ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/378486871

# AI-Enabled Computer Vision Framework for Automated Knowledge Extraction in Planetary Rover Operations

**Conference Paper** · October 2023

CITATION
1

READS
117

**9 authors**, including:

Steven Kay
GMV NSL
**19** PUBLICATIONS **49** CITATIONS
SEE PROFILE

Maciej Quoos
GMV
**3** PUBLICATIONS **2** CITATIONS
SEE PROFILE

Robert Field
GMV
**3** PUBLICATIONS **2** CITATIONS
SEE PROFILE

Maciej Prokopczyk
GMV
**1** PUBLICATION **1** CITATION
SEE PROFILE

All content following this page was uploaded by Steven Kay on 26 February 2024.

The user has requested enhancement of the downloaded file.

# AI-ENABLED COMPUTER VISION FRAMEWORK FOR AUTOMATED KNOWLEDGE EXTRACTION IN PLANETARY ROVER OPERATIONS

**Steven Kay[1,*], Maciej Quoos[2], Robert Field[1], Maciej Józef Prokopczyk[2], Matteo De Benedetti[1], Fakher Mohammad[3,4], Maciej Szreter[2], Yang Gao[3,4], and Evridiki V. Ntagiou[5]**

[1]*GMV NSL Ltd, Airspeed 2, Eighth Street, Harwell Campus, Oxfordshire, UK, OX11 0RL*
[2]*GMV Innovating Solutions Sp. z o.o., ul. Hrubieszowska 2, 01-209 Warsaw, Poland*
[3]*STAR LAB, School of Mechanical Engineering Sciences, University of Surrey, Guildford, UK, GU2 7XH*
[4]*Centre for Robotics Research, Department of Engineering, King's College London, London, UK, WC2R 2LS*
[5]*European Space Operations Centre, ESA, Darmstadt, DE, 64293*
[*]*Corresponding author: Steven Kay, skay@gmv.com*

## ABSTRACT

This paper details the architectural design and implementation of a new AI-enabled Computer Vision (AI-CV) Framework developed in ViBEKO activity. The AI-CV Framework combines an AI platform and a Mission Control System (MCS). The AI platform selected was ESA's AInabler infrastructure based on the Kubeflow platform, and ONE-CC was selected as the MCS, which combines the 3DROCS rover control environment with the standard EG(O)S-CC mission control system. Also detailed are two software prototypes which implement robotic operations use cases. The first use case considers terrain classification of rover image products to increase situational awareness for rover operators when planning safe paths to traverse. The second use case implements rover global localisation to visually display to the rover operator the current best estimate of the rover position global coordinate system, using the most recent rover and orbital image and products. The AI-CV Framework has been developed to TRL 4 during the ViBEKO activity and the implemented use cases demonstrate the overall feasibility of the solution.

Key words: Mission Control System, AInabler, Planetary Robotics, Terrain Classification, Global Localisation.

## 1. INTRODUCTION

Advancements in Artificial Intelligence (AI) and Deep Learning (DL) stand out as transformational technologies in the digital age, especially when applied to Computer Vision (CV) tasks such as: object detection, semantic segmentation, pose estimation, and many others. Computer Vision is a well-established technique used in space applications for various tasks ranging from interplanetary navigation to studying the formation of planets. Recent trends show an increased desire to apply powerful AI-based Computer Vision techniques in the space domain, both within the ground segment, and the flight segment.

Image products from planetary rover missions provide an invaluable resource for both operators and science teams when planning activities. In rover operations, image products are used for numerous tasks, including both long-term strategic and short-term tactical path planning, hazard avoidance and target identification. The situational awareness afforded directly through the visualisation of raw or processed image products is essential to maintain rover safety during operations. Conventionally, processing of the received image products has been largely an intensive manual, or semi-automated process, requiring considerable resources.

