Uncertainty Estimation for Planetary Robotic Terrain Segmentation

Marcus G Müller[1,2], Maximilian Durner[1], Wout Boerdijk[1], Hermann Blum[2], Abel Gawel[2], Wolfgang Stürzl[1], Roland Siegwart[2], Rudolph Triebel[1]

[1]German Aerospace Center (DLR)
Institute of Robotics and Mechatronics
Münchner Str. 20, 82234 Weßling, Germany
Contact: Marcus.Mueller@dlr.de

[2]Swiss Federal Institute of Technology (ETH)
Leonhardstrasse 21
8092 Zürich

***Abstract*— Terrain Segmentation information is crucial input for current and future planetary robotic missions. Labeling training data for terrain segmentation is a difficult task and can often cause semantic ambiguity. As a result, large portion of an image usually remains unlabeled. Therefore, it is difficult to evaluate network performance on such regions. Worse is the problem of using such a network for inference, since the quality of predictions cannot be guaranteed if trained with a standard semantic segmentation network. This can be very dangerous for real autonomous robotic missions since the network could predict any of the classes in a particular region, and the robot does not know how much of the prediction to trust. To overcome this issue, we investigate the benefits of uncertainty estimation for terrain segmentation. Knowing how certain the network is about its prediction is an important element for a robust autonomous navigation. In this paper, we present neural networks, which not only give a terrain segmentation prediction, but also an uncertainty estimation. We compare the different methods on the publicly released real world Mars data from the MSL mission.**

TABLE OF CONTENTS

1. INTRODUCTION

Robotic systems are more important than ever for exploring the extraterrestrial planetary surfaces. Equipped with scientific instruments, robots can help scientists understand the origins of the solar system in places that are currently inaccessible for humans. In recent decades, planetary robots have become more autonomous with every iteration. The most advanced system regarding the level of autonomy is arguably NASA's Mars helicopter, Ingenuity [1]. On the mission, it is not possible to directly control Ingenuity by a human operator and it has to perform most of its tasks by itself. The level of autonomy will only increase in future planetary exploration missions, particularly if entire robotic teams will be deployed as some researchers envision [2]. In order to increase autonomy, the robots would need to perceive and interpret their environments better than ever before.

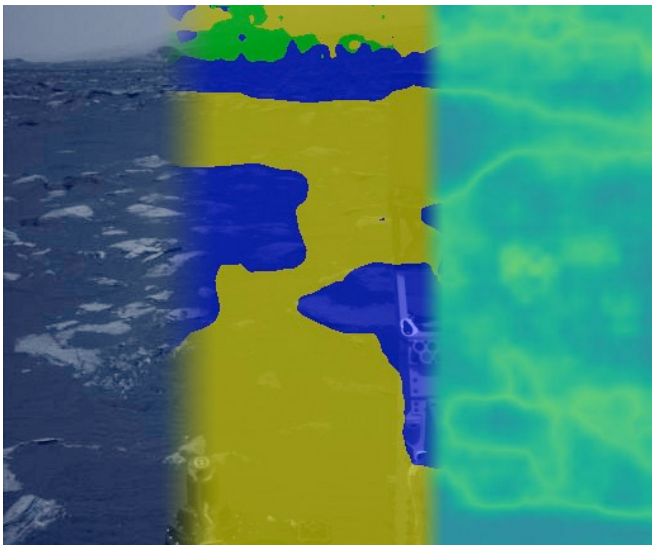

Figure 1. Example image of Mars. From left to right: raw sensor image, network prediction, and uncertainty prediciton.

Semantic information about the environment is crucial input for future planetary robotic missions. Of particular interest is the semantic information about the underlying terrain, which can support in the traversability analysis of the terrains and plan safe robotic paths, or discover scientifically interesting locations. Gathering information about the terrain from a distance is important for the safety of the robot. Therefore, classifying terrains with perception sensors is of high relevance. Based on the output of the perceptual terrain classification, a decision can be made on which terrain the robot can traverse or should avoid. The classification result can also help to improve the estimation of slip parameters.

Although deep neural networks have shown impressive results on the task of semantic segmentation, the task for terrain segmentation in the planetary context remains challenging. One reason is the lack of real-world data as the amount of collected data from extraterrestrial bodies is low. Even with the collected data, most of it is not annotated by any expert. Additionally, annotating the data based on predefined semantic classes is difficult in the unstructured environment in which they were collected. It is quite common to have label ambiguity in the data, meaning that it is not always clear if a pixel belongs to one class or another. This leads to

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

many images having large regions which are unlabeled. Last but not least, since space exploration is about making new discoveries, it is likely to be confronted with terrain classes that were not presented in the training dataset. These points and many more render terrain segmentation of the planetary environment difficult.

Since the safety of the robot is important for a mission, it is not enough to just predict whether a terrain is traversable or not. It is equally important to know how certain the method is about its prediction. If the robot predicts a terrain falsely as traversable, it can have problems traversing efficiently or might even get stuck, putting the entire mission at risk. For a robotic team, a wrong prediction can mean that team members are led to a wrong direction or to a point of interest that does not contain any scientifically relevant information.

