

Autonomous Rock Instance Segmentation for Extra-Terrestrial Robotic Missions

Maximilian Durner, Wout Boerdijk, Yunis Fanger, Ryo Sakagami,
David Lennart Risch, Rudolph Triebel, Armin Wedler

German Aerospace Center (DLR)
Institute of Robotics and Mechatronics
Münchner Str. 20, 82234 Weßling, Germany
Contact: Maximilian.Durner@dlr.de

Abstract—The collection and analysis of extra-terrestrial matter are two of the main motivations for space exploration missions. Due to the inherent risks for participating astronauts during space missions, autonomous robotic systems are often considered as a promising alternative. In recent years, many (inter)national space missions containing rovers to explore celestial bodies have been launched. Hereby, the communication delay as well as limited bandwidth creates a need for highly self-governed agents that require only infrequent interaction with scientists at a ground station. Such a setting is explored in the ARCHES mission, which seeks to investigate different means of collaboration between scientists and autonomous robots in extra-terrestrial environments. The analogue mission focuses a team of heterogeneous agents (two Lightweight Rover Units and ARDEA, a drone), which together perform various complex tasks under strict communication constraints. In this paper, we highlight three of these tasks that were successfully demonstrated during a one-month test mission on Mt. Etna in Sicily, Italy, which was chosen due to its similarity to the Moon in terms of geological structure. All three tasks have in common, that they leverage an instance segmentation approach deployed on the rovers to detect rocks within camera imagery. The first application is a mapping scheme that incorporates semantically detected rocks into its environment model to safely navigate to points of interest. Secondly, we present a method for the collection and extraction of in-situ samples with a rover, which uses rock detection to localize relevant candidates to grasp. For the third task, we show the usefulness of stone segmentation to autonomously conduct a spectrometer measurement experiment. We perform a throughout analysis of the presented methods and evaluate our experimental results. The demonstrations on Mt. Etna show that our approaches are well suited for navigation, geological analysis, and sample extraction tasks within autonomous robotic extra-terrestrial missions.

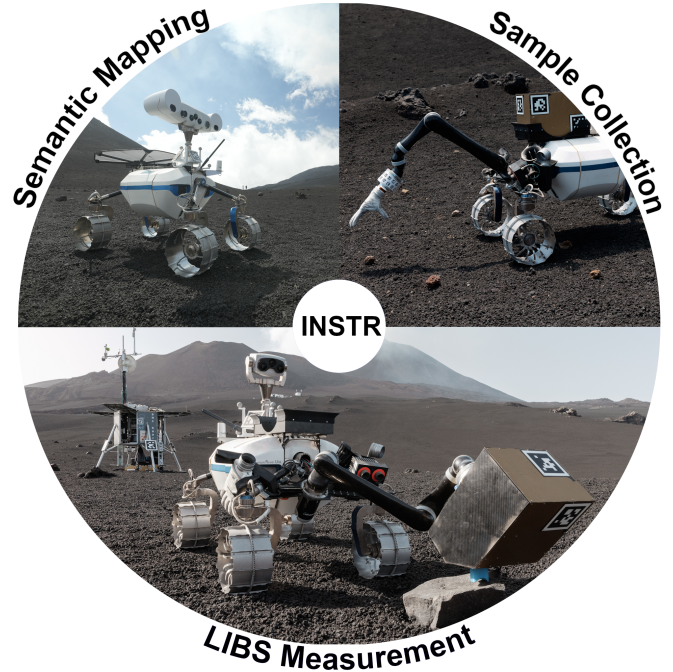


Figure 1: Our INSTR based rock segmentation approach was robustly run during the final demo mission of ARCHES. It was an important module in three mission critical applications: semantic mapping, in-situ LIBS analyses, and sample collection.

TABLE OF CONTENTS

1. INTRODUCTION.....	1
2. RELATED WORK	2
3. SYSTEM OVERVIEW	3
4. SEMANTIC MAPPING	4
5. SAMPLE COLLECTION AND ANALYSIS	5
6. EXPERIMENTS.....	8
7. SUMMARY	11
ACKNOWLEDGMENTS	11
REFERENCES	11
BIOGRAPHY	14

1. INTRODUCTION

Rocks are one of the major features present on the surfaces of celestial bodies that future space missions plan to explore (e.g., Mars or Moon). Besides being the principal impediments for rover traversal and having the potential to endanger rover safety, they can be important objects for both engineering and scientific activities. Firstly, the study of planetary geology can benefit greatly from the abundant information that rocks have to offer. The shape of rocks found on these celestial bodies provide valuable information about climate and erosion [1], [2]. Furthermore, as stated by Gor *et al.* [2], detected rocks can help with image compression and data priority as it is relayed to the Earth's control center [3], [4]. The locations and distributions of rocks also help for other autonomous capabilities such as site characterization[5] or target selection [6].

Considering the limited bandwidth, an on-board image analysis module is important for future space rover missions [7].

978-1-6654-9032-0/23/\$31.00 ©2023 IEEE

Further, the rising task complexity and operation distance of space missions, in combination with communication delays, requires modules to support rover autonomy in unknown, extra-terrestrial environments.

In this work, we present a rock segmentation module based on Instance Stereo Transformer (INSTR) [8], which is deployed on a *NVIDIA Jetson TX2* on each of our two Lightweight Rover Units (LRUs). The detection of rocks is an especially challenging task, due to the lack of a regular morphology regarding shape and size. Furthermore, rocks share important features, such as color and texture, with the surfaces they are on (e.g., gravel, soil), which makes them more difficult to detect. Thus, additional modalities such as depth sensing can help to obtain a more robust performance. The authors in [9] state, that 3D point clouds or range data are complementary cues about image intensity. To this end, we employ INSTR, which implicitly fuses RGB and disparity features without the necessity of high-quality depth data. In comparison to off-the-shelf solutions which have to be fine-tuned for a specific task (e.g. Mask-RCNN [10]), INSTR does not require any additional training data which makes it interesting in the planetary context given the scarce annotated training data. Additionally, the method showed empirically more robust results during our preliminary lab tests, and, importantly, also for the real mission demonstration - the latter being evaluated in Section 6.

We further present how the rock information is used in three crucial robotic applications for rover-based exploration of celestial bodies (see Figure 1). In the first application, we show that information about rock locations in the environment can be incorporated into the mapping process and used for navigation. In combination with the geometrical detection of non-traversable areas, this information enhances the safety of autonomous exploration and is a first step towards semantic navigation. Secondly, we highlight the important role of rock segmentation in the sample analysis. Thereby, one LRU, which is equipped with a robot arm, attaches a Laser Induced Breakdown Spectroscopy (LIBS) device that can be pointed towards rocks of interest to preform spectroscopic measurements¹. This is an established technique to analyze the elemental composition of rocks on Mars [12]. Thirdly, we present the task of sample collection, which is of particularly high interest given concurrent endeavours, like the Mars Sample Return Mission [1]. Considering known problems for space exploration robots, such as sporadic communication and missing data transmission, a skill to autonomously sample rock or soil is important. Based on detected rock instances, the LRU equipped with the robot arm is able to fully autonomously collect rocks or soil within the current scene.

In both the present and the future, space exploration activities heavily rely on mobile robotics. Aside from the actual development of the robotic systems, main challenges of these missions are the various ways of commanding those systems, using technologies ranging from teleoperation with human in the loop [13], through shared autonomy [14], to highly autonomous systems [15]. With lunar-analogue missions as here in the context of the Autonomous Robotic Networks to Help Modern Societies (ARCHES) project [16], [17], we want to give an outlook on future space exploration possibilities. Over the next decade, there will be a greater global commitment to lunar exploration, with humans returning to the lu-

nar surface, the ISS station reaching the end of its life, and its successor, the Lunar Gateway, being established in lunar orbit to pave the way for manned Mars exploration missions after 2030. Hence, the ARCHES demonstration missions serve to test and validate scientifically relevant mission scenarios; in particular, the validation and collection of valuable insights into our hard- and software developments for the use of robots in such challenging extraterrestrial environments is a central element. Based on the gathered experience and results, we aim to improve the modules' robustness, enabling their usage in future space missions. As an example, navigation components tested in the last analogue mission within the context of the predecessor project Robotic Exploration for extreme environments (ROBEX) [15] are now deployed in the Martian Moons eXploration (MMX), the first robotic space mission to return samples from Mars' largest moon Phobos [18].

Within the ARCHES project key challenges regarding the cooperative task-execution of heterogeneous robotic teams are the focus and current research. The agents autonomously collaborate to explore, collect and analyze samples, as well as deploy infrastructure and scientific instrumentations on the planetary surface. The applied methods and technologies will be relevant for robotic support and operation of permanent installations (e.g., the lunar village concept, or long-term scientific experiments). In this context, all three applications presented in this work are deployed during the final demonstration mission on Mt. Etna, Sicily. As our experiments show, we successfully demonstrated the robustness of the rock segmentation module even in this extreme environment. The stone segmentation for the semantic mapping was running during the complete mission, enhancing the safety of the rovers. Furthermore, we were able to complete several sample collection runs and multiple LIBS measurements.

The paper is organized as follows: in the next section, we give an overview on published rock detection approaches with a focus on planetary environments. Next, we introduce hardware and software modules most relevant for this work. In the following two sections, the applications containing INSTR are presented. All three applications were part of the final ARCHES demo mission on Mt. Etna, Sicily. We show and discuss the performance in Section 6.

2. RELATED WORK

The task of rock detection is addressed by a variety of publications. Thompson and Castaño [19] compare seven methods on four different rock datasets. During the Marsokhod rover field tests rocks were detected given their shadows [20]. As stated in [19], this seems to be one of the first rock detectors for autonomous science applications. The method uses a spherical lighting model to exploit the known sun angle. Although the approach only identifies a point on the rock instead of the complete rock outline, it can be directly used for sample analysis like the LIBS measurement [11]. Another approach tested in [19] is the Multiple Viola-Jones detector, a template-based rock detection algorithm that utilizes the template cascade proposed in [21]. The approach applies a supervised training strategy to create a cascade of filters to identify windows with rock candidates. Thus, instead of an pixel-wise instance mask, the method produces an approximate bound box for the rock.

