AGILE: GIScience Series, 3, 55, 2022. https://doi.org/10.5194/agile-giss-3-55-2022
Proceedings of the 25th AGILE Conference on Geographic Information Science, 2022.
Editors: E. Parseliunas, A. Mansourian, P. Partsinevelos, and J. Suziedelyte-Visockiene.
This contribution underwent peer review based on a full paper submission.
© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.

Off-Road Navigation Maps for Robotic Platforms using Convolutional Neural Networks

Raphael Prinz[1], Rizwan Bulbul[1], Joahnnes Scholz[1], Matthias Eder[2], and Gerald Steinbauer-Wagner[2]

[1]RG Geoinformation, Institute of Geodesy, Graz University of Technology, 8010 Graz, Austria
[2]Institute of Software Technology, Graz University of Technology, 8010 Graz, Austria

Correspondence: Johannes Scholz (johannes.scholz@tugraz.at)

Abstract. As part of AMADEE-20, an integrated Mars analog field mission in the Negev Desert in Israel conducted by the Austrian Space Forum, an exploration cascade for the remote sensing of extraterrestrial terrain was implemented. For this purpose, aerial robots were conceptualized, which were used in an iterative process to generate a navigational map for an autonomous ground vehicle. This work presents the process for generating navigation maps using multiple aerial image sources from satellites as well as from low orbiting aerial vehicles. First, Deep Learning methods are used to analyze a high altitude aerial images of a large area, creating a basis map for mission planning and navigation. Second, high resolution unmanned aerial vehicle (UAV) images were recorded on low altitude for a pre-defined area of interest, processed with Deep-Learning and Structure from Motion and used to update the basis map. This approach results in a high accuracy navigation map for autonomous, off-road robot navigation. Experiments during the AMADEE-20 mission in the Israeli Negev Desert validated the proposed methods by sending an autonomous ground vehicle through the environment using the generated map.

Keywords. remote sensing, deep learning, robotics, GeoAI

1 Introduction

The Mars Analog Research Mission (AMADEE) 2020, organized by Austrian Space Forum (OeWF) in cooperation with the Israel Space Agency, elaborated on an operational workflow for the exploration cascade of a future Mars mission, using robotic systems and remote sensing. The mission took place in October 2021 and was conducted in the Ramon Crater, Negev Desert in Israel. For the exploration cascade, aerial imagery is used to generate a navigation map in an iterative process which can be used by autonomous ground vehicles (AGVs) for global navigation and ground exploration. This work presents the iterative process to generate a navigational map for AGVs using aerial imagery retrieved from different altitudes and with different resolution.

The remainder of this paper is structured as follows. In the first section, the importance of Mars analog research is contextualized, and the role of Geosciences for the exploration of Mars described. This is followed by the exploration cascade of the AMADEE-20 mission, the preparation, and classification of the aerial images during the mission, followed by the creation of the navigational map. Concluding, the experiments conducted during the mission and future research topics are discussed.

1.1 Mars Analog Research

Terrestrial analog research missions provide the opportunity to test space exploration missions directly in the field, i.e equipment, strategies/procedures, and human behavior. They are an established tool applied in NASA's D-RATS and HI-SEAS missions, as well as the European MOONWALK and PANGEA projects, and an essential part of space exploration. Human-robotic Mars missions are likely to be launched within the next 2 to 3 decades and will include surface missions with at least 1 month in duration (Groemer and Ozdemir, 2020).

Using the keyword "Mars analog research" in Google Scholar yields 1,620 results up to the year 2010. This number is increased to 4,730 when applying the same search up to the year 2021. The Austrian Space Forum (OeWF) has conducted 12 Mars analog field campaigns from 2006 to 2018 (Groemer and Ozdemir, 2020).

1.2 AMADEE-20 Exploration Cascade

The "Exploration Cascade" (EC) is a term used by OeWF for tactically optimizing scientific procedures for pursuing a predefined scientific question. The strategic aims may be set before the flight mission architecture development (Groemer and Ozdemir, 2020).