Having an uncertainty map in addition to the classification result can provide the robot or scientist with an estimate on how reliable the predictions are, which can lead to better decisions. This work addresses the important topic of measuring uncertainty for terrain segmentation networks.

2. Related Work

In literature, one can find several works on terrain segmentation within the planetary context. A common approach is the pixel-wise classification by a Support Vector Machine (SVM). Shang and Barnes [3] apply a Fuzzy-rough feature selection followed by a SVM that assigns one out of nine terrain categories to each pixel. In [4], an Ant Colony Optimization based feature generation is used. Given the feature vector per pixel, a SVM is run with seven terrain classes.

While these methods operate in a two-stage manner, one can also apply end-to-end learning approaches. SPOC proposed by [5] is a terrain segmentation approach based on a fully-convolutional network. It is evaluated on data collected by a Mars rover and a Mars orbiter. Unfortunately, the data used by SPOC, as well as the network itself, is not open source, which makes it difficult to be compared to other methods. The authors of the AI4Mars dataset [6] apply a neural network on their collected dataset, which show a high performance on the labeled regions. The neural network in [7] uses a contrastive loss in order to learn the semantic classes. The authors train with data from AI4Mars and show that the amount of data needed for training is significantly less than the original one promoted in the dataset paper. To make use of the unlabeled regions, [8] proposes a self-learning approach. The authors also address the use of the prediction uncertainty in order to retrain the network. Although all of the above terrain segmentation approaches on Mars show promising results on the annotated regions, they do not show any results on the unlabeled regions. Since these areas make up a good portion of the image, it is important to also have some information about them for prediction purposes.

In literature, exhaustive research addresses the issues of estimating the prediction quality of a neural network, which is still an activate field of research. The direct calculation of the posterior of a prediction is intractable, which leads to various approximation approaches. One method to approximate the posterior distribution is to train the same network with different initial weights to form an ensemble[9]. During inference, the slightly differing predictions of all ensemble members due to other local optima being reached, are fused for the final prediction. Another form of creating an ensemble is to apply the Dropout technique. This technique, originally proposed in[10] as a regularization mechanism, can be used during inference to obtain an estimate of the prediction uncertainty[11]. An adapted version of the classical Dropout technique is proposed by [12], where the dropout hyper-parameter is directly learned. The advantages of this approach is its simplicity. However, it comes with the price that the inference procedure has to be done multiple times. Nevertheless, many works apply Dropout in the context of semantic segmentation [13][14]. In [14] the authors describe additional methods to derive uncertainty relevant for semantic segmentation, e.g. Laplace Approximation.

The term uncertainty can be further divided into model and data uncertainty. While the first represents the uncertainty of the model, which can, in principle, be decreased with more data, the other is expressing the uncertainty of the data itself. High uncertainty values can have various reasons. A major cause, especially in real-world scenarios, are samples that are out of the learned distribution during training. The detection of such cases is often referred to as novelty or anomaly detection, open-set prediction, or out-of-distribution detection. This ability is crucial for the task of outdoor terrain segmentation in unknown extraterrestrial environments. In this context, we are faced with semantic classes that were never seen before and thus, not part of our training data. As the ability of future space rovers to autonomously navigate is partially based on semantic information, the detection of unknown terrains is critical. Before a rover traverses novel terrain, it should communicate with operators on the ground. Furthermore, detecting areas with an unknown semantic class can also be interesting from a scientific point of view. One possible method to detect such samples is to introduce a separate network head, which predicts novel samples. The method shown in [15] uses a contrastive learning approach in order to detect anomalous samples. Hermann *et al.* [16] evaluate the final logits before the softmax layer to detect anomalies.

A major challenge for learning-based terrain segmentation methods in general is the limited amount of high-quality data. The amount of data is limited to the number of space missions executed in the past. For planetary environments, the focus was mainly on the planet Mars. However, despite multiple successful missions to the red planet over the past decades, it only represents a small fraction of celestial objects out there in the universe. Furthermore, most of the datasets normally lack required annotations due to enormous effort or ambiguous areas. One of the datasets collected by a Martian rover on Mars is AI4Mars [6]. Another dataset recorded by Curiosity is presented in [8]. While both datasets contain annotations, they differ in the labeled semantic classes as well as the used camera. Since collecting data from other planets and moons is not easily accesible, another option is to collect data from Earth in so-called analogue environments, which are similar to the ones found on other solar bodies [17]. The authors in [18] recorded data in Morocco, which features Mars-like environments. However, the data is not annotated and thus, cannot be directly used in a supervised learning scheme. Recently, several works address the labeled data bottleneck by training their networks only with synthetic data. Simulators based on modern graphic engines, such as [19], enable the generation of almost realistic data, which can be used for training neural networks. In [20], the blender based simulator OAISYS is presented. It synthesizes unstructured outdoor environments with a focus on planetary environments. Although simulators can help to overcome the data

shortage, they tend to introduce a new problem. The so-called sim-to-real gap defines the distribution difference between the simulated training and the real-world test data, which leads to low performance values. However, that simulators can be successfully used also for planetary applications is already shown among others in [20] and [21].