A common idea is the application of a Support Vector Machines (SVMs) to obtain pixel-wise classification. These methods do not reveal the contours of each rock instance.

¹In this work we highlight only those parts of the method that are relevant from a perception standpoint. The complete approach is presented in [11]

Nevertheless, one can obtain an estimate about the fractional coverage of rocks within the scene. In [22], features generated by a Fuzzy-rough feature selection are forwarded to a SVM. Similarly in [23], for each pixel a feature vector is created based on Ant Colony Optimization (ACO) methods and forwarded to a SVM that assigns one out of seven terrain categories to each pixel.

The *Rockfinder* algorithm [6] and its successor the *Rockster* method [24] rely on edge detectors (e.g., Canny or Sobel algorithm). Both works are designed by the Jet Propulsion Laboratory (JPL) and search for closed contours in the edge fragments. The basic idea is that the majority of closed edge shapes in extra-terrestrial environments are on top of the surfaces representing rocks. Another approach proposed by the JPL is the Smoothed Quick Uniform Intensity Detector (SQUID). The method relies on the identification of contiguous areas of constant pixel intensities.

In [25] the concept of superpixels is applied to segment rocks in grayscale images from the Moon. Given the superpixel, features like size, texture, and intensity are extracted and processed in cut-graphs. A similar approach is proposed by Xiao *et al.* [26]. Based on the variance and intensity of those, a contrast value is computed which is compared to the other superpixel regions. Finally, a segmentation of rocks and background is produced.

Several rock detectors [2], [27], [5] use stereo range data as input modality. Given depth data, the first step is to fit a plane in the scene via RANSAC or Hough-Voting. Secondly, the distance between each pixel and the planar surface is computed to identify image regions above the plane. The usage of 3D information is also proposed in [9]. The authors propose a combination of the mean-shift algorithm and plane fitting to detect small and large stones.

More recently, a gradient-region constrained level set method is presented by Yang and Kang [28]. Grimm *et al.* [29] present an interactive perception approach relying on depth data. By pushing rock piles, the method generates separate clusters which represent rock instances. Besides these methods one can observe the trend to deep-learning based approaches [30]. In [31] a Mask R-CNN is fine-tuned on manually annotated data recorded in the laboratory. Furlán *et al.* [32] adapt the U-net architecture to separate stones from the background in a Mars-like environment. The network was trained with only 300 images and shows promising results. In general, the current bottleneck of deep-learning based methods is the limited existence of suitable real-world data. Even though there are a few datasets from planetary [33], [34], [35] and analog environments [36], [37], [38], they lack the required annotations for instance segmentation. Recently, to overcome this shortage, several simulators [39], [40], [41], [42] are proposed. In [41] and [43] the photo-realistic simulator called Outdoor Artificial Intelligent Systems Simulator (OAISYS) is used to generate synthetic data to fine-tune existing network architectures. In this work, we follow another approach and apply an Unknown Object Instance Segmentation (UOIS) approach to segment rocks. As a consequence, we do not require any additional fine-tuning on context-specific data, which makes it especially valuable in the planetary context given the limited annotated data.

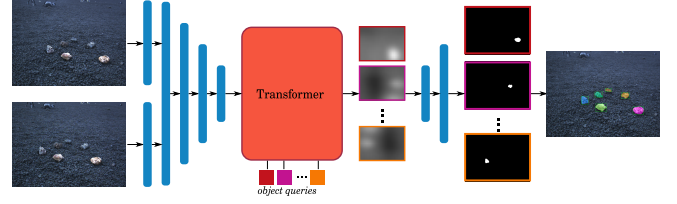


Figure 2: Given a stereo image pair, the transformer-based INSTR segments arbitrary objects, here rocks, on generic horizontal surfaces.

3. SYSTEM OVERVIEW

Rock Segmentation Approach

As readily mentioned, rocks have a highly irregular geometric structure. This makes conventional, model-based training of object detectors infeasible, since every newly encountered stone instance is unknown. Therefore, we tackle the problem from the point of *Unknown Object Instance Segmentation*, which aims at predicting instance masks of previously unencountered objects on dominant horizontal surfaces. While usually applied in in-door domains, we observe that the fundamental concept of objects on table-top surfaces is very similar to rock instances on (almost) flat terrain. We employ INSTR [8], a transformer-based architecture, which processes a pair of stereo images to *implicitly* reason about geometry in the scene. With this concept, INSTR outperforms other UOIS methods in contexts with potentially noisy or incomplete depth imagery, as quantitatively shown in [8]. After training on a vast amount of different object instances on in-door tables, the network can well estimate pixel-wise masks of arbitrary instances. The thereby learned concept of *objectness* enables it to generalize to novel scenes, even including particularly challenging ones such as lunar-like planetary environments (see Figure 2). Hence, our approach can be seen as learning-based successor of methods like [2], [27], [5], with the difference that here the extraction of a planar surface and the instance separation is simultaneously done in an end-to-end manner. Similar to these and other early methods [6], [24], [25], [26], we assume that all encountered object instances on the extraterrestrial surface are rocks². By following this principle we cast the problem to a pure instance segmentation. Please note, that if this assumption does not hold any longer, one can easily extend this approach with an additional classifier, as shown in [47], on-top to distinguish between object classes. INSTR is employed onto the rover’s onboard NVIDIA Jetson TX2, resulting on an average inference time of approximately 1s. The predictions are post-processed to filter out too small / big detections, which we further elaborate in the upcoming Section 5.

Space Rovers

The Lightweight Rover Unit (LRU) is a prototype rover that is designed for the application in extraterrestrial planetary exploration. During ARCHES demo mission, we deployed the INSTR network on two instances of the Lightweight Rover Unit, LRU1 and LRU2. Both rovers are built upon the same version of locomotion platform that consists of four wheels. Each of them is able to be individually controlled and allows the rover to maneuver through rough terrain [48], [49]. They are also equipped with similar computing hardware, including an Intel NUC i7 as the main on-board computa-

²Artificial objects, such as payload boxes or the lander are detected via our AprilTag [44] based method, that applies a multi-marker approach for a more robust estimation [45], [46]. With the knowledge about their locations, we can remove them from the scene leaving only rocks as possible objects.



(a) Large pan-tilt unit of LRU which includes a stereo camera pair for depth perception as well as spectral imaging cameras to gather additional information about the geological makeup of potential samples found in the environment



(b) View of the back side of LRU2 showing the Jaco2 arm with a robotic hand attachment performing a stone collection task. The rc-visard camera attached to the back provides visual feedback during manipulation.

Figure 3: Perception sensors on rovers used in ARCHES demo-mission.

tion device, an Intel Atom and BeagleBoneBlack board for motor control, and a Jetson board for deploying software that requires graphic card acceleration. The environment perception capabilities of the robots is purely vision based, provided by a pan-tilt camera unit. It is able to rotate with two degrees of freedom and includes a stereo camera pair for depth perception. Due to a filtered lens attached to these cameras the color is striped from RGB images that are used primarily for navigation tasks. However, there are also distinct differences between the two rovers that enable each of them to perform specific tasks which the other can not. These differences are mainly seen in the sensory stack as well as in the actuators on the robots.

The LRU1 is equipped with an extra large pan-tilt camera unit as shown in Figure 3a. In addition to the stereo camera pair, present in both rovers, this pan-tilt unit houses cameras for spectral imaging. It allows the rover to safely traverse the environment in search of scientifically interesting rock samples that are to be collected during the mission. The pan-tilt unit is able to perform a full 360 degree rotation which enables it to take round-view scans of the surrounding environment. Since LRU1 does not possess the means to collect samples by itself, it serves the purpose of a scouting robot in ARCHES and requires the assistance of a second rover for manipulation tasks.

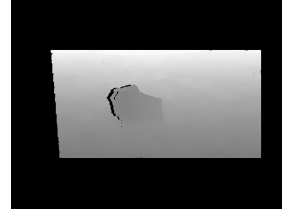
LRU2 is able to manipulate objects in the environment using its Jaco2 arm that is mounted to the back of the robotic platform. Depending on the task at hand, the rover is able to change the attached device at the end-effector autonomously. For collecting small rocks, a robotic hand is attached to the arm, while for sampling sand a shovel is used. Since the rover must turn with its rear side towards the objects it is supposed to manipulate, the pan-tilt unit is unable to capture images of the manipulation due to the rovers body being in the way. Thus, an additional *rc-visard 65* stereo camera³ is placed towards the back of the LRU2 as seen in Figure 3b. This camera module provides high resolution RGB images and corresponding depth images [50] of the scene.

LIBS Box

As stated in [19], rocks are ideal for conducting a spectrometer compositional analysis. The resulting information about the elemental composition is of high geological value. To this end, we conduct a LIBS analysis during the ARCHES mis-



Figure 4: The LIBS payload box with nozzle for constant distance between rock and spectrometer.



(a) Depth image recorded by the pan-tilt cameras used during navigation.



(b) Overlay of INSTR generated segmentation mask and the respective RGB image.

Figure 5: Depth image and segmented image recorded by pan-tilt unit

sion. As presented in [11], the LIBS instrument is integrated into one of the payload boxes, which can be manipulated and transported by LRU2. In general, the idea of the modular payload box design is to separate the rover and measurement instruments. Hence, the rover docks a specific box module only if required, which reduces the general weight of the rover and enables easy replacement or update of instruments. Figure 4 depicts the modular LIBS payload box. During measurement the module is picked by the arm mounted on LRU2 and placed on the object of interest. To ensure a constant distance between material and laser, the blue nozzle is mounted onto the hole of the laser.

