Rock Instance Segmentation from Synthetic Images for Planetary Exploration Missions

W. Boerdijk*[1,2], M. G. Müller*[1,3], M. Durner*[1,2], M. Sundermeyer[1,2], W. Friedl[1], A. Gawel[3], W. Stürzl[1], Z.-C. Márton[4], R. Siegwart[3], R. Triebel[1,2]

*Equal Contribution

[1]Institute of Robotics and Mechatronics
German Aerospace Center (DLR)
82234 Wessling, Germany
<first>.<second>@dlr.de

[2]Department of Computer Science
Technical University of Munich (TUM)
85748 Garching, Germany

[3]Autonomous Systems Lab
Swiss Federal Institute of Technology (ETH Zürich)
Switzerland

[4]Agile Robots AG
81477 Munich
Germany

Problem Description

With complexity and operation distance of space missions steadily rising, the demand of highly autonomous rovers increases likewise. An aspect of autonomous rovers that has been specifically attracting much attention from the space community is semi-autonomous sampling from celestial bodies. As a result, several missions are mainly based on returning different kind of samples, such as soil and rocks, back to Earth (e.g. [1, 2, 3]). Since communication delay and limited bandwidth often render a fully remote controlled rock sample extraction infeasible, it is necessary that the sample extracting robot is detecting samples autonomously to then send a list of potential rocks for selection and extraction to the ground team.

However, the autonomous detection of rocks is challenging, especially due to their unique and highly unstructured shape. Even more, existing datasets from planetary [4] and analog planetary environments [5, 6] do not provide the necessary annotations to train an instance segmentation approach.

To overcome the aforementioned challenges, we propose to generate photo-realistic renderings of a planetary environment using OAISYS [7] and train our Instance Stereo Transformer (INSTR) [8], as illustrated in Figure 1.

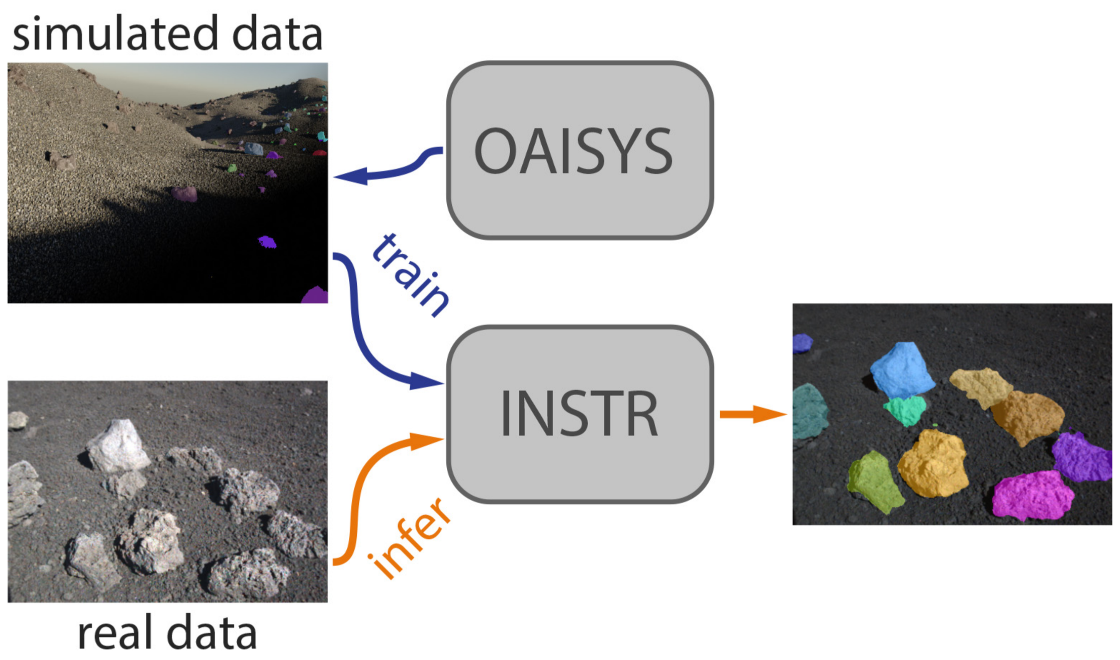

Figure 1: Illustration of the presented pipeline. OAISYS is used to create a training dataset with which INSTR is trained to segment arbitrary rocks.

INSTR for Rocks

Our aim is to segment arbitrary stone instances in an unknown environment for manipulation purposes. Such a task could readily be achieved even with limited amount of data available by fine-tuning an existing model on the task of rock instance segmentation. However, instead of solely providing RGB cues as in [7, 9], we argue for the importance of incorporating additional modalities such as depth, as stones and background often have similar color and texture. Since high-quality depth data cannot always be ensured, we propose to use the INSTR network [8], which implicitly fuses RGB and disparity information by operating on stereo images.

Originally, the network is trained on synthetic data to segment any unknown instance on a dominant horizontal surface (e.g. tables) in an indoor environment. Since rocks on a planar surface state a similar problem, the pre-trained INSTR is able to partially segment instances. To further improve, we propose to specialize the network on the underlying use-case. Therefore, we evaluate the effect of fine-tuning with photo-realistic synthetic data of stone instances and compare it to the pre-trained version.