3. Datasets

For this paper, we make use of the AI4Mars dataset [6]. The dataset contains data from the Mars Science Laboratory (MSL) mission, more precisely the Mars rover Curiosity. The data was captured by the NAVCAM [22], a grayscale camera. The dataset contains of about 16k training and 943 test images. It distinguishes between four semantic classes: sand, soil, bed rock and big rock. All regions, which are do not bleong to these semantic classes, or where the human annotators were unsure, are labeled as unlabeled. Also, regions which are further away than 30 meters from the rover and the rover itself are annotated as unlabeled as well. Due to semantic label ambiguity, the dataset contains large amounts of areas, which are not labeled as a known class. Figure 2 shows example images from the dataset.

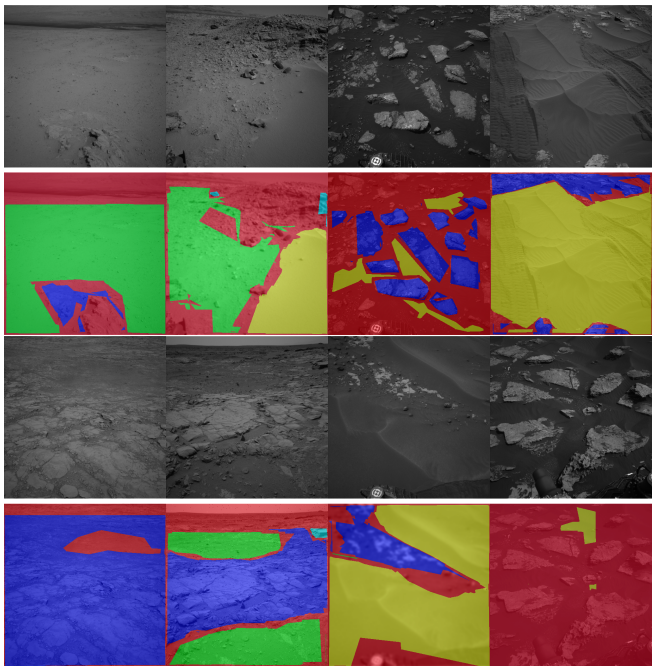

Figure 2. Example raw images and corresponding labels from AI4Mars dataset. The different colors represent the following semantic classes: green:soil; blue: bedrock; yellow: sand; cyan: big rock; red: unlabeled.

4. Terrain Segmentation with Uncertainty Prediction

In the following, we describe the three methods, which we use to obtain an uncertainty measure of our network. For all networks, we use the same DeepLabv3 backbone network [23], which only differs with respect to the specific technique used for obtaining the uncertainty measure. The DeepLabv3 network has proven to be capable for semantic segmentation tasks and belongs to one of the state of the art networks in its field.

Deterministic Approach

The most straight-forward way to obtain an uncertainty measure from the used network is to make use of the classification layer. Most networks will have a softmax layer at the very end, which contains the information about the prediction certainty. This method is often called deterministic approach, since it does not rely on any sampling. The advantage of this method is that the network does not have to be retrained in order to get this measure. The disadvantage, however, is that the outcome usually gives a rather poor estimate of the uncertainty prediction, which is typically surpassed by any other method which tries to approximate the posterior distribution. Nevertheless, due to its simplicity and the advantage that it is directly available for most networks, it is often used. Therefore, we include it in our evaluation as baseline. Figure 3 illustrates the schematic architecture of this method.

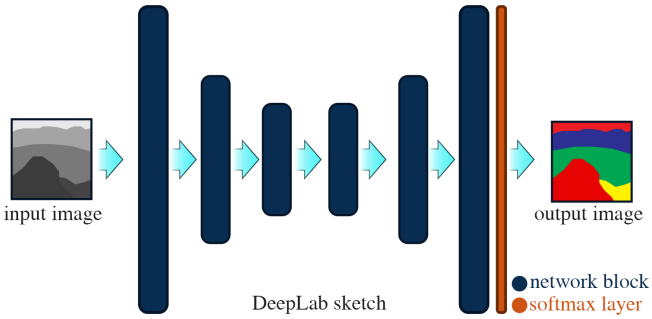

Figure 3. Illustration of the deterministic approach: The image is fed into the encoder stage of the segmentation network. The encoder is downsampling the image and pulling the semantic information from it, until it reaches the bottleneck (after the third layer in this example). After this, the semantic data is sent to the decoder part of the network, which is retaining the spatial information of the image. As the last layer, the network has a softmax layer (here in orange), which gives each pixel a class prediction and uncertainty value.