Several systems have been proposed in the past which aim to provide ground control system tools for rover operators to perform short-term tactical and long-term strategic planning for rover activities. 3DROCS is a system developed in the framework of the 'Ground Control Station For Autonomy' ESA GSTP Activity. Its main objectives are to reduce the tactical planning process time, increase user awareness on the system behaviour, improve the Activity Plan understanding and provide a unified interface for strategic and tactical planning [1]. The ESA METERON (Multi-purpose End-To-End Robotic Operations Network) initiative aimed to demonstrate the feasibility of controlling advanced robots on Earth using 'telepresence' control equipment, providing essential experience for planning and preparing real human exploration missions. In this context, METERON can be seen as a test-bed for future missions to the Moon, Mars and other celestial bodies. The METERON activities are organised in dedicated experiments, which combine testing operations, ground and space systems, technologies, and robotic systems in an environment as realistic as possible [2]. More recently, the Robotic Digital Twin [3] activity proposed a new framework combining engineering tools and AI techniques to allow the on-line update of system models, facilitate planning, enable monitoring and fault analysis in the context of space missions, particularly for robotic assets. Lastly, the recent establishment of the AInabler Platform as a Service (PaaS), developed in the AI4Ops activity [4] has provided ESA users the

ability to train and deploy AI models for mission operations, hosted on ESA infrastructure.

Vision Based Knowledge Extraction using Artificial Intelligence (ViBEKO) is an ESA-funded activity focusing on the design and prototyping of an AI-enabled Computer Vision (AI-CV) Framework, with a view to automate the extraction of operationally relevant information from mission products to enhance situational awareness in typical rover operations scenarios, particularly in tactical, activity planning cycle scenarios. Specifically, the ViBEKO activity has resulted in the following main contributions:

- An extensible AI-CV Framework for hosting machine learning models was developed, based on the typical AI and Computer Vision processing chain, consisting of: preparing and pre-processing image and associated metadata products, model training, evaluation and versioning, and deploying selected model for later inference.
- The AI-CV Framework was established using existing ground systems tools and Mission Operation Infrastructure (MOI) through interface extensions to support image and AI product exchange and storage.
- Two AI-driven software prototypes were developed and validated using the established AI-CV Framework, utilising visual information from rover and orbital image products for selected rover operations use cases.

This paper focuses on the architectural design and implementation of the AI-CV Framework developed in ViBEKO. Additionally, this paper details concrete examples of AI-CV framework usage, by means of describing the design and implementation of two software prototypes selected to address common rover operation use cases, namely: planetary rover terrain classification and global localisation.

## 2. AI-ENABLED COMPUTER VISION FRAMEWORK FOR OPERATIONS

The Architecture concept of the AI-CV Framework is highlighted in Figure 1, which showcases the main system components taking part in the vision-based knowledge extraction workflow aimed at enhancing operators capabilities. The suggested workflow follows the classical approach utilised in the processing of spacecraft telemetry. Within this proposed architecture, the origin of the image products is the space robotic asset. It could be a planetary rover, robotic arm, spacecraft or any other device with capabilities to produce images. The asset sends products to the ground segment, routing to the Mission Control System (MCS). An AI Platform interacts with MCS in order to download these mission products and perform vision-based knowledge extraction with the use of AI techniques. The results of AI model inference is returned to the MCS and visualised to the operator, thanks to the robotic extensions available in the MCS. The AI platform is also used for model training and evaluation, by requesting large datasets hosted by the MCS. Both the MCS and AI Platform components play key roles in fulfilling the activity objectives briefly outlined before. As such, this have been the main focus of the ViBEKO activity, further detailed below. Integration of the other systems highlighted in the diagram could be considered in the future to increase the overall TRL of the solution.