Generating Training Data

Since datasets for the described use-case are scarce (or not available), we are using OAISYS [7] to synthesize a dataset. It is a simulator which can auto generate photo-realistic outdoor environments and is specialized for planetary use-cases. One can provide textures for the underlying terrain and a set of objects, which are scattered on the surface. To emulate a similar environment as our target domain, Mt. Etna, we use three gravel textures as terrains and 14 different kind of rocks as mesh assets. We apply the particle system option of OAISYS to distribute the rocks over the surface. To create a realistic composition, we previously adjusted the color of all assets to be similar. Finally, we configure a stereo set-up and render ~1800 color images alongside the respective ground truth depth and rock instance masks (see Figure 2).

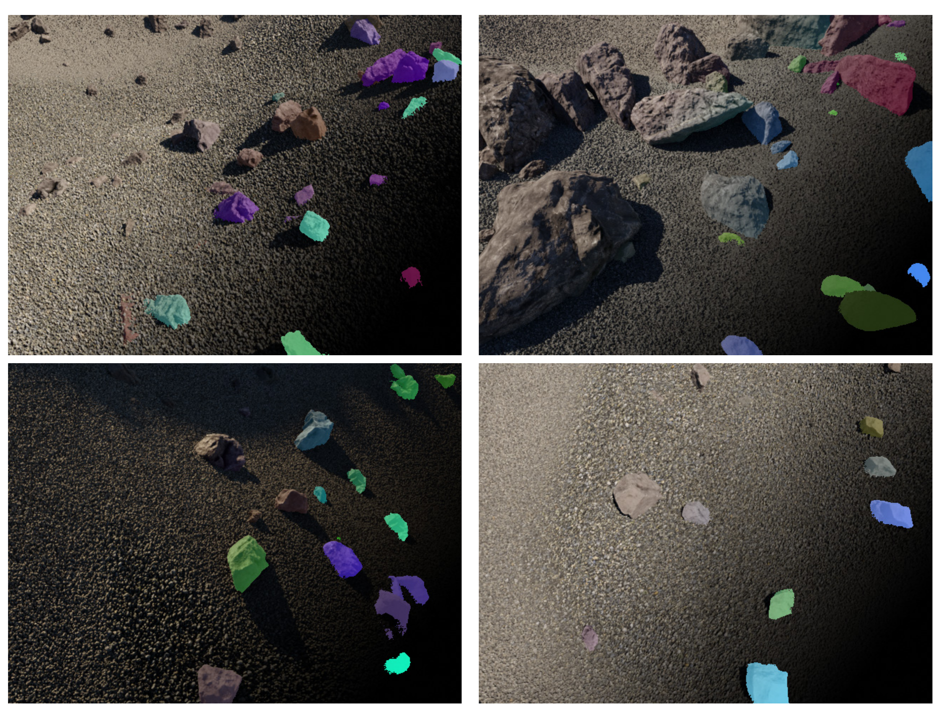

Figure 2: Example images of the training dataset created by OAISYS. Color images partly overlayed by ground truth instance segmentation map (colors are assigned randomly).

Evaluation

We compare the pre-trained version of INSTR to different fine-tuning strategies: (a) training all layers, (b) only training the decoder, (c) transformer + (b), (d) only the encoder. All models are evaluated on a real dataset (26 images, manually annotated) recorded on a site on the volcano Mt. Etna in Sicily, Italy. While (a) results in the best performance (Tab. 1), (b) and (c) vastly degrade. We hypothesize that this is due to the distance change between the terrain dataset and the original indoor dataset, which makes an adaption of the encoder weights inevitable. This is confirmed by (d), where we freeze everything except the siamese ResNet encoder, and achieve similar results to (a). See Figure 3 for quantitative results.

Table 1: *mIOU* [%] of fine-tuning approaches on Mt. Etna test data.

pre-trained	(a)	(b)	(c)	(d)
53.55	**63.88**	36.55	36.60	62.25

Figure 3: Qualitative results of INSTR pre-trained (left) and fine-tuned with the synthetic data (right) (colors area assigned randomly).

References

[1] T. Morota *et al.*, "Sample collection from asteroid (162173) ryugu by hayabusa2: Implications for surface evolution," *Science*, 2020.

[2] B. K. Muirhead, A. Nicholas, and J. Umland, "Mars sample return mission concept status," in *2020 IEEE Aerospace Conference*, 2020.

[3] M. J. Schuster *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*, 2020.

[4] K. Wagstaff *et al.*, "Deep mars: Cnn classification of mars imagery for the pds imaging atlas," *AAAI Conference on Artificial Intelligence*, 2018.

[5] L. Meyer *et al.*, "The MADMAX dataset for visual-inertial rover navigation on Mars," *Journal of Field Robotics*, 2021.

[6] M. Vayugundla *et al.*, "Datasets of long range navigation experiments in a moon analogue environment on mount etna," in *International Symposium on Robotics*, 2018.

[7] M. G. Müller *et al.*, "A Photorealistic Terrain Simulation Pipeline for Unstructured Outdoor Environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2021.

[8] M. Durner *et al.*, "Unknown Object Segmentation from Stereo Images," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2021.

[9] F. Schenk *et al.*, "Automatic much pile characterization from UAV images," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2019.

Workshop on Advances in Space Robotics and Back to Earth
IEEE/RSJ International Conference on Intelligent Robots and Systems
Online, October 01st, 2021