Entropy with Dropouts

Another option to get an uncertainty measure for the network prediction is the use of dropouts [24]. Dropouts activate and deactivate connections of a neural network randomly. With that, dropouts are an effective methods to help regularizing a neural network in the training phase. They force the neural network to distribute learning over the entire network, which makes it more robust and also helps it to generalize better. However, they have also been shown to be useful during inference for estimating the prediction uncertainty of a network. For estimating the uncertainty with dropouts, the dropouts are active in the training phase as well as in the inference phase. Instead of only once, the image sample is passed multiple times through the neural network to get multiple predictions. One can then calculate the predictive entropy as well as the mutual information in order to get a measure of the underlying uncertainty. Eq. (1) is used to calculate the predictive entropy according to [13].

$$\hat{H}[y|x, D_{\text{train}}]] = \\ -\sum_{c} \left(\frac{1}{T}\sum p(y=c|x,\hat{w}_t)\right) \log\left(\frac{1}{T}\sum p(y=c|x,\hat{w}_t)\right) \tag{1}$$

Eq. (2) represents the formula for the mutual information according to [13].

$$\hat{I}[y, w|x, D_{\text{train}}]] = \hat{H}[y|x, D_{\text{train}}]] + \sum_{c,t} \frac{1}{T} \sum p(y = c|x, \hat{w}_t) \log p(y = c|x, \hat{w}_t) \tag{2}$$

According to [13], predictive entropy represents the predictive uncertainty and the mutual information the epistemic uncertainty. The epistemic uncertainty is a measure for the model uncertainty, whereas the predictive uncertainty is a combination of epistemic and aleatoric uncertainty. The aleatoric uncertainty represents the data uncertainty. Figure 4 illustrates the schematic architecture of a semantic segmentation network with dropouts.

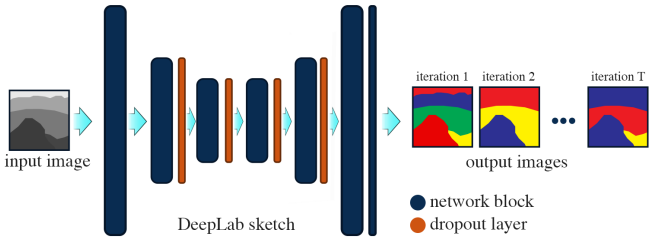

Figure 4. Uncertainty with dropouts: Instead of passing the sample only once trough the network as illustrated in Figure 3, the sample is passed multiple times. For each pass, the predictive output will be slightly different. The uncertainty of the prediction can be calculated using the predictive entropy and mutual information.

Contrastive Loss

When applying a contrastive loss, the semantic classes are viewed as clusters in an embedding space. The aim of the loss is to group samples as close as possible together, which belong to the same semantic class and repel samples, which are not. The method presented in [15] represents such an approach. Such a framework is particular suitable for detecting out-of-distribution samples. If a sample is far away from all other clusters, then it can be argued that it does not belong to any of the learned clusters and the sample can be assigned as out-of-distribution (ood) sample. The same principle can be used to measure how certain the network is about its prediction. The authors of [15] derived an uncertainty measure calculated from the embedding space, which we use for the task of terrain segmentation. However, we cannot use the proposed loss directly since it does not deal with unlabeled regions. The loss proposed by [15] comprises of two parts: discriminative cross entropy and variance loss. In order to use this loss, we have to mask out the areas which are unlabeled. Otherwise, the network would learn to directly group and move the unlabeled regions in the origin of the embbedding space. By doing so, the network would directly learn the unlabeled regions as another class. To prevent this, we mask the unlabeled areas and do not use them for the loss.

In order to get an uncertainty measure from this method, the authors of [15] propose an anomalous probability, which again comprises of two parts: metric-based maximum softmax probability and Euclidean distance sum.

Figure 5 illustrates the schematic architecture of a semantic segmentation network with contrastive loss.

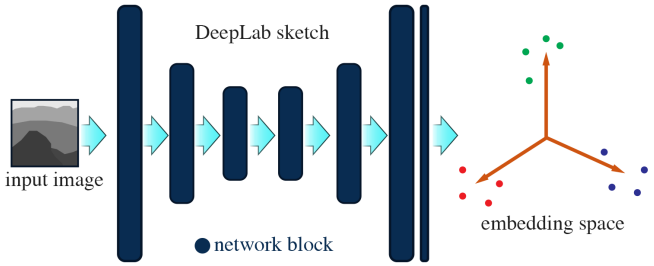

Figure 5. Uncertainty with semantic loss: The outcome of this network is a vector in an embedding space. Depending on the location of the sample in the embedding space, the sample either belongs to one of the known classes or is considered as an anomalous sample. Due to the loss, the anomalous samples will be more likely group around the origin of the embedding space. Note, that they should do so without having direclty learned the unlabeled regions.

5. Experiments

We first evaluate the accuracy of the different presented approaches. Here we are mainly concerned about the mean Intersection over Unit (IoU) per class. In another experiment we then compare the confidence prediction of the networks. Last but not least, we have a qualitative look at the out of distribution detection capability of the different approaches.