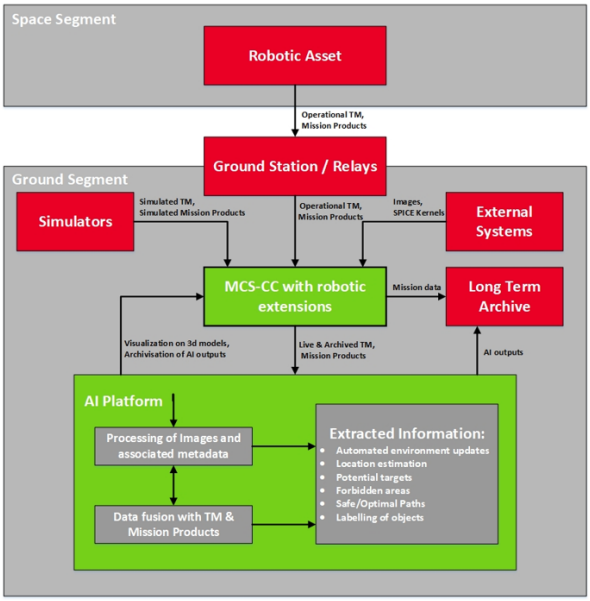


*Figure 1. High level perspective of the AI-CV Framework Architecture*

### 2.1. Mission Control System (MCS)

The Mission Control System (MCS) selected for the activity is ONE-CC, which was developed as part of the ONEOPSDS activity, where the key functionality of the 3DROCS control environment was included into the standard EG(O)S-CC mission control system. EG(O)S-CC is a new generation mission control system designed and developed by ESA, based on the EGS-CC solution providing a modular design for creating customised MCS instances for particular missions. The main part of the EGS-CC instance is the backend running components responsible for key functions, such as telemetry and telecommanding chains, access to data and file archives, and maintaining the mission definitions. Frontend clients can connect to the backend and provide users with a GUI, enabling respective visual capabilities. For ONE-CC, in addition to the standard EGS-CC views, the 3DROCS views were added, including 3D visualisations of the controlled assets, the planetary environment, captured image products, and engineering state of the asset subsystem.
The ONE-CC MCS required customisation to align with VIBEKO's specific requirements, namely: enabling the retrieval of image data from the MCS and integrating AI-generated results back into the MCS for operator utilisation. To achieve this, the creation of new extensions was essential. These extensions encompassed components developed within the scope of the VIBEKO project, such as facilitating access to the filespace and injecting telemetry through REST queries.

### 2.2. AI Platform

The AI platform selected for integration within the ViBEKO AI-CV Framework is ESA's AInabler instance of Kubeflow. Kubeflow, is an open-source machine learning platform designed for Kubernetes-based container orchestration, and is the back-end of ESA's AInabler Platform as a Service (PaaS) solution. Through the AI4Ops activity [4], ESA AInabler instances allow users to streamline their machine learning workflows by having access to ESA infrastructure and data silos to extract the most current data for model training.

The AInabler system is a mature and robust AI platform, but with limited underlying interfaces which handle communication to external systems. To facilitate communication with the MCS, new extensions in the form of a set of REST API interfaces were developed in ViBEKO which allow the transfer of data required for both model training and inference. The ViBEKO instance of AInabler is where the majority of infrastructure development efforts were made, in the form of pipelines: a series of steps written in Python which can be encapsulated and re-used as individual components. For example, the pre-processing step for resizing an image would be a single step in a larger pipeline that includes other processing steps in a directed acyclic graph (DAG). Pipelines were chosen for implementing the software prototypes for their reusability and representativeness, using a periodic running schedule any new data made available from the MCS. In doing so, operators can always have up-to-date predictions available to them on-demand. The ViBEKO pipeline was split into a series of tasks which follow typical Computer Vision and Machine Learning processing chains, namely:

- **Download**: checks the MCS for any new images or telemetry data, and downloads these products if available.
- **Pre-processing**: prepares the downloaded products to be used for later training or inference.
- **Model Training or Inference**: the AI-CV Framework mandates periodic inference, by default nightly. But should the user wish to train a new model, then a training step facilitates this, replacing the inference task. Model training is performed by requesting large datasets, which are provisioned through the MCS, and the corresponding hyperparameters and associated metadata from the training session are uploaded back to the MCS for later use.
- **Post-processing**: before uploading to the MCS, some models may require post-processing steps to format the prediction results before they can be visualised by the MCS frontend.
- **Upload**: uploads the results in the form of predictions from inference (visualisation results for the MCS to visualise), or training products (e.g., hyperparameters etc.,) to the MCS backend.

## 3. USE CASE OVERVIEW

### 3.1. Terrain Classification

Terrain classification is an essential function to ensure that a planetary rover can safely negotiate different terrains. Automating the identification of terrain types from rover images from ground segment workflows can both improve tactical planning efficiency and improve situational awareness for rovers operators. This is achieved by training models to learn different geological characteristics of the concerned planetary surface. Consequently, this use case aims to identify and build upon the best-performing deep learning models, and prototype these inside the AI-CV Framework, to perform semantic segmentation of planetary surfaces.

A number of recent planetary terrain classification methods have been proposed using a variety of models, using both orbital and surface image products. Terrain classification using orbital images is often employed in the analysis of planetary landing site candidates. The NOAH-H project [5] aimed to create a comprehensive set of ontological classes encompassing diverse surface textures in select areas of the surface of Mars. A deep learning-based terrain classification system was employed to categorise the various types of terrains, using the Google DeepLab model, the solution yielded an mIoU of 74.15%. Also, in [6], the authors benchmark different state-of-the-art semantic segmentation models for planetary safe site landing, including U-Net and DeepLabv3. Ultimately, ConvDeconv exhibited the best performance in terms of accuracy and had the shortest inference time. Additionally, it proved to be more computationally efficient and memory-friendly, attaining a pixel accuracy of 95% and an mIoU of 89%.