Scientific operations in multidisciplinary campaigns present challenges in the coordination workflows. The workflow defines when and where to deploy instruments, and when the data are to be expected to arrive at the Mission Support Center on Earth. Of particular interest is the question of how quick data processing can lead to knowledge that influences decision-making of the flight planners. The EC was first demonstrated during the AMADEE-18 field campaign in Oman Groemer et al. (2020) and was now further researched in AMADEE-20. During AMADEE-20 the EC with the experiments therein, aims to study the geology of the analog site by investigating geological samples and their composition. These results shall provide insights into geomorphological processes, whether aqueous formations have been present and what forms of life could have dominated during different periods of time.

The planning and control of the AMADEE-20 mission including the exploration cascade is performed by the Remote Science Support (RSS) Team of the OeWF, which is part of the Mission Support Center (MSC) located in Innsbruck/Austria. The RSS tasks include the tactical sequence of instrument deployment (i.e. spatial and temporal question in relation to the location of pre-defined regions of interest). The deployment, performed by human analog astronauts, is scheduled in the daily and mission support activities.

2 Semantic Segmentation of Aerial Images

Exploration of the Martian surface requires identifying interesting locations based on satellite reconnaissance. This task is the basis for mission planning, which includes detailed navigational planning for the mobile robotic platform. Convolutional Neural Networks (CNN) with a U-NET architecture (Ronneberger et al., 2015) are trained on a high resolution orthoscopic image of the mission environment to determine the traversability of the ground for the rover. The parameters for this analysis include obstacles as well as rims, tracks, and roads. For the AMADEE-20 mission area, bushes, buildings, boulders, and bodies of water are identified and classified as non-traversable for the rover, whereas roads and homogeneous terrain are suitable for rover movement. The objective is, to generate a map of the area, that provides navigational information including slope as well as a detailed surface model of the area. Hence, this provides accurate terrain information concerning the suitability for the movement of the rover.

Two Models are used for the segmentation. The first model was trained on larger features (macro-features), in order to extract larger rock formation and river beds. The second one (micro-features), was designated to detect small obstacles, like bushes and boulders as well as water bodies, tracks, and roads. Bushes and water bodies, roads and gravel roads represent obstacles, not present on the Martian surface, but due to their considerable influence on the navigation in the test area, they need to be included as well.

Before the mission start, a map of the test area is produced using the low-resolution, aerial images and elevation data, which are provided before the mission start. The generated low-resolution map is used for global route planning before the mission. During the experiments, aerial images of the test area in the vicinity of the planned route were taken by a UAV to produce a high-resolution, up-to-date map for the driving mission of the rover. The process used to generate the maps is shown in Figure 1. In both maps, the features derived from the neural networks are combined with elevation data. The macro-feature model was only deployed for the low-resolution map, while the micro-feature model was also combined with the high-resolution data. The detected obstacles were integrated with obstacles derived from a normalized difference surface model.

As depicted in the process defined in Figure 1, the information derived from the neural networks is combined with elevation data and standard-deviation filtered images to generate two navigational maps, on two resolution levels.

2.1 Convolutional Neural Networks

Deep Convolutional Neural models roughly mimic the nature of mammalian visual cortex, and by are the most promising for visual tasks (Ciresan et al., 2011). CNN is superior to other deep network algorithms due to its ability to preserve the geometry of the image (i.e. the 2-D format) while maintaining the interconnection between pixels and spatial information (Rezaee et al., 2018).

In CNN's, Filters are used to extract localized features from an input image. The filter outputs are repeatedly subsampled and re-filtered resulting in a deep feed-forward network architecture in which the output feature vectors are eventually classified. These filters are randomly initialized and changed in a supervised way using back-propagation (Ciresan et al., 2011).

The filters are followed by a so-called pooling layer, that reduces the size of data, and it preserve the most important information, such as the geometry of the input data. In each pooling layer, a particular number is determined by subsampling, such as using maximum value or linar combination (Ciresan et al., 2011; Lee et al., 2016).

