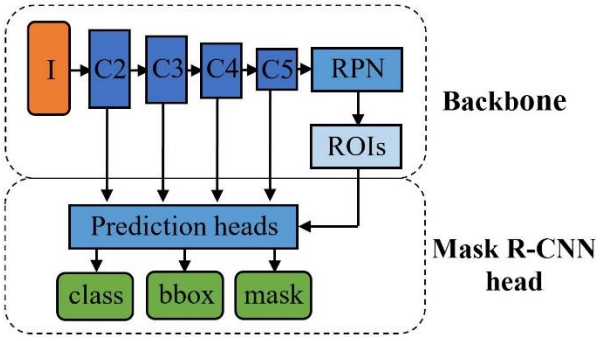


Figure 12: Neural network structure overview. Backbone: fully-convolutional network that extracts high-level features from the input image (C2-C5: convolutional blocks; RPN: region proposal network; ROIs: regions of interest). Mask R-CNN head: network section that generates object proposals (class, bounding box and mask) combining the features extracted by the backbone with the areas of interest generated by the RPN.

As a proof of concept, we trained on the LabelMars [9] dataset, using 4877 images for training and 100 images for evaluation. The total training time for this dataset was approximately 13 hours on two RTX 2080 Ti GPUs.

Based on this preliminary test run, we identified some criticalities in the LabelMars dataset, which limit the system's performance: i) low number of samples, ii) partial and inconsistently annotated data, iii) presence of very small objects, iv) object classes not well distinguishable based on visual cues only. These issues are addressed in the simulated data generation pipeline of Mars in order to produce sample image sets which are more suitable to DL approaches.

2.5.3 Inference

During inference, the CNN core module is fed with high resolution images, which are split into partially overlapping tiles. Tiling is necessary to maintain sufficient resolution, as a full high-resolution image would be downscaled too much by the preprocessing stages of the network.

The output of Mask R-CNN is a set of detection candidates, characterized by an object class, a confidence score, a bounding box and a binary mask indicating which pixels belong to the detected object. We discard detections with a confidence score lower than an acceptance threshold τ and forward the remaining detections to the subsequent modules of the STC system.

The following Figure 13 shows the detection results for the network trained on LabelMars. It includes the bounding boxes, confidences, classes, and qmasks.

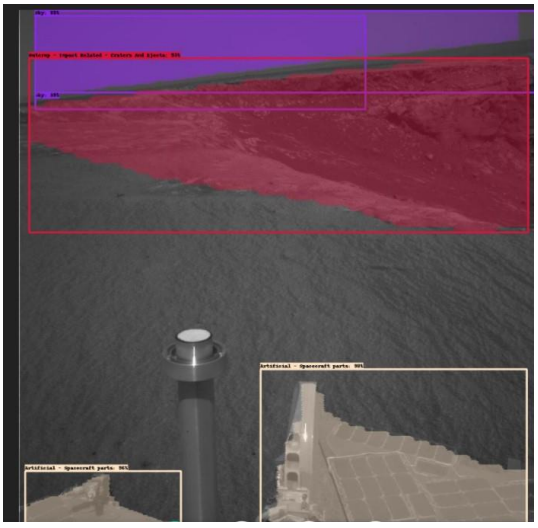


Figure 13: Example of visualized detection output on image from the LabelMars dataset. Image data credits: NASA/JPL/CalTech/ASU.

3. SUMMARY AND PROSPECTS

3.1 Status and Conclusions

We present an efficient way to generate large amounts of images for the training of DL systems to support autonomous scientific target selection in future rover missions. It is based on available high-resolution Martian surface reconstructions and newly created shatter cone models. The viewer PRo3D, developed for planetary science explorations, was extended for the efficient mass production of different DL training sets.

The generation of simulated training data is currently ongoing with several improvements of the data base and simulation elements. The interfaces to the training and inference components are already verified, and functional tests using publicly available manually annotated training sets were successful, which, on the other hand, showed several shortcomings of such manually annotated data in terms of consistency, precision, and amount of data. Follow-on activities for more general use cases are currently discussed with a first design phase still in the scope of the Mars-DL activity which will terminate in December 2020.