4. SEMANTIC MAPPING

The first practical application of INSTR within the ARCHES project leverages the semantic terrain information extracted from segmented images to improve the robots behavior during navigation. To be more precise, it mitigates issues where a rover drives onto larger rocks in the environment while maneuvering towards a goal location. This is important since the wheels of the LRUs tend to get stuck in the creases of these stones, especially when performing in-place rotations. Applying high torque forces to the wheels while stuck may cause damage to the rovers hardware. Thus, the avoidance of such rocks is crucial for a safe navigation.

Traditional mapping approaches use depth sensing capabilities (cf. Figure 5a), to create a geometric representation of the environment. While these methods are able to detect larger obstacles geometrically, they cannot distinguish between a pile of gravel, over which the rover can drive with ease, and a narrow stone, which causes the aforementioned issues. To mitigate this, we build upon the purely geometric mapping approach [48] by adding a semantic layer to the local maps used for obstacle avoidance.

Figure 6 illustrates the full data processing pipeline used for the semantic mapping. The semantic layer is created by processing the image stream captured by the navigation stereo camera located in the pan-tilt unit of the LRU. Since the segmentation using INSTR is significantly slower than

³<https://roboception.com/en/rc-visard-en/>

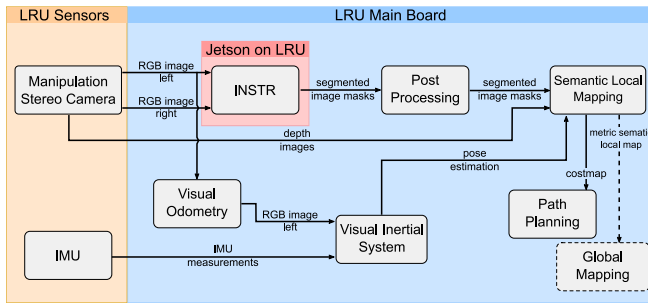


Figure 6: Diagram showing the data-flow used in the semantic mapping system. Dashed lines represent future work.

the camera frequency we sub-sample the image pairs to about 0.6 Hz. The mapping system is designed to be able to handle the asynchronous integration of semantic data into the geometric map, even with the additional time delay, that the segmentation algorithm introduces. After the instance segmentation masks are computed, a response is sent back to the main On-Board Computer (OBC) for additional computations. This post-processing consists of filtering out objects that are too small to be considered obstacles worth avoiding. The size estimation is based on the obstacles 3D size which is computed by combining the segmented data with the respective depth image that is generated by Semi-Global Matching (SGM) [50], using the stereo data of navigation cameras.

For our application scenario of navigating in moon-like environments, we only distinguish between rocks and gravel as terrain types. Hence, we are able to use the output of INSTR directly by assuming that detected objects are always going to be rocks. However, this approach could be adapted by adding a classifier into the processing pipeline to allow for different treatment of distinct semantic classes. Since the instance information is not relevant for this navigation use-case, we combine all of the masks into a single image that includes all obstacles. An example image of the camera image overlaid with the mask is shown in Figure 5b.

The semantically segmented images are then combined with the depth images to create point-clouds that include the terrain information as labels. Note that this is a simple arithmetic operation, where depth artifacts do not cause immediate problems. For Deep Neural Networks, partially incomplete inputs are usually more difficult to be processed, which is stated as a main motivation for the stereo imagery for INSTR [8]. Furthermore, points that are farther away from the camera than a threshold distance are filtered out, since the accuracy of both the semantic and geometric information decreases with increasing distance. Simultaneously, we employ a Visual Inertial System (VIS) that fuses Inertial Measurement Unit (IMU) and Visual Odometry (VO) data to give us a relative pose estimate from the starting location. This localization is used to register the point-clouds to each other over time. Finally, the points are aggregated into cells of a grid map. This map has a separate cost layer which describes the costs of traversing each grid cell. Thereby, the cost is high if the geometric makeup of the terrain is considered too steep, as well as when the semantic label shows the terrain to be a rock. Thus, the rovers are able to plan paths around the objects that are determined to be rocks. A projection of the costmap into the camera frame during buildup of the map is shown in Figure 7a. The final map is shown in a top-down view in Figure 7b, where areas with

geometrically or semantically detected obstacles are marked. Including the semantic data gathered in these local maps into a global mapping scheme is future work that we plan to integrate for the next mission.

5. SAMPLE COLLECTION AND ANALYSIS

In addition to the semantic mapping, INSTR is further utilized for rock sample collection and in-situ LIBS analyses. In comparison to the semantic mapping, the additional technical challenge for these tasks is that the rover needs to plan collision-free motions of the robotic arm and to manipulate the rock in a suitable manner. Therefore, the perception components are required to estimate geometry of rocks and to reason feasible approaches for the gripper or contact points for the LIBS instrument. For these tasks, different processes centered around INSTR need to coordinate with each other as shown in Figure 8. The *gripper approach point estimation* process is utilized for collecting the rock samples, while the *Octomap generation* and the *LIBS contact point sampling* is specifically employed for the spectrometer measurements.

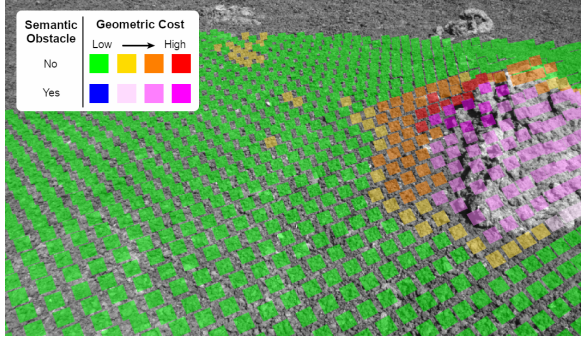
The *world model* is as a central knowledge representation of the physical, real world, and serves also for managing the model of the rocks. It helps to avoid duplicated data representation and processing. Furthermore, it allows for simple synchronization of the rock model with the motion planner [51] or any other software components.

The *sample selection interface* is employed to include the scientific operators into the decision-making activities of which rock to collect or apply the LIBS measurements. The graphical user interface (GUI), shown in Figure 9, provides a drop-down menu for selecting a camera system and input forms to specify parameters. Once the segmentation process is triggered, the GUI overlays the received segmentation mask over the camera stream. The operators can intuitively click on the image to select a rock for scientific activities. The “Send sample info” button at the bottom generates an object into the world model. A unique ID is given to the object, which is provided for the modelling processes to recognize the rock of interest.

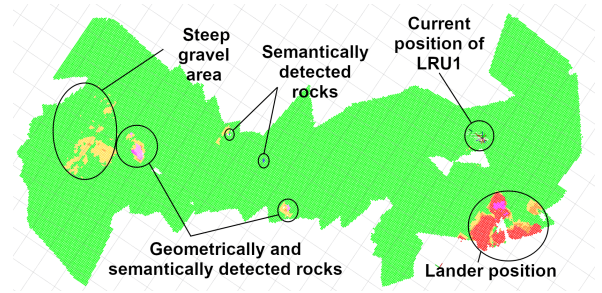
Sample Collection

The highly distinct nature of a rock’s shape makes model-based pose prediction for designated grasp estimation impossible. To this end, we present a top-down grasping approach that can adapt to the varied geometric nature of rocks, and is solely based on the readily available stereo imagery. Together with predictions from INSTR, unsuitable candidates are excluded (too small/big) based on their estimated size. Then, an approach point a_i is calculated for every valid stone instance i in 3D space, from which the arm moves vertically downward until contact is reached and the grasp is executed.

Approach Pose—The estimation of both size and an approach point requires knowledge about the object’s 3D shape. Ideally, one would calculate a 3D bounding box from which size (e.g., the diagonal of the bounding box) and an approach point (the horizontal center of the box plus an upward offset) can be deduced. Yet, shape irregularities such as bulges are invisible to the camera when located on the back side of the stone, and thus they render a precise 3D bounding box estimation impossible. To this end, we propose to derive the approach point merely from the 2D bounding box of the segmented rock in the image plane, and all depth values within the predicted INSTR mask. Formally, let a_i denote



(a) Overlay of semantically annotated local map over camera image during buildup of map.



(b) Top-down view of the local map created in one run using the semantic mapping approach during the ARCHES mission. Several areas are highlighted, including the starting point at the lander, purely semantically detected smaller rocks, a sloping area with gravel, as well as larger stones detected semantically and geometrically.

Figure 7: Semantic grid map during buildup and top-down view. Hereby, green points are traversable terrain. Shades of yellow \rightarrow orange \rightarrow red denote progressively more difficult to traverse terrain purely determined on basis of geometric data. Shades of purple are used to illustrate objects that are additionally detected as semantic obstacles. Blue regions are detected only semantically and would not be considered obstacles geometrically.