Inside a CNN, images are transformed into a sequence of increasingly abstract representations. The deeper one descends into this hierarchy, the spatial resolution of the representation decreases while the diversity of fea-

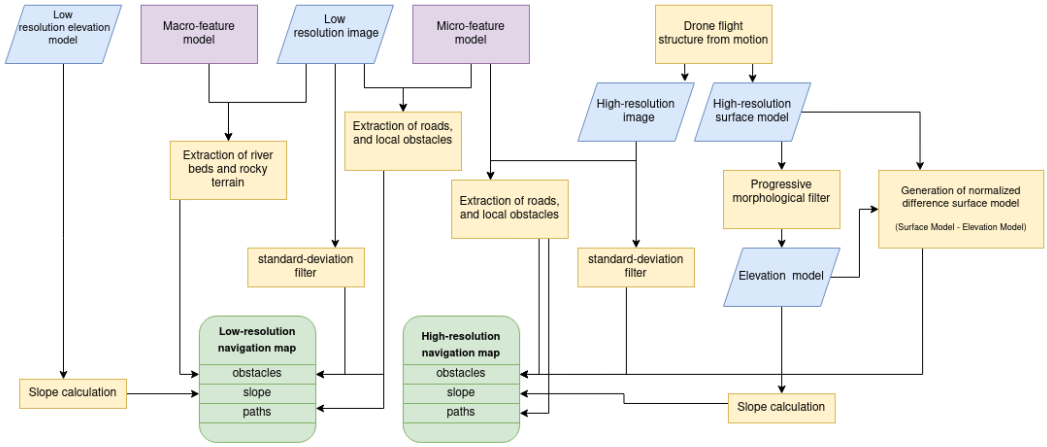

Figure 1. Process to produce navigational maps from aerial images. The low-resolution navigation map can be used for pre-mission route planning. The high resolution map is generated from UAV footage, and provides accurate and up-to-date information for the driving experiments.

tures and their invariance to transformation increase. This way, CNN's learn contextual features at different scales (Behnke, 2003).

CNNs are often used in classification tasks, that classify the image in a single class label. U-NET's (utilized in this particular work) are designed for biomedical image processing, and require the localization of a label, i.e., every pixel in the output image should be assigned a label (Ronneberger et al., 2015).

Labeling is also a recurrent problem in remote sensing (Maggiori et al., 2016). As Maggiori et al. (2016) and Rezaee et al. (2018) successfully show, CNNs can be utilized for the classification of satellite imagery. For the classification of complex wetlands, Rezaee et al. (2018) demonstrate that CNNs can have a better classification performance than Random Forest, even with a pixel resolution of 5m.

2.2 Model Architecture

U-NET was designed with limited amounts of labeled sample-data in mind (Ronneberger et al., 2015). Thus minimizing the work of labeling data manually. The network architecture consists of a contracting path, down-sampling the input, while increasing the feature channels, and an up-sampling path that reduces the feature-channels while increasing the spatial resolution (Ronneberger et al., 2015).

Both the macro- and micro-feature models followed the same architecture. The model was implemented using the Keras and TensorFlow Python libraries. The macro-feature model uses an input size of 512x512 pixel, while the micro-feature model uses 128×128 pixel. The down-sampling path of the model consists of two 3×3 convolutions followed by a rectified linear unit (ReLU) and a max pooling layer with a stride of 2. This schema (convolution-block) is repeated 5 times with an increasing size of feature channels. The 16 filter channels at the start, are increased in increments of multiples of two, until a maximum of 256 channels is reached. The up-sampling path, consists of a transposed convolution layer followed by the concatenation of the corresponding output from a previous down-sampling convolution block, in order to consider features from multiple layers. This is followed by two 3×3 convolution and ReLU. This schema is repeated five times.

2.3 Training

For the training of the macro-feature model, 147 representative samples of size 512x512 pixels have been labeled manually using conventional GIS-methods. Similarly, the micro-feature model - with a sample size of 736 images, each 128×128 pixels - was labeled manually. Figure 2 shows samples used to train the models. The location of the samples are chosen with respect to the occurrence of features.