3.2 Mars-DL Use Cases

For past and present missions the project will help explore and exploit existing millions of planetary surface images that still hide undetected science opportunities. The Mars-DL activity assesses the feasibility of machine-learning based support during and after missions by automatic search on planetary surface imagery to raise science gain, meet serendipitous opportunities and speed up the tactical and strategic decision-making during mission planning. Mars-DL training and validation will explore the possibility to search scientifically interesting targets across different sensors, investigate the usage of different cues such as 2D (multispectral / monochrome) and 3D, as well as spatial relationships between image data and regions thereon.

During operations of forthcoming rover and lander missions the approach can help avoid the missing of opportunities that may occur due to tactical time constraints preventing an in-depth check of images.

The digitized shatter cones can also be used for public presentations, research, or as 3D printed objects for interactive purposes in exhibition. The system can further be used for training, public outreach, and testing of 3rd party image analysis methodologies.

3.3 Ongoing Work and Prospects

The background data sets are currently explored for optimum regions to place the simulated objects. Various parameters are to be adapted such as intrusion depth, angular variation of viewpoints and textural variations on the objects.

Future work in the simulation domain includes the design of a shader that combines shadows cast from SCs with the image texture of the surrounding surface and a fully automatic color adjustment.

A side activity is the testing of *style transfer* to enhance the realism in the rendered scenes, particularly for the inserted artificial objects – still keeping their distinct textural & shape features valid. Works on unsupervised generative-adversarial neural image-to-image translation techniques (generative adversarial network / GAN; [11]) have already been started. This will further improve the realistic placement of the simulated objects into the background environment, to avoid raising the interest of the DL system to detect “unusual” objects rather than distinguish the objects based on their textural and shape appearance relevant to their scientific origin.

One important item for future usability is the incorporation of other scientific phenomena such as layers, meteorites, or even serendipitous findings without an initial model. The embedding of albedo maps also for the background will further raise the degree of freedom of the simulation.

References

[1] Paar G., Muller J.P., Tao Y., Pajdla T., Giordano M., Tasdelen E., Karachevtseva I., Traxler C., Hesina G., Tyler L., Barnes R., Gupta S., Willner K. PRoViDE: Planetary Robotics Vision Data Processing and Fusion. In EPSC Abstracts, Vol. 10, EPSC2015-345, European Planetary Science Congress 2015.

[2] French B.M., and Koeberl C. (2010) The convincing identification of terrestrial meteorite impact structures: What works, what doesn't, and why. Earth-Science Reviews 98, 123–170.

[3] Barnes R., Sanjeev G., Traxler C., Hesina G., Ortner T., Paar G., Huber B., Juhart K., Fritz L., Nauschnegg B., Muller J.P., Tao Y. and Bauer A. Geological analysis of Martian rover-derived Digital Outcrop Models using the 3D visualisation tool, Planetary Robotics 3D Viewer - Pro3D. In Planetary Mapping: Methods, Tools for Scientific Analysis and Exploration, Volume 5, Issue 7, pp 285-307, July 2018.

[4] He K., Gkioxari G., Dollár P., Girshick R. (2017). Mask R-CNN. Proceedings of the IEEE International Conference on Computer Vision (ICCV).

[5] Szegedy C., Ioffe S., Vanhoucke V., Alemi A. (2016). Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. arXiv preprint arXiv:1602.07261

[6] Huang J., Rathod V., Sun C., Zhu M., Korattikara A., Fathi A., Fischer I., Wojna Z., Song Y., Guadarrama S., Murphy K. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. CVPR 2017

[7] Lin T.-Y., Maire M., Belongie S., Hays J., Perona P., Ramanan D., Dollar P., Zitnick C. L. (2014). Microsoft COCO: Common objects in context. ECCV.

[8] Ning Q. (1999) On the momentum term in gradient descent learning algorithms. Neural networks: the official journal of the International Neural Network Society, 12(1):145–151

[9] Schwenzer, S. P., Woods, M., Karachalios, S., Phan, N., Joudrier, L. (2019). LabelMars: Creating an extremely large Martian image dataset through machine learning. Proc. 50th Lunar and Planetary Science Conference.

[10] Acton Jr., C.-H., Charles, H. (1996) Ancillary data services of NASA's navigation and ancillary information facility. *Planetary and Space Science*, 44(1): 65-70.

[11] Zhu, J.-Y. et al. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. ICCV 2017, 2223-2232.

Acknowledgements

The Mars-DL study is receiving funding from the Austrian Space Applications Programme (ASAP14) funded by BMVIT. JR, VRVis, NHM, and SLR co-finance the activity. Initial training data from the LabelMars data set was received from ESA.