Prediction Accuracy

In this experiment, we evaluated how accurate the mentioned method can predict the known semantic classes. For that, we run the networks on the AI4Mars test data and apply the well known IoU metric. The IoU does not take regions, which are unlabeled, into consideration. Table 1 lists the class specific IoUs for each method. The overall IoU for 'deterministic', 'dropout', 'contrastive' are respectively: 0.80475, 0.81053, 0.77439. It can be seen that all approaches have a relative high overall IoU. As reported before by other researchers [6], the IoU for the class Big Rock is the weakest. This is due to an unbalancing in the dataset, which we tried to reduce with class weightings. However, it is also due to the fact that the dataset having many images, where big rocks are very similar to the class bedrock. Therefore, the networks sometimes predict bedrock instead of big rock. That might also be the reason why potential big rocks are often unlabeled in the training and test images, since the human labelers did not agree on the same semantic class. To increase the performance in such areas, the use of the depth map might be helpful as well as the prediction over multiple image frames. Interestingly, the overall IoU is the highest when applying dropouts. Since dropouts can also be used as regularization during training, it might have had a positive effect as well.

Table 1. IoU per semantic class

Model	Soil	Bedrock	Sand	Big Rock
Deterministic	0.9352	0.8093	0.8696	0.6047
Dropout	0.9404	0.9285	0.8910	0.4820
Contrastive	0.9365	0.8925	0.8624	0.4060

Uncertainty Prediction

The network is supposed to be accurate if it is certain about its prediction, whereas the network shall be uncertain in cases where its predictions are inaccurate. The PAvPU metric [13] is a metric to measure exactly this behavior.

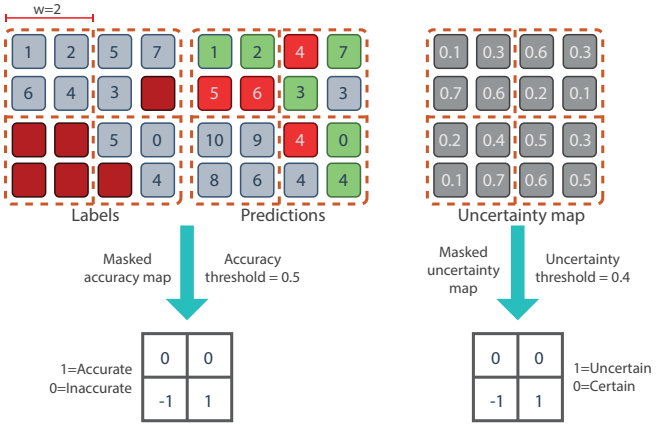

Figure 6. Example of the masked PAvPU metric. The matrices from left to right represent the ground truth labels, predictions and uncertainty map. The red pixels in the labeled map are unlabeled entries. The window size in the example is set to 2.

Since the PAvPU is not taking unlabeled regions in an image into account, we cannot directly apply it in our case. Therefore, we propose a masked version of the PAvPU metric. In each window, only the regions are taken into account, which have valid labels. If all regions in a window are unlabeled, the values for the accuracy map and uncertainty map are set to -1 and not further taken into account. Figure 6 illustrates an example of the masked PAvPU metric.

Figure 7 shows the PAvPU curve for the three evaluated methods. A window size of 2 was chosen.

It can be seen that the dropout approach provides the best performance. Both the dropout and contrastive method surpass the deterministic approach. The dropout method performs better than the contrastive approach. One of the reasons might be that the contrastive training and uncertainty map are more tailored towards recognizing anomalous samples, meaning samples that are not included in the training set.

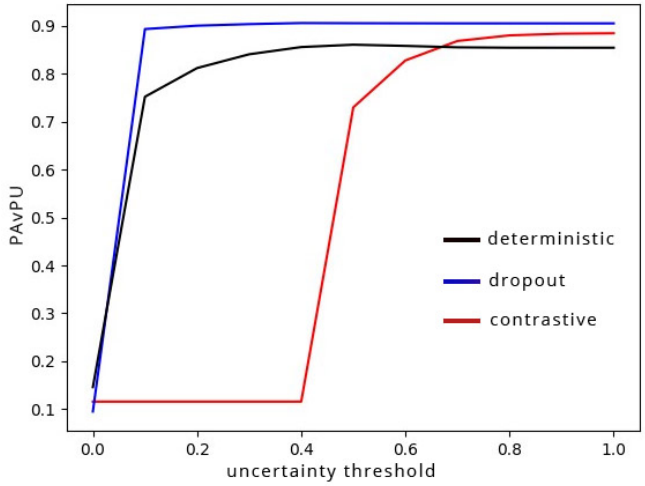

Figure 7. The PAvPU curve for the three methods.

Out-of-distribution Prediction

As mentioned before, neither the IoU, nor the PAvPU tells anything about the performance of the network in the unlabeled regions. However, the prediction of such regions are often most crucial in practice. They either represent areas, where even the human experts were not able to assign a class, or areas which are transition from one label to the other. These areas can therefore be very dangerous for the robot, since a network might predict the unlabeled regions as traversable, although it is not. Looking at the predictions in the unlabeled regions of the images, Figure 8, it becomes clear that the predictions of the network looks almost arbitrary. As a result, it is of great value if a network can detect areas, which are not part of the training dataset. This task is known as out-of-distribution task. In this experiment, we have a qualitative look at the out-of-distribution detection capabilities of the different approaches.