Using rover image products, [7] proposed a unique hybrid attention-based semantic segmentation approach for surface-based terrain classification, featuring a dual-branch network. This method effectively integrates both the broader global context and the finer local context of unstructured terrains. A merging module is employed to combine the contextual information from these two branches to produce the final segmentation through a newly designed loss function. The performance of this method is assessed on both a newly generated panoramic dataset called MarsScapes and the publicly available AI4Mars dataset. The method obtains a 60% mIoU on the MarsScapes dataset, and a 91% mIoU on the AI4Mars dataset. It's important to emphasize that the computational performance of this approach has not been verified. Additionally, the specific testing set configuration, performance metrics, and number of classes considered for training are not specified. A semantic segmentation network is introduced in [8] with a focus on limited labeled data of the Mars terrain. The approach employs semi-supervised learning, where an unsupervised model trained on an unlabeled dataset is adapted to a supervised network using a small amount of labeled data. Testing is conducted on the AI4Mars dataset, which includes four classes: soil, bedrock, sand, and big rock. To en-

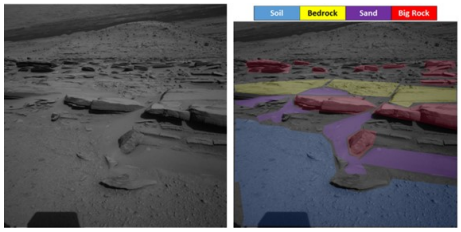


*Figure 2. AI4Mars dataset image labelling of soil, bedrock, sand and big rock [9].*

hance classification, two additional classes representing the rover itself and distances beyond 30 meters were incorporated, creating a six-class model. Notably, the latter two classes are excluded from testing and evaluation metrics. The results indicate an impressive pixel-level accuracy of 97.5% on the M3 testing set of the AI4Mars dataset, surpassing the plain supervised learning accuracy of 95% on the same dataset.

**Datasets**: NASA published the AI4Mars dataset [9], available at [10]. It consists of 35K images captured from MER and MSL-Curiosity rovers {Spirit navcam (3K), Opportunity navcam (6K), Curiosity navcam (17K) and Curiosity mastcam (9K)} and includes 326K full image labels. However, only a subset (MSL) of 16K images are readily usable for training semantic segmentation models. AI4Mars considers fewer and simpler labels in comparison to LabelMars [11]. The four labels included in the AI4Mars dataset are Soil, Bedrock, Sand, and Big Rock with different data proportions. The rarest class is Big Rock and the most common one is Bedrock. Besides the four classes; sky, distances further than 30 meters and the rover hardware are assigned a class label 255 and might be ignored during the training and testing process. Three holdout testing sets M1, M2, M3 each with 322 images are provided. The M3 set is considered the benchmark, hence recommendable for model testing.

The AI4Mars dataset, consisting of 16k Curiosity images, was divided into a 90% training set and a 10% validation set. Testing is conducted using the holdout sets M1, M2, and M3 provided, as previously described. For testing, areas of images containing rover hardware, pixels at distances beyond 30 meters, and unlabeled pixels are considered as a single class, resulting in a five-class segmentation model.

**Model Overview**: Following a review of the current state-of-the-art, models based on U-Net and DeepLabV3+ were selected for further evaluation and comparison. Method evaluation is done on the publicly available dataset AI4Mars[10]. Both models described below are based on encoder-decoder architectures.

Various pre-processing techniques were investigated to assess the potential to improve model performance. Furthermore, in the pursuit of additional performance improvement, especially for rare classes, various augmentation methods were employed, including the application of a GAN-based model, namely SemanticStyleGAN, to

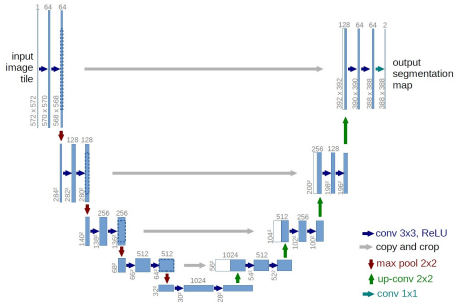


*Figure 3. U-Net architecture [12].*

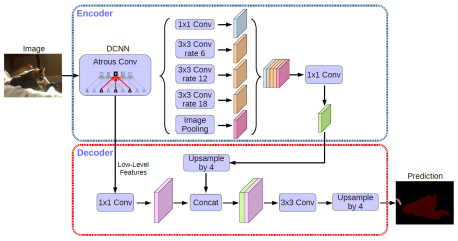


*Figure 4. DeepLabV3+ architecture [13].*

generate synthetic images.