The macro-feature model uses 118 samples for training and 29 samples for validation, for the micro-feature model 589 samples are used for training and 147 for validation. The number of samples in the micro-feature model is higher because the included features require greater accuracy of the model for navigation purposes of the rover. While rough terrain and river basins are important for global route planning, the detection of micro-features is of immediate importance for the local scope of the driving experiments. For example, a dry riverbed can be traversed by the rover but should generally be avoided, while a large boulder should always be avoided.

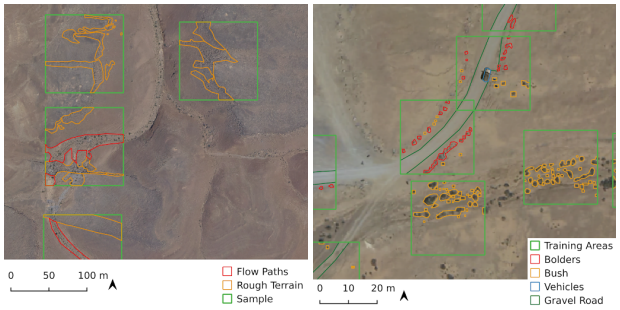

Figure 2. Samples for the training of the macro-feature model (left), the green rectangles denote the extent of the sample while red areas show flow paths and orange areas rough/rocky terrain. The samples for the micro-feature model (right), this includes boulders (red) as well as bushes (orange), gravel roads (dark green) and vehicles (blue). Vehicles were detected to later remove them from roads.

The input data is augmented with a random flip around the horizontal axis and a random brightness adjustment to avoid overfitting.

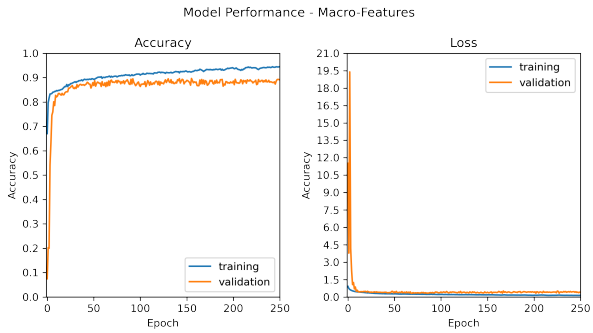

Figure 3. Model performance of the macro-feature model.

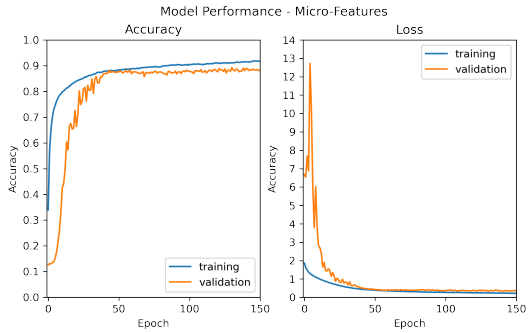

Figure 4. Model performance of the micro-feature model.

Training on both the macro- and micro-feature models is done for 250 epochs, the macro-feature model achieves an accuracy of 0.9443 on the training dataset and an accuracy of 0.8920 on the validation dataset.

The micro-feature model is trained for 250 Epochs and reaches a maximum accuracy of 0.9284 on the training dataset and 0.8936 on the validation dataset (see figure 4). A visual interpretation, depicted in Figure 6, shows that the class of gravel roads contains the largest false-positives by area.

2.4 Prediction

The orthoscopic image of the AMADEE-20 mission area has a size of 250002 pixels. To classify every pixel with the trained models, the classification is done with overlapping tiles with the same size as the training images.

The macro-scale model (see Figure 5) shows problems in distinguishing between river basins and large stone fields terrain. It is assumed, that this is because of a similar texture that both display. A higher number of training samples could have improved the quality, but due to time restrictions before the mission, it was not possible to evaluate on this hypothesis. Nonetheless, numerous non-traversable areas are identified. The pixels identified as flow paths and rocky terrain areas are joined as “rough terrain”.