Prediction results with the corresponding semantic and uncertainty maps can be seen for 'deterministic', 'dropout', and 'contrastive' respectively in Figures 8, 9, and 10. As can be seen, all methods have usually a higher uncertainty in regions where they have false predictions and a low uncertainty if the prediction is correct. This shows that it is of advantage to calculate any of the uncertainty maps for an actual robotic mission. If a terrain region is unlabeled, the networks do their best to predict any of the known classes. In areas, which are out-of-distribution, like the robot class or the sky, the performance is quite different. For the sky class, the networks usually predict either sand or soil. This makes sense, since the homogeneous texture of the sky is mainly represented by such classes. However, it can be seen that the quality of uncertainty prediction for the sky is quite different. In the first image, the deterministic method is certain about its soil prediction. The same applies for the dropout method, which in general seems to be more responsive to the borders of two semantic classes instead of entire areas. Only the contrastive method shows high uncertainty values for the sky. Similar result can be seen for the unknown class of the robot in image 6. Both the deterministic and dropout method do not capture this out-of-distribution area well. Instead, the contrastive method is doing a much better prediction of the unknown area. From these examples it can be seen that the contrastive method is able to better detect out-of-distribution areas than the other two methods.

6. Conclusion

In the publication, we demonstrated the necessity of uncertainty estimation accompanied with the terrain segmentation prediction for space rovers. We showed in this publications several methods how to obtain such an uncertainty measure and compared the different methods with each other. In future work, it will be interesting to see how to further use the uncertainty maps for other tasks as well. One of such tasks might be to use it for pseudo labeles as some researchers have already proposed, or to use it for perception aware path planning and mapping.

Acknowledgments

This work was supported by the Helmholtz Association, project ARCHES (www.arches-projekt.de/en/, contract number ZT-0033).

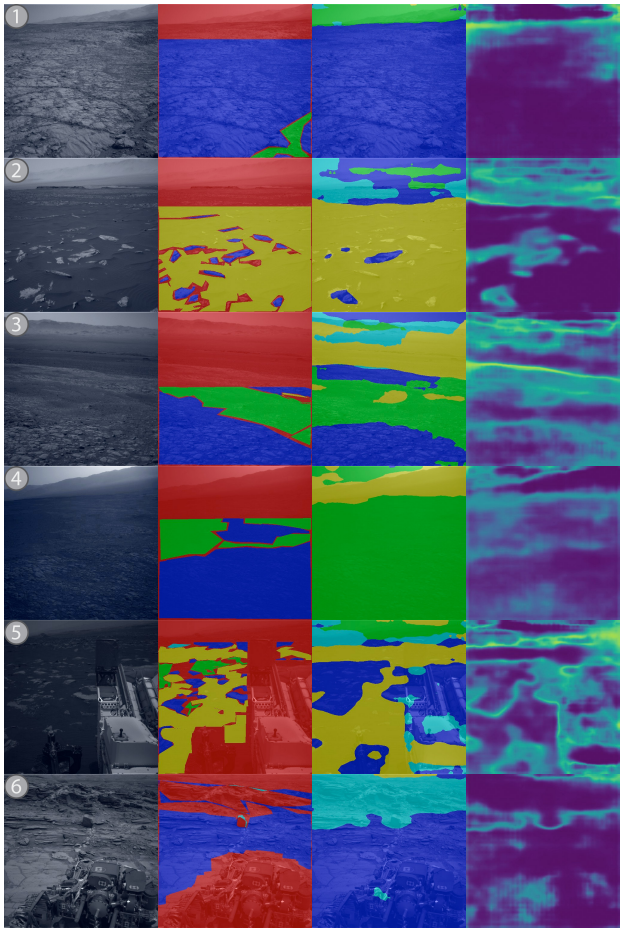

Figure 8. Qualitative results from the deterministic method. Left to right: raw image, ground truth, prediction, and uncertainty map. [color code: the brighter the more uncertain]

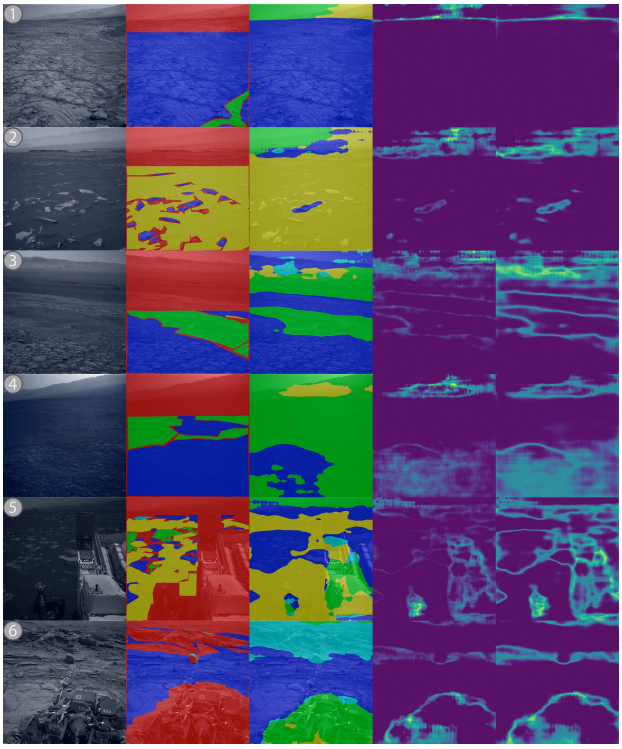

Figure 9. Qualitative Results from the dropout method. Left to right: raw image, ground truth, prediction, predictive entropy, and mutual information.