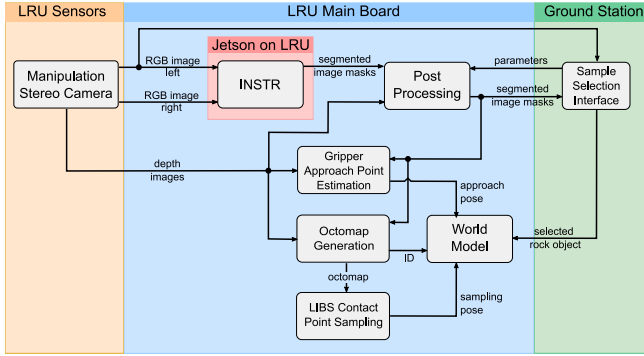


Figure 8: Diagram showing the data-flow used in the rock sample collection and the LIBS measurements

a 3D approach point from which a robotic arm can move vertically downward to grasp a stone - that is, the x and y coordinates of \mathbf{a}_i are the horizontal center of the stone's 3D bounding box. Let further $c_{i,x}$, $c_{i,y}$ denote the center of the 2D bounding box in the image plane of the detected stone. With the help of the traditional pinhole model, a relationship between the 2D and 3D information can be derived:

$$\mathbf{a}_i = \mathbf{K}^{-1} \begin{pmatrix} c_{i,x} \\ c_{i,y} \\ 1 \end{pmatrix} d_i + s \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \quad (1)$$

where \mathbf{K} is the intrinsic camera matrix, d_i is the distance from camera to \mathbf{a}_i , and s is a scalar factor to move the approach point atop the object. In other words, \mathbf{a}_i is the intersection of the ray through the 2D bounding box center, and the 3D approach direction. Since \mathbf{a}_i is thus located somewhere inside the stone, d_i cannot be calculated directly. Hence, we estimate the position $\hat{\mathbf{a}}_i$ by selecting a pre-defined percentile at the depth values \mathbf{d}_i belonging to the detected stone - that is, all depth values on the INSTR mask:

$$d_i \approx \hat{d}_i = f(\mathbf{d}_i, p), \quad (2)$$

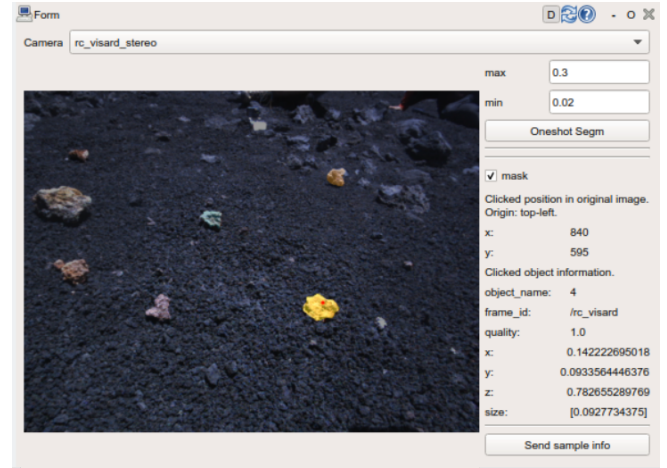


Figure 9: Screenshot of the sample selection interface. The red dot on the rock segmented with a yellow mask indicates the clicked point by the operators.

with $f(\mathbf{d}_i, p)$ returning the p -th percentile of the masked depth values \mathbf{d}_i . For a visualization, see Figure 10. To estimate the stone's size we take the diagonal of the projected 2D bounding box.

Percentile Derivation—The selection of p is crucial for a good approximation of the stone's location and primarily depends on the camera angle, the stone's shape, and its position relative to the robot arm. To derive an intuition for a suitable p and quantitatively evaluate the resulting accuracy, a synthetic scene is re-created in BlenderProc [40], mimicking the camera intrinsics and extrinsics of the LRU2's rc_visard 65 (see Figure 11). To realistically emulate the presence of stones, we explore different settings of object types, size and placement and refer the reader to Table 1 for further details.

For each possible combination the object is placed on the floor in front of the rover and a depth map is rendered. Finally, approach points are calculated for different percentile

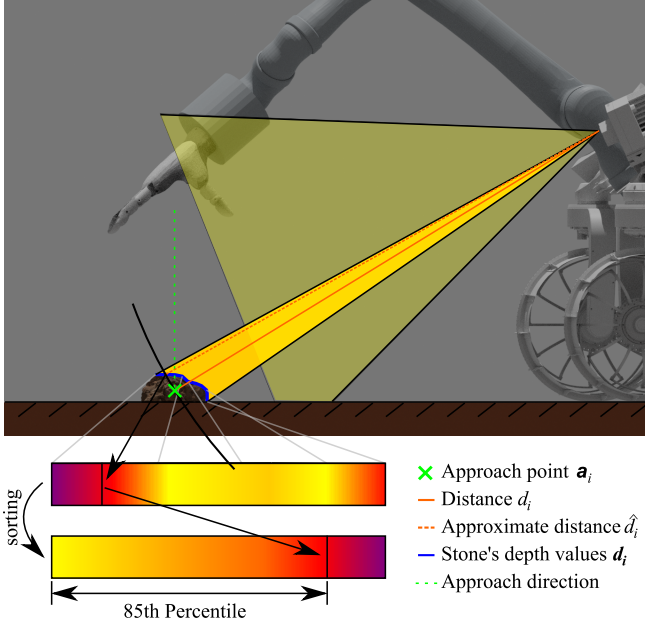


Figure 10: 2D sketch of the approach pose derivation via percentiles from a stone’s depth values (best viewed magnified and in color). Bright colors in the depth bars denote closer distance values to the camera.

Table 1: Object placement settings. Three different primitives are used (cube, icosahedron, cylinder). The object z -location refers to the offset applied to the object’s center relative to the ground plane. The distance to camera is chosen such that the robot can well reach the object for grasping.

Setting	Values
Primitive / rotation	\square , \square , \square , \square , \square , \square
Object size [cm]	6, 8, 10
Object z -location [cm]	-2, 0, 2
Distance to camera [cm]	70, 75, 80, 85, 90, 95, 100

values and the absolute error between the estimated point and the actual bounding box center, both in the xy -plane, is calculated. As can be seen in Table 2, a percentile of $p = 85$ results in the overall best approximation across three different shape primitives.

To test our primitive-based hypothesis we perform the same analysis on site with a set of 36 real rock samples collected from the volcanic environment [52]. Based on the respective results listed in Table 3, a depth-selection percentile of 85 indeed results in lowest error for rock sample grasping.

Up to now, the primarily focus of this section was rock detection, pose estimation and extraction. Yet, the pipeline can equally well be used for sand sample collection, which is triggered upon a user click on any part of the image *but* the stone predictions. Now, instead of the gripper, a shovel is mounted, and the approach point is calculated in similar fashion as in (1), with the only difference being that $c_{i,x}$, $c_{i,y}$ are the clicked point instead of the bounding box’s center.

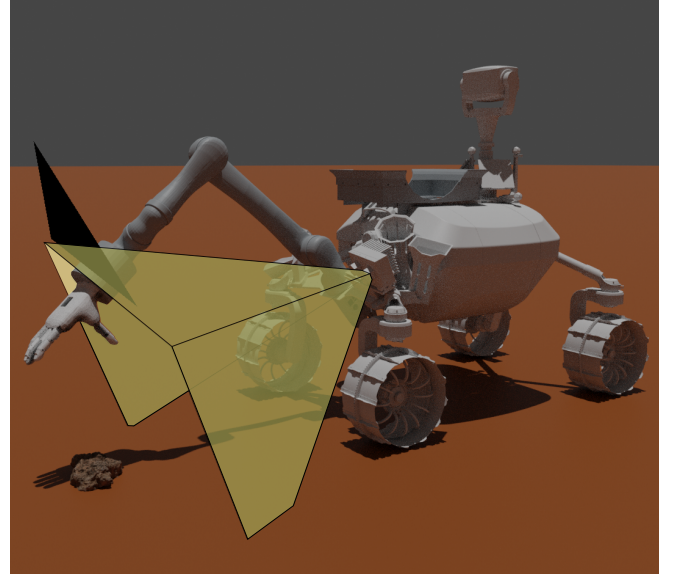


Figure 11: Modeling different primitives in Blender for the LRU2’s rc_visard 65 sensor, here with an exemplary stone. The location of the stone is selected such that it is reachable with the robotic arm.

Table 2: L1 distance [cm] between predicted horizontal approach point and original value for given percentiles across various primitives. The 85th percentile results in the lowest error overall.





































Primitive	Percentiles				
	75	80	85	90	95
\square	6.04	4.44	4.40	7.95	15.59
\square	9.11	6.64	7.27	12.16	18.47
\square	26.39	23.72	20.72	16.68	8.41
\square	23.12	15.08	8.09	9.4	23.85
\square	42.21	37.69	31.88	24.31	14.11
\square	21.10	12.56	11.09	21.12	35.68
Mean	21.36	16.69	13.91	15.27	19.36

LIBS Measurement

Similar to the sample collection, the rock to be analyzed with the LIBS instrument is selected via the Graphical User Interface (GUI). Figure 12 depicts the perception pipeline to determine the sample spot. To determine a suitable location on the rock of interest, we first compute a 3D representation of the rock instance. Given one or multiple depth images, we generate a so-called Octomap [53]. This representation describes the world by occupied voxels, with a certain resolution (here 1cm). The occupancy probability of each of these voxels is computed from the depth measurements by taking prior probability and the sensor model into account. Since we only regard contact locations on the stone, the area covered by the Octomap is defined by the 2D segmentation mask of the stone predicted by INSTR. Hence, we only consider pixels of the depth map which are inside the instance mask. A positive side-effect is the reduced computation speed due to the low amount of relevant voxels.

To ensure reliability during the measurement, we define two criteria for the sampling spot. First, the nozzle of the LIBS

Table 3: L1 distance [cm] between predicted horizontal approach point and original value for given percentiles across various real-world rock samples (best viewed magnified). The mean values for the given percentiles are 10.8, 9.8, 9.3, 9.6 and 11.6 cm; the 85th percentile results in the lowest error overall.