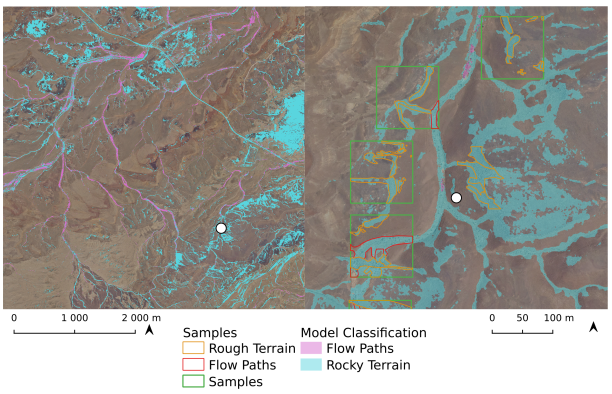

Figure 5. Flow paths and rocky terrain detected by the macro-feature model. The model has difficulties distinguishing between both features.

The micro-feature model (Figure 6) successfully identifies roads, boulders, water bodies and bushes. The separation of gravel roads from the barren land in Ramon Crater proves difficult for the CNN, as colors and textures of roads exhibit a low variability, challenging detection by the model in specific geographic regions. For further processing, the gravel roads need to be extracted manually.

2.5 Additional Information

In addition to semantic segmentation, a standard-deviation kernel-filter (Figure 7) with a size of 5×5 pixels was applied to the existing orthoscopic image. This method estimates the heterogeneity of an area. The reason behind this is, that areas with a high variance in the image provide evidence, that obstacles - like groups of boulders and rocky terrain - exist. While this method does not identify the objects on the ground, it still provides useful information and proves successful in identifying areas that should be avoided by the rover – due to their low suitability for

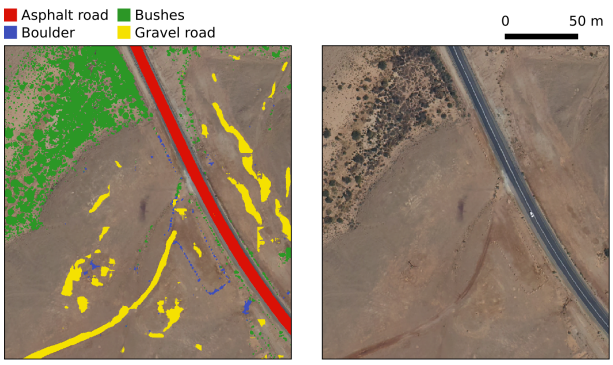

Figure 6. Features identified by the micro-feature model, asphalt roads (red), boulders (blue) and bushes (green) showed adequate accuracy in the classification while gravel roads shows poor performance.

navigation. A digital elevation model is used to calculate a slope raster image, where pixels with a value larger than 20 degrees of inclination are flagged as non-traversable by the rover.

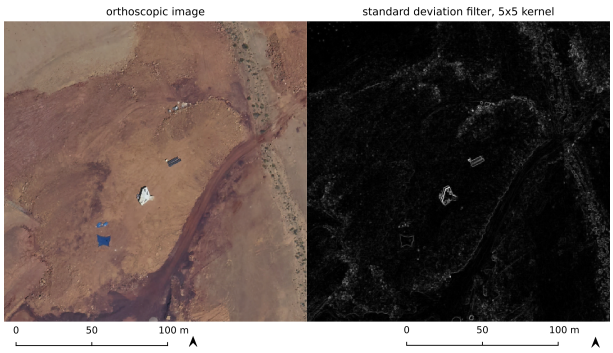

Figure 7. The area around the base area (left), the same area processed with a standard-deviation filter (5x5 pixels). The edges clearly show obstacles that can not be traversed by the robot.

2.6 Lower Altitude UAV Data

Another critical part of the EC tool chain is the deployment of UAVs in lower altitude, to obtain more details of the area of interest. While higher altitude missions provide an overview of the entire area at low resolution, high-resolution images are created with the low-altitude flights. In addition, specific zones, that are of special interest, require ego-motion free images and close up views.