**U-Net Architecture**: The U-Net is an encoder-decoder-based architecture originally used for medical/biomedical image segmentation [12]. The encoding and decoding paths give the model its U shape. To enable reusing learned features during down sampling, the feature maps in the down sampling path are concatenated with their mirrored counterparts among the up sampling path to catch various levels of abstractions (see grey arrows in Figure 3). Different CNN backbones, such as VGG, ResNet, Inception etc., can be employed in U-Net to evaluate encoding performance in different applications.

**DeepLabV3+ Architecture**: The DeepLabV3+ is the latest version from the popular DeepLab family, which includes DeepLabV1 - 3, and others [13]. The model is also based on an encoder-decoder architecture and uses atrous/dilated convolutions which increases the field of view without increasing the number of parameters. It uses filters at multiple sampling rates to capture objects and multiscale image contexts, and combines cascaded and parallel modules of dilated convolutions (see Figure 4). The encoder includes CNN backbone networks such as ResNet-101 or Xception, and the Atrous conv layers.

**SemanticStyleGAN**: The SemanticStyleGAN [14] is one of the most effective GAN models for generating images and their associated semantic segmentation masks. While commonly employed for generating and mixing facial features, its versatility extends to various other domains. Notably, this method not only generates new images but also produces corresponding semantic masks, making it a preferable method, avoiding additional labelling. Subsequently, this model is used to create synthetic images, thereby enlarging the dataset utilised, to


*Figure 5. SemanticStyleGAN generated images and masks.*

train the terrain classification models. To assess the quality of these generated images, standard metrics like Inception Score (IS) and Fréchet Inception Distance (FID) are used. Figure 5 shows an example of a generated image along with its corresponding mask.

**Data Processing**: Normalisation and centering processes were employed in all our experiments. Nevertheless, it's worth noting that standardisation led to a slightly diminished performance, and Contrast Limited Adaptive Histogram Equalization (CLAHE) did not yield any performance improvements in the limited experiments conducted and therefore dropped from the pre-processing pipeline. The volume of training data was increased by applying data augmentation. Straightforward augmentations introduced multiple variations of the original image, including flipping, cropping, rotating, random zooming, and random contrast adjustments. Additionally, advanced augmentation techniques using generative models such as SemanticStyleGAN are used to produce additional synthetic images.

**Training with GAN Models**: Two SemanticStyleGAN models, G1 and G2, are trained using different subsets from the AI4Mars dataset. G1 is trained on a subset consisting of 2,226 images, ensuring that the rare class "Big Rock" is consistently included. On the other hand, G2 is trained on the entire training set, which comprises a total of 14,457 images. Subsequently, each of these models generates 4,000 new synthetic Mars images. In two separate experiments, these synthetic images are added to the training set to assess their impact on overall performance and on the performance of the rare class.

**Results & Discussion**: Overall, DeepLabV3+ outperformed the U-Net model significantly in terms of both pixel accuracy and inference time, while achieving a similar mIoU on the testing sets (see Table 1). Consequently, DeepLabV3+ was selected for subsequent experiments. Training the models for 50 epochs yielded satisfactory results, but further improvements might be attainable with longer training durations. Basic augmentation techniques such as flipping, cropping, and zooming did not enhance performance, whereas employing advanced augmentation techniques based on GAN models improved the IoU on the testing sets.

The DeepLabV3+ model achieved a maximum segmentation accuracy of 95% on the AI4Mars testing set M3 (see Table 1). Additionally, augmenting the data with the GAN model G2 (see Table 2), trained on the entire dataset, increased the mIoU by approximately 3%, reaching a value of 61%. Notably, training the GAN model solely with images containing the rare class "Big Rock" improved the IoU for this specific class by 1-2%.