Stone	Percentiles					Stone	Percentiles					Stone	Percentiles				
	75	80	85	90	95		75	80	85	90	95		75	80	85	90	95
	11.8	11.2	10.9	11.5	13.3		10.9	9.2	7.9	7.6	8.9		10.9	9.9	9.4	9.5	11.8
	10.1	9.9	10.3	11.4	14.4		9.8	8.9	8.4	8.2	9.6		13.0	13.3	13.4	14.3	16.0
	10.2	9.0	8.1	8.0	9.4		11.3	9.9	8.8	8.5	9.7		9.7	8.9	8.3	8.0	9.8
	9.1	8.5	8.4	9.3	11.9		11.4	10.5	9.9	10.0	11.6		9.4	8.2	8.1	8.8	11.1
	12.6	12.0	11.7	12.1	14.2		10.1	9.2	8.7	8.8	10.8		9.9	9.1	9.0	9.9	12.7
	10.1	9.0	8.3	8.6	10.9		10.7	11.0	11.6	12.9	15.7		13.1	11.9	11.2	11.4	12.9
	12.0	11.2	10.7	10.9	12.0		9.4	7.4	5.7	5.1	7.1		10.4	9.5	9.3	9.6	11.2
	9.7	7.9	6.2	5.6	7.1		10.1	8.8	8.4	9.2	11.8		11.9	10.0	8.1	6.6	7.0
	11.4	10.5	10.3	11.3	13.9		9.3	7.3	6.2	6.1	8.1		12.4	11.7	11.4	12.0	13.8
	8.3	7.2	7.2	8.5	11.6		13.5	12.0	10.3	9.3	9.9		10.9	10.6	10.9	11.6	13.5
	9.3	7.8	6.9	7.4	9.9		10.8	10.2	9.8	10.7	13.7		11.5	10.7	10.6	11.4	13.8
	13.0	12.7	12.7	13.3	14.9		8.7	7.4	6.6	6.3	8.0		9.9	9.3	9.3	10.0	13.0

box should approach the sampling point from above. Due to the high weight of the measurement box, the robot arm operates on its upper limit. In combination with the extreme sun radiation on Mt. Etna, the arm runs the risk of overheating. To counteract this, we aim to rest the major weight of the box on the stone via the nozzle (see Figure 4) during measurement. Additionally, this also reduces the power consumption which is an important aspect for mobile rovers.

We express this criteria by a score value s_u for the candidate point \mathbf{v} . Therefore, we consider the normal vector $\mathbf{n} \in \mathbb{R}^3$. In general, this vector describes the normalized sum of the direction to every free neighboring voxel. Please note that unobserved voxels (e.g., inside the stone) do not count as free. Ideally, the normal vector should point upwards, which leads to the following score equation:

$$s_u = 1 - \min\left(\frac{2}{\pi} \arccos\left(\mathbf{n} \cdot \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}\right), 1\right). \quad (3)$$

The value is in the range $[0, 1]$ with 1 being the best. It is to be noted, with this criteria we constrain ourselves to rocks which are on a planar surface with respect to the robot. However, given a stronger robot arm, we could soften or even ignore this criteria.

The second criteria is a flat surface around the sampling spot. This ensures a stable rest position of the instrument's nozzle on the stone. Similar to the first constraint we also compute a score for this one. We compute the maximum distance between a plane defined by the candidate point \mathbf{v} and every occupied voxel \mathbf{p} within the radius r_s . Let $\mathcal{P} = \{\mathbf{p} \mid \text{dist}(\mathbf{p}, \mathbf{v}) < r_s\}$ define the set of voxels within the radius to the candidate point \mathbf{v} . Then we can compute the maximum surface distance for \mathbf{v} by

$$d_{\max} = \max_{\mathbf{p} \in \mathcal{P}} |\mathbf{n} \cdot (\mathbf{p} - \mathbf{v})|. \quad (4)$$

With this we define the flatness score for the candidate voxel \mathbf{v} as

$$s_f = \min\left(1 - \frac{d_{\max}}{2r_s}, 1\right). \quad (5)$$

The final score for each voxel is the minimum of both scores:

$$s_{\mathbf{v}} = \min(s_u, s_f). \quad (6)$$

The voxel with the highest score is the preferred location to take a measurement. Since various other factors are not considered by the presented score function (e.g., reachability of sample spot), we consider the 10 highest locations.

6. EXPERIMENTS

The peak of the ARCHES project was a 4 week mission on Mt. Etna, Sicily. The location was chosen because of its lunar-like environment in terms of geological structure. During the mission, we executed all three applications several times, demonstrating the robustness of all involved hard- and software components. In the following, the findings of the results within the final demo mission are discussed.

Mission Set-Up

Before we elaborate on the experiments, a short overview of the conducted ARCHES demo mission is given. Generally, the basic idea is, that a scientifically related geological mission should be executed as precursor, without the lunar gateway and robots are operated from earth (Mission I - Geological Mission I (GEO-I)), followed by a second mission (Mission II - GEO-II) which collects the samples from the prior missions with already explored sites and positions [54]. Mission III deals with the task of establishing a permanent scientific installation, here a Low Frequency Array (LoFar) telescope on the far side of the moon [55].

Although the navigation ran during all missions related to the LRU systems, we here focus on Mission I - GEO-I, since there all three applications are used. There both rovers, LRU1 and LRU2, operated semi-autonomously to reach, inspect, and collect targets of scientific relevance. The geological interesting locations were situated a maximum of 50 meters away from the Lander.

LRU1, equipped with a range of scientific visual sensors in its pan-tilt unit (cf. Figure 3a), autonomously navigated towards

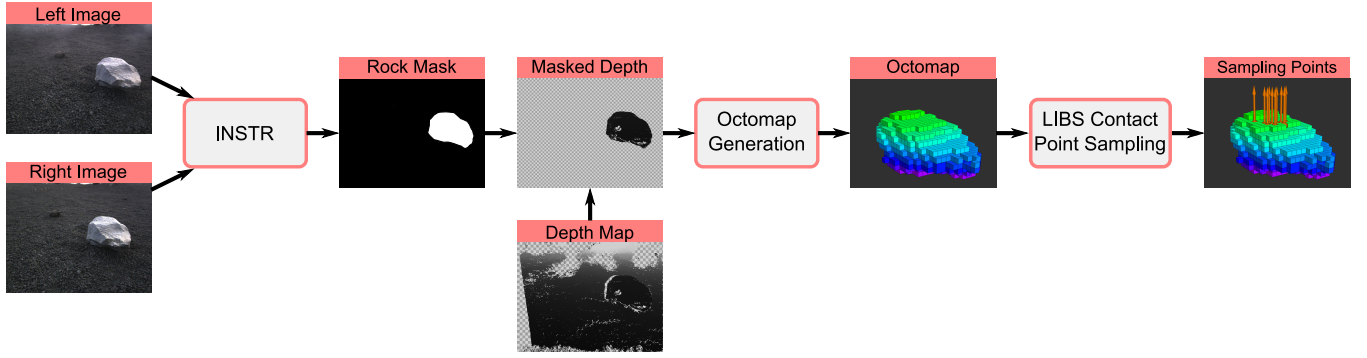


Figure 12: Perception pipeline to determine LIBS sampling spots on the rock of interest. The rock mask is fused with the depth map. The masked depth map is forwarded to the Octomap generator. Finally, the sample points are determined based on two criteria.

3 different locations to generate panoramic scans. Given the presence of isolated stones or regions with bedrocks, each of the three goals was positioned in hazardous and difficult-to-reach areas. However, during the whole mission we achieved safe navigation.

For the LRU2 rover, GEO-I started with picking the sample box from the Lander, followed by navigating to the relevant sample sites. At each sample site, the operator decided via a GUI whether a specific rock or a soil sample was collected. Based on the decision, the rover attached a robotic hand or a shovel and executed the sample collection.

In the second part of GEO-I a geochemical in-situ analysis was conducted. Therefore, LRU2 replaced the sample box with the LIBS payload module from the Lander and drove to one of the three the relevant location. After reaching the target position, the operator selected a stone of interest. Next the rover docked the LIBS device, placed it on a suitable surface on the selected stone and performed the LIBS measurement.

A complete demonstration of the mission had a total duration of 3 hours. In the final demo, both rovers autonomously navigated and manipulated without interruptions, demonstrating outstanding robustness for all software and hardware components involved in the process.

Experimental Results

In the following, the performance of INSTR during the demonstration week is evaluated. In total, we report results on 62 images for autonomous sample extraction and 17 images for LIBS measurement. For the navigation part, we evaluate a subset of 49 samples. To provide accurate ground truth masks, we manually annotate every stone instance in the RGB frame. Rock masks that are outside the desired grasping size are discarded (see Section 5). Note that for the stone segmentation during LIBS measurement, we only consider the biggest rock and mask out other rocks.

After matching every detected instance with its corresponding ground truth, we compute the binary True Positives (TPs), False Positives (FPs), and False Negatives (FNs) for all pixels labeled as object in an image. Remaining ground truth or prediction instances are matched with a zero mask. The *precision* and *recall* for a matched pair (g_i, p_i) can then be

computed via:

$$\begin{aligned} \text{precision}(g_i, p_i) &= \frac{\text{TP}}{\text{TP} + \text{FP}}, \\ \text{recall}(g_i, p_i) &= \frac{\text{TP}}{\text{TP} + \text{FN}}. \end{aligned} \quad (7)$$

We follow [8] and calculate the Intersection over Union (IoU) in a *size-sensitive* way, i.e. we summarize TP, FP and FN scores for all objects in a scene:

$$\text{IoU}(g, p) = \frac{\sum_i g_i \cap p_i}{\sum_i g_i \cup p_i} = \frac{\sum_i \text{TP}_i}{\sum_i \text{TP}_i + \text{FP}_i + \text{FN}_i} \quad (8)$$

We also compute the *F1* score by:

$$\text{F1}(g, p) = \frac{\sum_i 2\text{TP}_i}{\sum_i 2\text{TP}_i + \text{FP}_i + \text{FN}_i}. \quad (9)$$

It can be shown, that the IoU and the *F1* score are positively correlated, meaning that if method I is better than method II in one metric, it is also better in the other one. The main difference is the larger penalization of a single instance in the IoU, leading to possibly deviating results taking the average score over a set of inferences. The final scores are computed by averaging across all scenes.