For the EC in AMADEE-20, a UAV was chosen which features a high-resolution camera in the visual spectrum, and an Inertial Measurement Unit (IMU). Both sensors are used as input for visual-inertia algorithms to guide the aerial vehicle autonomously from start to destination, including take-off and landing following the provided exploration path. Lower altitude imagery was done by the AEROSCAN UAV experiment which aims to design, manufacture and test an autonomous (dawn-sunset flight time) solar-powered, convertible vertical take off and landing UAV, that carries cameras and a range of atmospheric sensors.

For a specific area of interest, the UAV is used to take 230 images of an approximately 450m by 500m area. With the we reconstruct the surface using structure from motion. The software Agisoft Metashape calculates a point cloud, a high-resolution surface model and and orthographic imagery from the collected image data (see Figure 8). The Progressive Morphological Filter developed by Keqi Zhang et al. (2003) helps to classify the point cloud into ground and none ground, allowing for further obstacle detection. Although this filter was developed with LiDAR data in mind, descent results could be achieved with the structure from motion data (Figure 8 - ground andoObstacles). After filtering, we calculate an elevation model with the obtained ground-points. The surface and the elevation model are subtracted from each other to obtain a raster map, that represents obstacles for the rover. The elevation model is utilized to compute the slope of the terrain.

Additionally, the micro-feature CNN gets tested on the orthoscopic image generated from the UAV image data. As seen in Figure 9, numerous obstacles can be identified, and a combination with the structure from motion data, resulted in a detailed map of obstacles. These obstacles are combined with the slope and standard-deviation in order to generate a navigational map for the mobile robotic platform (Figure 10).

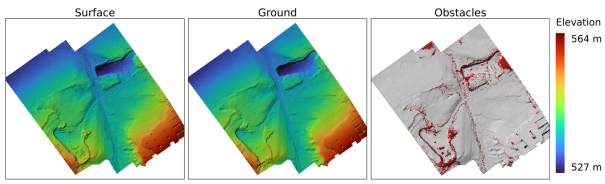

Figure 8. Three different maps created with structure from motion - the digital surface model (denoted as Surface), the digital terrain model (denoted as Ground), and the obstacles detected by subtracting the elevation model from the surface model.

2.7 Data and Software Availability

The training data and the software developed, are available on Github: https://github.com/raphi-web/AMADEE20-GeoAI/. Unfortunately, the full size orthoscopic image cannot be provided on Github, due to legal restrictions.

3 Creation of the Navigation Map

The objective of the navigation map is, to provide a detailed map of the area's traversability for the mobile robotic platform. This is done by converting the information from the navigation map directly into costs, so that

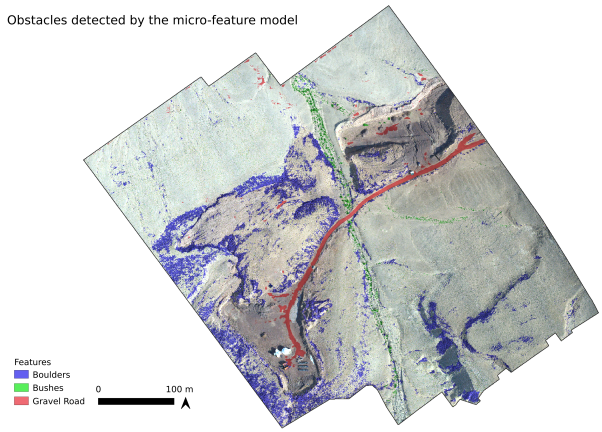

Figure 9. Navigational parameters detected by the micro-feature model. The background image is the orthographic image generated with structure from motion.