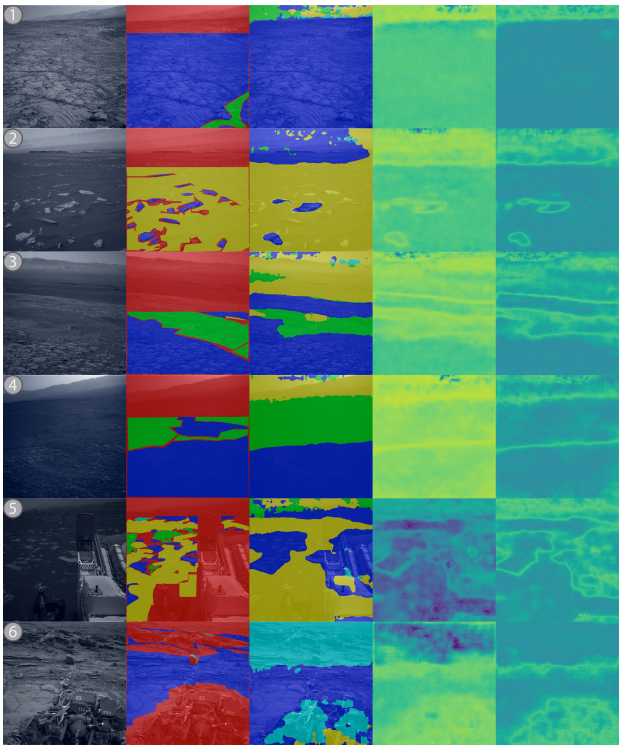

Figure 10. Qualitative Results from the contrastive method. Left to right: raw image, ground truth, prediction, metric-based maximum softmax probability, and Euclidean distance sum.

REFERENCES

[1] T. Tzanetos, M. Aung, J. Balaram, H. F. Grip, J. T. Karras, T. K. Canham, G. Kubiak, J. Anderson, G. Merewether, M. Starch, M. Pauken, S. Cappucci, M. Chase, M. Golombek, O. Toupet, M. C. Smart, S. Dawson, E. B. Ramirez, J. Lam, R. Stern, N. Chahat, J. Ravich, R. Hogg, B. Pipenberg, M. Keennon, and K. H. Williford, "Ingenuity mars helicopter: From technology demonstration to extraterrestrial scout," in *2022 IEEE Aerospace Conference (AERO)*, 2022, pp. 01–19.

[2] M. J. Schuster, M. G. Mller, S. G. Brunner, H. Lehner, P. Lehner, R. Sakagami, A. Dmel, L. Meyer, B. Vodermayer, R. Giubilato, M. Vayugundla, J. Reill, F. Steidle, I. von Bargen, K. Bussmann, R. Belder, P. Lutz, W. Strzl, M. Smek, M. Maier, S. Stoneman, A. F. Prince, B. Rebele, M. Durner, E. Staudinger, S. Zhang, R. Phlmann, E. Bischoff, C. Braun, S. Schrder, E. Dietz, S. Frohmann, A. Brner, H. Hbers, B. Foing, R. Triebel, A. O. Albu-Schffer, and A. Wedler, "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.

[3] 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.

[4] 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.

[5] B. Rothrock, R. Kennedy, C. Cunningham, J. Papon, M. Heverly, and M. Ono, "Spoc: Deep learning-based terrain classification for mars rover missions," *AIAA SPACE*, 2016.

[6] 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)*, 2021, pp. 1982–1991.

[7] E. Goh, J. Chen, and B. Wilson, "Mars terrain segmentation with less labels," 2022. [Online]. Available: https://arxiv.org/abs/2202.00791

[8] J. Zhang, L. Lin, Z. Fan, W. Wang, and J. Liu, "S^5mars: Self-supervised and semi-supervised learning for mars segmentation," 2022. [Online]. Available: https://arxiv.org/abs/2207.01200

[9] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," 2016. [Online]. Available: https://arxiv.org/abs/1612.01474

[10] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *Journal of Machine Learning Research*, vol. 15, no. 56, pp. 1929–1958, 2014. [Online]. Available: http://jmlr.org/papers/v15/srivastava14a.html

[11] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian Approximation: Insights and Applications," in *International Conference on Learning Representations (ICLR) Workshop*, 2016, p. 10.