Since the above definitions of IoU and F1 can be interpreted in a *semantic* way, we additionally list the Panoptic Quality (PQ) [56] that instead computes the mean across detected instances that are matched with a valid, non-zero ground truth:

$$\text{PQ} = \frac{\sum_{(g_i, p_i) \in \text{TP}} \text{IoU}(g_i, p_i)}{|\text{TP}| + \frac{1}{2}|\text{FP}| + \frac{1}{2}|\text{FN}|}. \quad (10)$$

Again, the final scores are computed by averaging across all scenes.

The mean values for each task are depicted in Table 4, where we also list the performance of a Mask-RCNN [10] trained on synthetic OAISYS data generated with assets provided by the authors⁴. Figure 13 depicts exemplary qualitative results.

For a more intuitive understanding, we depict the absolute number of correctly identified stones for a particular IoU threshold in Figure 14.

⁴<https://github.com/DLR-RM/oaisys>

Table 4: Quantitative evaluation of INSTR during the demonstration week for the tasks of navigation, autonomous sample collection and LIBS measurement.

Task	No. Images	Method	IoU [%]	Precision [%]	Recall [%]	F1 [%]	PQ [%]
Navigation	49	Mask-RCNN	22.23	48.08	28.48	31.89	20.90
		INSTR	51.87	95.31	53.42	63.82	46.60
Autonomous Sample Collection	62	Mask-RCNN	70.61	95.25	73.19	82.20	72.07
		INSTR	70.24	94.87	72.84	81.30	71.53
LIBS Measurement	17	Mask-RCNN	85.59	92.63	92.28	92.11	84.48
		INSTR	94.16	98.06	95.95	96.95	88.93
Mean (per task)		Mask-RCNN	59.48	78.65	64.65	68.73	59.15
		INSTR	72.09	96.08	74.07	80.69	69.02

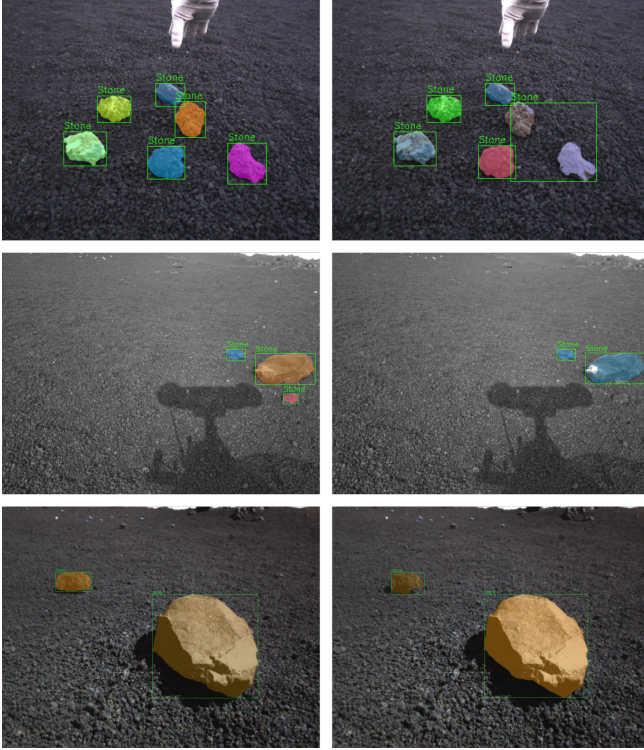


Figure 13: Qualitative results of INSTR employed during sample extraction (top), navigation (middle) and LIBS measurement (bottom). Left images depict ground truth annotations, while right images show INSTR predictions. Colors are assigned randomly. Bounding boxes are added for enhanced visual experience.

In total, the network was triggered more than 1,000 times during the main demonstration week; a major part stemming from navigation requests. It demonstrates good performance across all tasks, and particularly excels during the LIBS measurement, where stone sizes are considerably larger. An interesting observation from Table 4 is the overall high precision, indicating few background pixels labeled as stones. For navigation imagery, the mean IoU and Panoptic Quality is substantially worse; a fact which potentially can be contributed to the filters on the LRU1’s camera. While of great benefit for geological examination, INSTR has not been exposed to such altered data during training, thus potentially being less accurate on the respective images. Undetected instances, as depicted in Figure 15, are often due to smaller

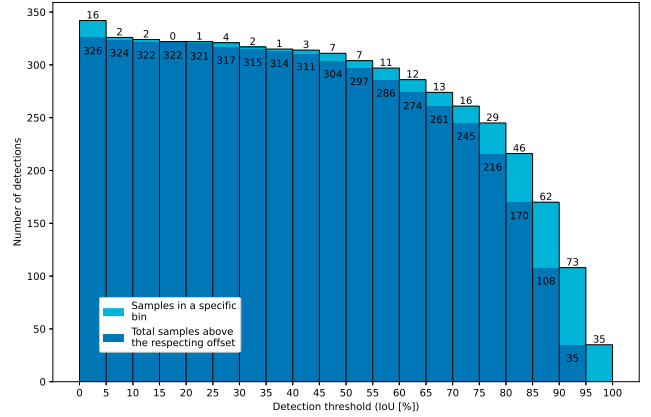


Figure 14: Absolute number of correctly detected stones for a particular IoU (best viewed magnified). The light-blue bar parts (and numbers above the bars) denote the detections in the particular range; the dark blue bars (and numbers inside the bars) describe the cumulative counts of detections with the respective threshold or higher.



Figure 15: Failure cases of INSTR during deployment on Mt. Etna.

size or because of rock instances being embedded into the gravel. Overall, the Mask-RCNN performs en par with INSTR except for the LIBS scenes, where multiple times stones are not detected due to their unexpected large size compared to the training data. While such aspects can be very well modeled with OASYS, they have to be known beforehand. This underlines the beneficial usage of INSTR for the task at hand, where generalization onto novel scenes is considerably less constrained on the training data.

Semantic Mapping—In Table 5 the percentages of semantically detected obstacles in the gridmap aggregated according to the geometric cost values found at the same positions are shown. The geometric costs are split into three categories according to the percentage of the maximum cost: low (50%-30%), medium (70%-50%), high ($\geq 70\%$). In almost half of the semantically predicted rocks, the geometric obstacle detection did not consider this as a rock. This indicates the significant amount of additional, non redundant information provided by the semantic annotations to the geometric map.

Table 5: Geometric cost at points in the environment which are detected to be semantic obstacles. The geometric cost is binned into four categories which are defined in relation to the maximum cost that is considered. High cost is defined to be higher than 70% of the maximum cost, medium cost is in the range of 70%-50% and low cost is between 50%-30%. Parts of the environment which are below 30% of maximum cost are considered to be negligible in terms of traversability.

Geometric Cost at Semantic Obstacles			
negligible	low	medium	high
47.5%	24.6 %	14.8%	13.1%

Rock Grasping—Besides the vision related metrics, a crucial performance indicator is the mask prediction quality with respect to grasping. Given the assumption, that the method’s input stereo images are the only source of information, we compute the absolute error between the estimated grasping approach point \hat{a}_i (see Equation (1)) obtained via the ground truth mask and the INSTR predicted one. We evaluated a subset of 202 rock samples from the sample collection use-case⁵ As illustrated in Figure 16, most of the samples (150) result in an error of 2 cm or even smaller (116 samples ≤ 1 cm). With rocks of interest roughly in the range of 8 to 12 cm of size and the robot hand with a maximal spread of around 15 cm, this error was small enough to successfully grasp – during our final demonstration (at the 29th of June) we were able to successfully collect all three rocks without a failure case. For larger errors the robustness of the grasp dramatically decreased, leading to rocks falling out of the hand or even not being grasped at all. Such errors are mainly caused by oversegmentation as can be exemplarily seen in the upper right image in Figure 13.

7. SUMMARY

In this work we presented three autonomous robotic capabilities leveraging a rock instance segmentation approach. The discussed applications highlighted the necessity of a robust and precise rock detector for planetary exploration missions. The learning-based approach, which exploits stereo imagery, was deployed on a Jetson TX2 on our two prototype rovers running at a speed of ~ 1 Hz. We conducted experiments at a Moon analog site on Mt. Etna and successfully showed the suitability of our applied components. The final evaluation of our rock segmentation method on the challenging lunar-analogue Etna data shows promising results. It should demonstrate the ability to detect, analyze and collect rocks in the context of a field test. Nevertheless, it also showed shortcomings and gave valuable insights for future missions.

⁵Please note, that not all of these stone samples were actually picked during the mission. They should only give a general impression of the mask quality.

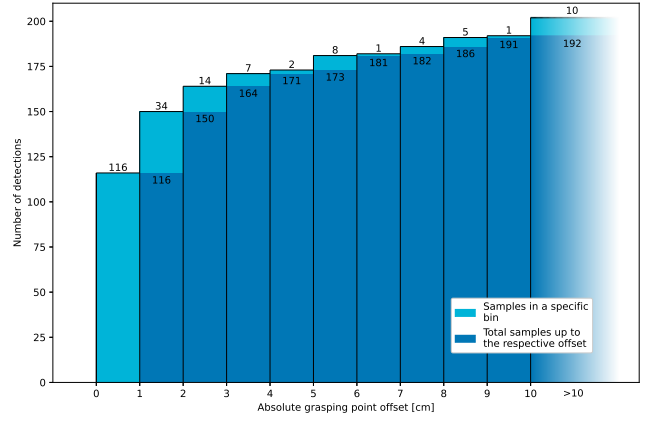


Figure 16: Absolute offset between grasping point obtained via the ground truth mask and the predicted mask in the camera coordinates.

We will address current weaknesses such as the partial mis-detection of bedrocks, improved detection within the navigation context and further increase the robustness of the presented components. This includes the fusion of several frames over time, and a more extensive post-processing to reduce implausible detection cases.