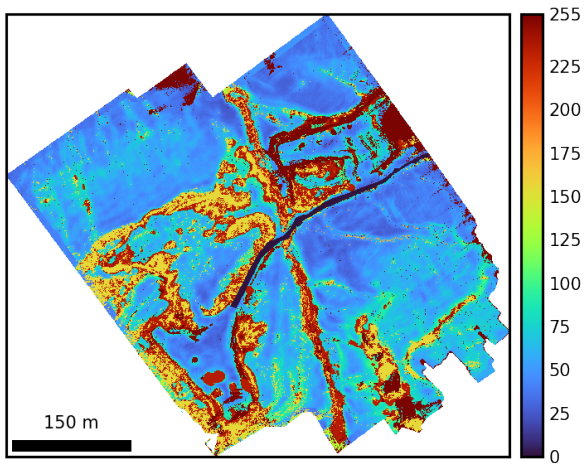

Figure 10. Navigational Map for the base station perimeter. The map is a fusion of the navigation parameters detected via GeoAI and structure from motion

they can be used to find the optimal route between two points Stentz (1997). The resulting cost map serves as basis for path planning in the project.

The navigational map (Figure 10) is represented by a single band raster image with a resolution of 0.5m per pixel and integer values ranging from 0 to 255. Value 0 denotes the lowest possible cost to traverse a pixel, e.g., the highest traceability, while the values larger than 114 denote non-traversable obstacles. Pixels that form paths are assigned the lowest values in the range from from 0-20 - with represents the slope of the pixel. Values from 21 to 114 represent offroad areas, these values were created by normalizing the slope and the standard-deviation raster between 0 and 1, adding both together and rescaling the result to the aforementioned interval. Both the data sets a) the large-extent orthographic image and b) the high-resolution UAV data are processed independently of another, resulting in two cost maps. They are merged to provide a detailed map for the field trial. Figure 10 shows an excerpt of the used cost map for navigation.

The mobile robot uses the cost map as the basis for the A* algorithm Hart et al. (1968) in order to find the shortest path to the destination point. The result of the planning component is a low-resolution path, provides a coarse direction to the desired target. This coarse resolution path is not continuously updated, only re-planned on request, as the calculation of long paths (up to 2 km on a 25 km^2 map) is a resource intensive task. Local navigation is used to consider the robot's surroundings, and kinematic constraints. As the resolution of the base map is rather poor ($0.5m$) for precise navigation, an internal 3D representation of the environment with a higher resolution (0.05m) is created using 3D LiDAR scanners and stereo cameras. These sensors constantly map the robot's surroundings to identify traversable areas in detail.

The computational cost of the calculations are as follows. The computational environment consists of a PC with an i5-10400 Hexacore with 2.9 GHz with 16GB RAM. The graphic processor is NVIDIA RTX 360 with 12GB memory. The running time for training the micro-feature model took 12min 35sec. The classification of the orthoscopic image took 1h 12min.

4 Evaluation in Field Trial

The AMADEE-20 mission area covers an area of 5×5 km around the base station, having a resolution of 0.5 m per pixel. The generated cost map, which is part of the EC for the AMADEE-20 Mars analog mission, is evaluated during the field tests in October 2021. Within the AMADEE-20 EC, the AEROSCAN UAV was launched by astronauts during EVAs, and supervised by the remaining crew in the habitat. In order to allow the in-situ exploration of the EC by the EXOSCOT rover team, the recorded images were processed into georeferenced maps using the process described in section 2.4. After the selection of a target area our approach calculated a navigation solution, which guides the robot to the selected region. During the rover's traverse, camera images and 3D point clouds (stereo camera, 3D LiDAR) together with a GNSS ground truth pose are recorded. The EXOSCOT rover was deployed 7 times, with a maximum distance of 2 km during AMADEE-20.

5 Conclusions

The experiments for the proposed robotic EC are regarded as a success story, as the EC was successfully tested during the field trials of the AMADEE-20 mission Didari et al. (2022). Here, the rover was able to explore the habitat autonomously, based on the spatial data created with the approach described in the paper. The success is measured by the fact, that the rover reached the defined targets in the habitat and made it home (i.e. to the base station)

safely, without being damaged or stuck. The presented work shows that convolutional neural networks, in conjunction with elevation data, can significantly improve the creation and the quality of navigational maps for outdoor robot navigation. Workflows that are already tested under real-world scenarios, can help to shape and manage future exploration missions of planets.