[12] Y. Gal, J. Hron, and A. Kendall, "Concrete Dropout," *arXiv:1705.07832 [stat]*, May 2017, arXiv: 1705.07832. [Online]. Available: http://arxiv.org/abs/1705.07832

[13] J. Mukhoti and Y. Gal, "Evaluating bayesian deep learning methods for semantic segmentation," 2018. [Online]. Available: https://arxiv.org/abs/1811.12709

[14] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" 2017. [Online]. Available: https://arxiv.org/abs/1703.04977

[15] J. Cen, P. Yun, J. Cai, M. Y. Wang, and M. Liu, "Deep metric learning for open world semantic segmentation," *ArXiv*, vol. abs/2108.04562, 2021.

[16] H. Blum, M. G. Mller, A. Gawel, R. Siegwart, and C. Cadena, "Scim: Simultaneous clustering, inference, and mapping for open-world semantic scene understanding," 2022. [Online]. Available: https://arxiv.org/abs/2206.10670

[17] 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.

[18] L. Meyer, M. Smíšek, A. F. Villacampa, L. O. Maza, D. Medina, M. J. Schuster, F. Steidle, M. Vayugundla, M. G. Müller, B. Rebele, A. Wedler, and R. Triebel, "The MADMAX dataset for visual-inertial rover navigation on Mars," *Journal of Field Robotics*, 2021, in press.

[19] M. Sewtz, H. Lehner, Y. Fanger, J. Eberle, M. Wudenka, M. G. Mller, T. Bodenmller, and M. J. Schuster, "Ursim - a versatile robot simulator for extra-terrestrial exploration," in *2022 IEEE Aerospace Conference (AERO)*, 2022, pp. 1–14.

[20] M. G. Müller, M. Durner, A. Gawel, W. Stürzl, R. Triebel, and R. Siegwart, "A Photorealistic Terrain Simulation Pipeline for Unstructured Outdoor Environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2021.

[21] 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," in *Advances in Space Robotics and Back to Earth (IROS WS)*, Oktober 2021. [Online]. Available: https://elib.dlr.de/144626/

[22] J. Maki, D. Thiessen, A. Pourangi, P. Kobzeff, T. Litwin, L. Scherr, S. Elliott, A. Dingizian, and M. Maimone, "The mars science laboratory engineering cameras," *Space Science Reviews*, vol. 170, 09 2012.

[23] L. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *CoRR*, 2017.

[24] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," vol. 15, no. 1, p. 19291958, jan 2014.

BIOGRAPHY

Marcus G. Müller *is a researcher in the department of "Perception and Cognition" at the German Aerospace Center (DLR) since 2016 and Ph.D. student at ETH Zurich. He is the leader of the MAV Exploration Team at the Institute of Robotics and Mechatronics (DLR-RM), where he is working on autonomous navigation algorithms for MAVs. Before joining DLR, he conducted research at the Jet Propulsion Laboratory (JPL) of NASA in Pasadena, USA, where he worked on visual inertial navigation for MAVs and on radar signal processing. Marcus received his Master's and Bachelor's degree in Electrical Engineering from the University of Siegen, Germany.*

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.*

Hermann Blum *received his BSc and MSc in 2016 and 2018 from ETH Zrich, completing parts of these studies at Imperial College London. He is a PhD candidate at the Autonomous Systems Lab at ETH Zrich and recently joined the Computer Vision and Geometry Group of ETH Zrich. His research focuses on machine learning for semantic and geometric scene understanding, enabling manipulation and other safety critical robotic applications in proximity to humans.*

Abel Gawel *is currently a Principal Researcher in Computer Vision and Machine Learning with the Huawei Zurich Research Center. Before that he was a Senior Scientist at the Autonomous Systems Lab of ETH Zurich. He received the PhD from ETH Zurich in 2018 and was a visiting Postdoctoral Fellow in the CRI group at NTU Singapore in 2019. His research interests include SLAM, high-accuracy localization, object recognition, and semantic scene understanding with application in construction robotics, industrial inspection, and search and rescue robotics. Prior to joining ETH in 2014, he worked for Bosch Corporate Research and the BMW group.*

Wolfgang Stürzl *is a senior research scientist in the department of "Perception and Cognition" at the Institute of Robotics and Mechatronics of the German Aerospace Center (DLR). His research interests include computer vision for mobile robots, in particular using multi-camera and wide-angle imaging systems, and bio-inspired visual navigation of flying systems.*

Roland Siegwart *Roland Siegwart (1959) is full Professor of Autonomous Systems at ETH Zurich since July 2006 and Founding Co-Director of the Wyss Zurich. From January 2010 to December 2014, he took office as Vice President Research and Corporate Relations in the ETH Executive Board. He is member of the board of directors of various companies, including Komax and NZZ. He received his Diploma in Mechanical Engineering in 1983 and his Doctoral Degree in 1989 from ETH Zurich. He brought up a spin-off company, spent ten years as professor at EPFL Lausanne (1996-2006) and held visiting positions at Stanford University and NASA Ames.*

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.*