ACKNOWLEDGMENTS

We would like to thank especially the Helmholtz association for understanding and supporting the postponement of the project twice in a row. Furthermore, we would also like to thank the local entities in Italy for understanding our proposal and supporting the missions with logistics, accommodations and licenses. The “Parco di Etna”, “INGV Catania”, “Funevia di Etna”, “Giro Calabria”, “Glasser Sondertransporte” and the “Baia Verde Hotel” are given credits as well. This work was supported by the Helmholtz Association, project alliance ROBEX (contract no. HA-304), project ARCHES (contract no. ZT-0033), and project iFOODis (contract no. KA-HSC-06_iFOODis).

REFERENCES

- [1] B. K. Muirhead, A. Nicholas, and J. Umland, “Mars Sample Return Mission Concept Status,” in *2020 IEEE Aerospace Conference*, Mar. 2020, pp. 1–8.
- [2] V. Gor, E. Mjolsness, R. Manduchi, R. Castano, and R. Anderson, “Autonomous Rock Detection for Mars Terrain,” in *AIAA Space 2001 Conference and Exposition*. American Institute of Aeronautics and Astronautics, Aug. 2001.
- [3] N. Kouadria, K. Mechouek, S. Harize, and N. Doghmane, “Region-of-interest Based Image Compression Using the Discrete Tchebichef Transform in Wireless Visual Sensor Networks,” *Computers & Electrical Engineering*, vol. 73, pp. 194–208, Jan. 2019.
- [4] K. L. Wagstaff, R. Castano, S. Dolinar, M. Klimesh, and R. Mukai, “Science-based Region-of-Interest Image Compression,” in *35th Lunar and Planetary Science Conference*, 2004.
- [5] D. R. Thompson, T. Smith, and D. Wettergreen, “Data

- Mining During Rover Traverse: From Images to Geologic Signatures,” in *International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS)*, 2005, p. 8.
- [6] R. Castano, M. Judd, T. Estlin, R. Anderson, D. Gaines, A. Castano, B. Bornstein, T. Stough, and K. Wagstaff, “Current results from a rover science data analysis system,” in *2005 IEEE Aerospace Conference*, Mar. 2005, pp. 356–365.
 - [7] D. Thompson, S. Niekum, T. Smith, and D. Wettergreen, “Automatic detection and classification of features of geologic interest,” in *2005 IEEE Aerospace Conference*, Mar. 2005, pp. 366–377.
 - [8] M. Durner, W. Boerdijk, M. Sundermeyer, W. Friedl, Z.-C. Márton, and R. Triebel, “Unknown Object Segmentation from Stereo Images,” in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2021, pp. 4823–4830.
 - [9] K. Di, Z. Yue, Z. Liu, and S. Wang, “Automated Rock Detection and Shape Analysis from Mars Rover Imagery and 3D Point Cloud Data,” *Journal of Earth Science*, vol. 24, no. 1, pp. 125–135, Feb. 2013.
 - [10] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 2980–2988.
 - [11] P. Lehner, R. Sakagami, W. Boerdijk, A. Dömel, M. Durner, G. Franchini, A. F. Prince, K. Lakatos, D. L. Risch, L. Meyer, B. Voderkmayer, E. Dietz, S. Frohmann, F. Seel, S. Schröder, H.-W. Hübers, A. Albu-Schaffer, and A. Wedler, “Mobile Manipulation of a Laser-induced Breakdown Spectrometer (LIBS) for Planetary Exploration,” in *2023 IEEE Aerospace Conference (under review)*, 2023.
 - [12] S. Maurice, R. C. Wiens, M. Saccoccio, B. Barraclough, O. Gasnault, O. Forni, N. Mangold, D. Baratoux, S. Bender, G. Berger *et al.*, “The ChemCam Instrument Suite on the Mars Science Laboratory (MSL) Rover: Science Objectives and Mast Unit Description,” *Space Science Reviews*, vol. 170, no. 1, pp. 95–166, Sep. 2012.
 - [13] M. Panzirsch, H. Singh, M. Stelzer, M. J. Schuster, C. Ott, and M. Ferre, “Extended Predictive Model-Mediated Teleoperation of Mobile Robots through Multilateral Control,” in *2018 IEEE Intelligent Vehicles Symposium (IV)*, Jun. 2018, pp. 1723–1730.
 - [14] N. Y. Lii, D. Leidner, P. Birkenkamp, B. Pleintinger, R. Bayer, and T. Krueger, “Toward scalable intuitive telecommand of robots for space deployment with METERON SUPVIS Justin,” in *The 14th Symposium on Advanced Space Technologies for Robotics and Automation (ASTRA)*. Leiden, The Netherlands: European Space Agency, Jun. 2017.
 - [15] A. Wedler, M. Vayugundla, H. Lehner, P. Lehner, M. J. Schuster, S. G. Brunner, W. Stürzl, A. Dömel, H. Gmeiner *et al.*, “First Results of the ROBEX Analogue Mission Campaign: Robotic Deployment of Seismic Networks for Future Lunar Missions,” in *Proceedings of the International Astronautical Congress, IAC*, vol. 68. Adelaide, Australia: International Astronautical Federation (IAF), Sep. 2017.
 - [16] M. J. Schuster, M. G. Müller, S. G. Brunner, H. Lehner, P. Lehner, R. Sakagami, A. Dömel, L. Meyer, B. Voderkmayer, R. Giubilato *et al.*, “The ARCHES Space-Analogue Demonstration Mission: Towards Heterogeneous Teams of Autonomous Robots for Collaborative Scientific Sampling in Planetary Exploration,” *IEEE Robotics and Automation Letters (RA-L)*, vol. 5, no. 4, pp. 5315–5322, Oct. 2020.
 - [17] A. Wedler, M. G. Müller, M. Schuster, M. Durner, S. Brunner, P. Lehner, H. Lehner, A. Dömel, M. Vayugundla, F. Steidle *et al.*, “Preliminary Results for the Multi-Robot, Multi-Partner, Multi-Mission, Planetary Exploration Analogue Campaign on Mount Etna,” in *Proceedings of the International Astronautical Congress, IAC*, Oct. 2021.
 - [18] M. Vayugundla, T. Bodenmüller, M. J. Schuster, M. G. Müller, L. Meyer, P. Kenny, F. Schuler, M. Bihler, W. Stürzl, B.-M. Steinmetz, J. Langwald *et al.*, “The MMX Rover on Phobos: The Preliminary Design of the DLR Autonomous Navigation Experiment,” in *2021 IEEE Aerospace Conference*, Mar. 2021, pp. 1–18.
 - [19] D. R. Thompson and R. Castano, “Performance Comparison of Rock Detection Algorithms for Autonomous Planetary Geology,” in *2007 IEEE Aerospace Conference*, Mar. 2007, pp. 1–9.
 - [20] V. C. Gulick, R. L. Morris, M. A. Ruzon, and T. L. Roush, “Autonomous Image Analyses During the 1999 Marsokhod Rover Field Test,” *Journal of Geophysical Research: Planets*, vol. 106, no. E4, pp. 7745–7763, Apr. 2001.
 - [21] P. Viola and M. Jones, “Rapid Object Detection Using a Boosted Cascade of Simple Features,” in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1. IEEE Comput. Soc, 2001, pp. I–511–I–518.
 - [22] C. Shang and D. Barnes, “Fuzzy-rough Feature Selection Aided Support Vector Machines for Mars Image Classification,” *Computer Vision and Image Understanding*, vol. 117, no. 3, pp. 202–213, Mar. 2013.
 - [23] A. Rashno, M. Saraee, and S. Sadri, “Mars Image Segmentation with Most Relevant Features among Wavelet and Color Features,” in *2015 AI & Robotics (IRANOPEN)*, Apr. 2015, pp. 1–7.
 - [24] R. Castano, T. Estlin, D. Gaines, C. Chouinard, B. Bornstein, R. C. Anderson, M. Burl, D. Thompson, A. Castano, and M. Judd, “Onboard Autonomous Rover Science,” in *2007 IEEE Aerospace Conference*, Mar. 2007, pp. 1–13.
 - [25] X. Gong and J. Liu, “Rock Detection via Superpixel Graph Cuts,” in *2012 19th IEEE International Conference on Image Processing*, Sep. 2012, pp. 2149–2152.
 - [26] X. Xiao, H. Cui, M. Yao, and Y. Tian, “Autonomous Rock Detection on Mars through Region Contrast,” *Advances in Space Research*, vol. 60, no. 3, pp. 626–635, Aug. 2017.
 - [27] J. Fox, R. Castano, and R. Anderson, “Onboard Autonomous Rock Shape Analysis for Mars Rovers,” in *Proceedings, IEEE Aerospace Conference*, vol. 5. IEEE, 2002, pp. 5–2052.
 - [28] J. Yang and Z. Kang, “A Gradient-Region Constrained Level Set Method for Autonomous Rock Detection from Mars Rover Image,” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-2/W13, pp. 1479–1485, Jun. 2019.
 - [29] R. Grimm, M. Grotz, S. Ottenhaus, and T. Asfour, “Vision-Based Robotic Pushing and Grasping for Stone Sample Collection under Computing Resource Con-