Further research works should focus on the classification accuracy of a feature-class as a factor that influences the navigation. More advanced cost maps could be created if the rover could process a cost map from a given set of input layers on the fly, and learning from previous rover deployments - i.e. which feature classes to avoid, and which can be traversed. This would include autonomous spatio-temporal learning capabilities for rovers. Weight-factors tailored to the robot's profile could be utilized to further advance the navigation. As an example, a robot with a continuous track - like a tank - could maneuver more difficult terrain than a robot with four wheels.

Acknowledgements. This work was partially supported by the Austrian Research Promotion Agency (FFG) in the project RoboNav - in the funding scheme "Austrian Space Applications Program (ASAP) 2019".

References

Behnke, S.: Hierarchical neural networks for image interpretation, no. 2766 in Lecture notes in computer science, Springer, Berlin New York, 2003.

Ciresan, D. C., Meier, U., Masci, J., Gambardella, L. M., and Schmidhuber, J.: Flexible, high performance convolutional neural networks for image classification, in: Twenty-second international joint conference on artificial intelligence, 2011.

Didari, H., Eder, M., Grömer, G., Halatschek, R., Özdemir-Fritz, S., Prinz, R., Scholz, J., and Steinbauer-Wagner, G.: The AMADEE-20 Robotic Exploration Cascade: An Experience Report, in: Advances in Service and Industrial Robotics, edited by Müller, A. and Brandstötter, M., pp. 477–484, Springer International Publishing, Cham, 2022.

Groemer, G. and Ozdemir, S.: Planetary Analog Field Operations as a Learning Tool, Frontiers in Astronomy and Space Sciences, 7, 32, https://doi.org/10.3389/fspas.2020.00032, 2020.

Groemer, G., Gruber, S., Uebermasser, S., Soucek, A., Lalla, E. A., Lousada, J., Sams, S., Sejkora, N., Garnitschnig, S., Sattler, B., and Such, P.: The AMADEE-18 Mars Analog Expedition in the Dhofar Region of Oman, Astrobiology, 20, 1276–1286, https://doi.org/10.1089/ast.2019.2031, pMID: 33179971, 2020.

Hart, P. E., Nilsson, N. J., and Raphael, B.: A Formal Basis for the Heuristic Determination of Minimum Cost Paths, IEEE Transactions on Systems Science and Cybernetics, 4, 100–107, https://doi.org/10.1109/TSSC.1968.300136, 1968.

Keqi Zhang, Shu-Ching Chen, Whitman, D., Mei-Ling Shyu, Jianhua Yan, and Chengcui Zhang: A progressive morphological filter for removing nonground measurements from airborne LIDAR data, IEEE Transactions on Geoscience and Remote Sensing, 41, 872–882, https://doi.org/10.1109/TGRS.2003.810682, 2003.

Lee, C.-Y., Gallagher, P. W., and Tu, Z.: Generalizing Pooling Functions in Convolutional Neural Networks: Mixed, Gated, and Tree, in: Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, edited by Gretton, A. and Robert, C. C., vol. 51 of *Proceedings of Machine Learning Research*, pp. 464–472, PMLR, Cadiz, Spain, https://proceedings.mlr.press/v51/lee16a.html, 2016.

Maggiori, E., Tarabalka, Y., Charpiat, G., and Alliez, P.: Fully convolutional neural networks for remote sensing image classification, in: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 5071–5074, IEEE, Beijing, China, https://doi.org/10.1109/IGARSS.2016.7730322, 2016.

Rezaee, M., Mahdianpari, M., Zhang, Y., and Salehi, B.: Deep Convolutional Neural Network for Complex Wetland Classification Using Optical Remote Sensing Imagery, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 11, 3030–3039, https://doi.org/10.1109/JSTARS.2018.2846178, 2018.

Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, arXiv:1505.04597 [cs], http://arxiv.org/abs/1505.04597, arXiv: 1505.04597, 2015.

Stentz, A.: Optimal and Efficient Path Planning for Partially Known Environments, pp. 203–220, Springer US, Boston, MA, https://doi.org/10.1007/978-1-4615-6325-9_11, 1997.