- straints,” in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2021, pp. 6498–6504.
- [30] X. Ran, L. Xue, Y. Zhang, Z. Liu, X. Sang, and J. He, “Rock Classification from Field Image Patches Analyzed Using a Deep Convolutional Neural Network,” *Mathematics*, vol. 7, no. 8, p. 755, Aug. 2019.
- [31] F. Schenk, A. Tscharf, G. Mayer, and F. Fraundorfer, “Automatic Muck Pile Characterization from UAV Images,” *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-2/W5, pp. 163–170, May 2019.
- [32] F. Furlán, E. Rubio, H. Sossa, and V. Ponce, “Rock Detection in a Mars-Like Environment Using a CNN,” in *Pattern Recognition*, ser. Lecture Notes in Computer Science, J. A. Carrasco-Ochoa, J. F. Martínez-Trinidad, J. A. Olvera-López, and J. Salas, Eds. Springer International Publishing, 2019, pp. 149–158.
- [33] K. Wagstaff, Y. Lu, A. Stanboli, K. Grimes, T. Gowda, and J. Padams, “Deep Mars: CNN Classification of Mars Imagery for the PDS Imaging Atlas,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, Apr. 2018.
- [34] J. Zhang, L. Lin, Z. Fan, W. Wang, and J. Liu, “S5Mars: Self-Supervised and Semi-Supervised Learning for Mars Segmentation,” Jul. 2022.
- [35] R. M. Swan, D. Atha, H. A. Leopold, M. Gildner, S. Oij, C. Chiu, and M. Ono, “AI4MARS: A Dataset for Terrain-Aware Autonomous Driving on Mars,” in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jun. 2021, pp. 1982–1991.
- [36] P. Furgale, P. Carle, J. Enright, and T. D. Barfoot, “The Devon Island rover navigation dataset,” *The International Journal of Robotics Research*, vol. 31, no. 6, pp. 707–713, May 2012.
- [37] M. Vayugundla, F. Steidle, M. Smisek, M. J. Schuster, K. Bussmann, and A. Wedler, “Datasets of Long Range Navigation Experiments in a Moon Analogue Environment on Mount Etna,” in *ISR 2018; 50th International Symposium on Robotics*, Jun. 2018, pp. 1–7.
- [38] L. Meyer, M. Smisek, A. Fontan Villacampa, L. Oliva Maza, D. Medina, M. Schuster, F. Steidle, M. Vayugundla, M. G. Müller, B. Rebele, A. Wedler, and R. Triebel, “The MADMAX Data Set for Visual-inertial Rover Navigation on Mars,” *Journal of Field Robotics*, Mar. 2021.
- [39] A. Jain, J. Balaram, J. Cameron, J. Guineau, C. Lim, M. Pomerantz, and G. Sohl, “Recent Developments in the ROAMS Planetary Rover Simulation Environment,” in *2004 IEEE Aerospace Conference*, vol. 2, Mar. 2004, pp. 861–876 Vol.2.
- [40] M. Denninger, M. Sundermeyer, D. Winkelbauer, Y. Zidan, D. Olefir, M. Elbadrawy, A. Lodhi, and H. Katam, “BlenderProc,” *arXiv:1911.01911 [cs]*, Oct. 2019.
- [41] M. G. Müller, M. Durner, A. Gawel, W. Stürzl, R. Triebel, and R. Siegwart, “A Photorealistic Terrain Simulation Pipeline for Unstructured Outdoor Environments,” in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Sep. 2021, pp. 9765–9772.
- [42] M. Sewtz, H. Lehner, Y. Fanger, J. Eberle, M. Wudenska, M. G. Müller, T. Bodenmüller, and M. J. Schuster, “UR-Sim - A Versatile Robot Simulator for Extra-Terrestrial Exploration,” in *2022 IEEE Aerospace Conference*, Mar. 2022, pp. 1–14.
- [43] W. Boerdijk, M. G. Müller, M. Durner, M. Sundermeyer, W. Friedl, A. Gawel, W. Stürzl, Z.-C. Marton, R. Siegwart, and R. Triebel, “Rock Instance Segmentation from Synthetic Images for Planetary Exploration Missions,” p. 3, 2021.
- [44] J. Wang and E. Olson, “AprilTag 2: Efficient and robust fiducial detection,” in *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Daejeon, South Korea: IEEE, Oct. 2016, pp. 4193–4198.
- [45] C. Nissler, S. Büttner, Z.-C. Marton, L. Beckmann, and U. Thomasy, “Evaluation and improvement of global pose estimation with multiple AprilTags for industrial manipulators,” in *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*, Sep. 2016, pp. 1–8.
- [46] C. Nissler, *Environment- and Self-Modeling through Camera-Based Pose Estimation*, U. Thomas, Ed. Shaker Verlag, Dec. 2019, vol. 2.
- [47] W. Boerdijk, M. Durner, M. Sundermeyer, and R. Triebel, “Towards Robust Perception of Unknown Objects in the Wild,” in *ICRA Workshop on Robotic Perception and Mapping: Emerging Techniques*, 2022.
- [48] M. J. Schuster, S. G. Brunner, K. Bussmann, S. Büttner, A. Dömel, M. Hellerer, H. Lehner, P. Lehner, O. Porges, J. Reill *et al.*, “Towards Autonomous Planetary Exploration: The Lightweight Rover Unit (LRU), its Success in the SpaceBotCamp Challenge, and Beyond,” *Journal of Intelligent & Robotic Systems*, vol. 93, no. 3-4, pp. 461–494, Mar. 2019.
- [49] A. Wedler, B. Rebele, J. Reill, M. Suppa, H. Hirschmüller, C. Brand, M. Schuster, B. Vodermayr, H. Gmeiner, A. Maier *et al.*, “LRU – Lightweight Rover Unit,” in *Symposium on Advanced Space Technologies in Robotics and Automation (ASTRA)*, May 2015.
- [50] H. Hirschmüller, “Semi-Global Matching - Motivation, Developments and Applications,” in *Photogrammetric Week 11*, D. Fritsch, Ed. Wichmann, Sep. 2011, pp. 173–184.
- [51] P. Lehner and A. Albu-Schäffer, “The Repetition Roadmap for Repetitive Constrained Motion Planning,” *IEEE Robotics and Automation Letters*, vol. 3, no. 4, pp. 3884–3891, Oct. 2018.
- [52] W. Boerdijk, M. M. Müller, M. Durner, and R. Triebel, “ReSyRIS: A Real-Synthetic Rock Instance Segmentation Dataset for Training and Benchmarking,” in *2023 IEEE Aerospace Conference (under review)*, 2023.
- [53] A. Hornung, K. M. Wurm, M. Bennewitz, C. Stachniss, and W. Burgard, “OctoMap: an Efficient Probabilistic 3D Mapping Framework Based on Octrees,” *Autonomous Robots*, vol. 34, no. 3, pp. 189–206, Apr. 2013.
- [54] A. Wedler, M. G. Müller, M. J. Schuster, M. Durner, P. Lehner, A. Dömel, M. Vayugundla, F. Steidle, R. Sakagami, L. Meyer *et al.*, “Finally! Insights into the ARCHES Lunar Planetary Exploration Analogue Campaign on Etna in summer 2022,” in *Proceedings of the International Astronautical Congress, IAC*, vol. 73. Paris, France: International Astronautical Federation, IAF, 2022.

- [55] E. Staudinger, R. Pöhlmann, Z. Siwei, A. Dömel, M. J. Schuster, A. Dammann, and A. Wedler, “Radio-Localization and Multi-Robot Technologies for Low-Frequency Radio Arrays: Results from a Space Analogue Campaign on Mt. Etna,” in *Proceedings of the International Astronautical Congress, IAC*. Paris, France: International Astronautical Federation, IAF, Sep. 2022.
- [56] A. Kirillov, K. He, R. Girshick, C. Rother, and P. Dollar, “Panoptic Segmentation,” in *Proceedings of the 2019 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2019*. IEEE Comput. Soc, 2019, pp. 9404–9413.

BIOGRAPHY



Maximilian Durner received his B.Sc. and M.Sc. degree in Electrical Engineering from the Technical University of Munich in 2014 and 2016, partially studying at the Politecnico di Torino, Italy and the Universidad Nacional de Bogota, Colombia. Since then he is a researcher at the Institute of Robotics and Mechatronics, German Aerospace Center (DLR). He is the leader of the research group on semantic scene analysis, where he focuses on object-centric perception for mobile manipulation.



Wout Boerdijk is a PhD student at the Technical University of Munich and a research scientist at the German Aerospace Center, where he is part of the Perception and Cognition department in the Institute of Robotics and Mechatronics. His research interests include computer vision methods for learning of and interacting with objects.



Yunis Fanger works at the Institute for Robotics and Mechatronics at the German Aerospace Center (DLR) as a scientific researcher since 2020. He received his M.Sc. degree in electrical engineering from the Technical University of Munich in 2019 with the specialization on robotics and automation. His research focuses on the topic of semantic mapping in distributed robotic systems.



Ryo Sakagami is a researcher at the Department of Cognitive Robotics at the Institute of Robotics and Mechatronics, German Aerospace Center (DLR) since 2020. His research focus is on the world model for autonomous, intelligent robotic systems, especially with mobile manipulation capabilities. He received his master’s degree in engineering from the University of Tokyo in 2020.



David Lennart Risch completed his B.Sc. as part of a dual studies program at the German Aerospace Center (DLR) in 2021. He has continued working in the department of Perception and Cognition at the Institute of Robotics and Mechatronics and is currently studying for his M.Sc. in Robotics, Cognition, Intelligence at the Technical University of Munich (TUM). His research interests include online modeling of the environment to improve the robustness of manipulation tasks.



Rudolph Triebel leads the department of Perception and Cognition at the DLR Institute of Robotics and Mechatronics. He received his PhD in computer science in 2007 from the University of Freiburg, Germany and the habilitation in 2015 from Technical University of Munich (TUM). Before working at DLR, he was a postdoctoral researcher at ETH Zurich and at the University of Oxford, UK. From 2013 to 2021, he was also appointed as a lecturer in computer science at TU Munich. Since the beginning of 2022, he is appointed as a guest professor in the TUM School of Engineering and Design.



Armin Wedler received his Diploma in Mechanical Engineering and Bachelor in Robotics from Leibniz University of Hanover in 2004 and his PhD in 2010. Starting in 2006, he worked for Leibniz University of Hanover until he joined the Institute of Robotics and Mechatronics, German Aerospace Center (DLR), in 2008. Since then, he has been focusing on the design and development of advanced space robotics, planetary exploration and mobile systems. He has worked as project manager, technical manager and member of project teams for numerous industrial and scientific robotic projects and is coordinator of the DLR groups for mobile robots (since 2015) and for the planetary exploration group (since 2014).