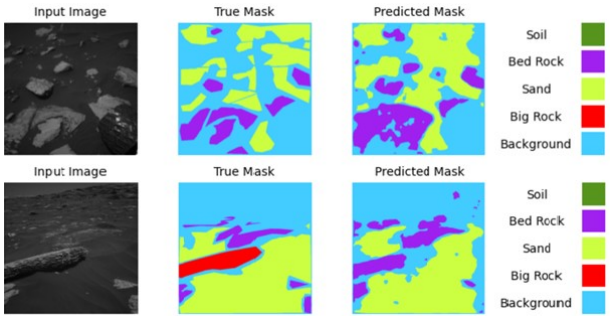


*Figure 6. DeepLabV3+ model predictions (Predicted Mask) and ground truth (True Mask) on AI4Mars.*

Using deeper backbone models like ResNet101 in DeepLabV3+ led to a modest enhancement in performance when contrasted with shallower models like ResNet50. However, it's crucial to weigh this marginal improvement against the increased complexity these deeper backbones bring to the model.

Figure 6 displays the mask predictions of the DeepLabV3+ model alongside the ground truth on the AI4Mars Dataset. In general, the model demonstrates impressive accuracy when predicting classes like Soil, Bed Rock, Sand, and Background. Nevertheless, its performance in predicting the Big Rock class falls short of expectations, frequently resulting in confusion with the Bed Rock class.

*Table 1. U-Net and DeepLabV3+ evaluation results; Acc,IoU, mIoU, inference time (s) on AI4Mars dataset.*

| Model | Backbone | Test set | Acc | Infer. time | mIoU | IoU | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Soil | Bedrock | Sand | Big Rock |
| U-Net | MobileNetV2 | M1 | 0.80 | 33 | 0.56 | 0.76 | 0.63 | 0.74 | 0.11 |
| | | M2 | 0.86 | 32 | 0.54 | 0.77 | 0.53 | 0.77 | 0.10 |
| | | M3 | 0.92 | 32 | 0.45 | 0.63 | 0.36 | 0.71 | 0.08 |
| U-Net | VGG16 | M1 | 0.82 | 77 | 0.57 | 0.77 | 0.62 | 0.75 | 0.12 |
| | | M2 | 0.88 | 78 | 0.55 | 0.78 | 0.52 | 0.79 | 0.12 |
| | | M3 | 0.92 | 78 | 0.45 | 0.64 | 0.35 | 0.71 | 0.10 |
| DeepLabv3+ | Resnet50 | M1 | 0.84 | **4** | 0.57 | 0.78 | 0.64 | 0.77 | 0.10 |
| | | M2 | 0.90 | **4** | 0.55 | 0.79 | 0.54 | 0.79 | 0.08 |
| | | M3 | 0.95 | **4** | 0.44 | 0.63 | 0.37 | 0.69 | 0.07 |
| DeepLabv3+ | Resnet101 | M1 | **0.85** | 10 | **0.58** | 0.78 | 0.66 | 0.77 | 0.09 |
| | | M2 | **0.91** | 10 | **0.55** | 0.79 | 0.54 | 0.80 | 0.08 |
| | | M3 | **0.95** | 10 | **0.45** | 0.64 | 0.35 | 0.72 | 0.07 |

### 3.2. Global Localisation

The ability to globally localise a planetary rover on the surface is critically important for every exploration mission. The objective of the global localisation function is to determine the position of the rover with respect to some global coordinate system. It is generally used to

*Table 2. DeepLabV3+ models result using GAN generated images. Backbone = ResNet50, input image size = 256x256.*

| Model | Trained on | Test set | Acc | mIoU | IoU | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | | Soil | Bedrock | Sand | Big Rock |
| DeepLabv3+ | AI4Mars | M1 | 0.84 | 0.57 | 0.78 | 0.64 | 0.77 | 0.10 |
| | | M2 | 0.90 | 0.55 | 0.79 | 0.54 | 0.79 | 0.08 |
| | | M3 | 0.95 | 0.44 | 0.63 | 0.37 | 0.69 | 0.07 |
| DeepLabv3+ | AI4Mars+G1 | M1 | 0.86 | 0.58 | 0.78 | 0.66 | 0.78 | **0.11** |
| | | M2 | 0.91 | 0.55 | 0.78 | 0.54 | 0.79 | **0.1** |
| | | M3 | 0.95 | 0.44 | 0.63 | 0.35 | 0.7 | **0.08** |
| DeepLabv3+ | AI4Mars+G2 | M1 | 0.86 | **0.61** | 0.78 | 0.66 | 0.77 | 0.09 |
| | | M2 | 0.91 | **0.58** | 0.79 | 0.54 | 0.80 | 0.06 |
| | | M3 | 0.95 | **0.47** | 0.62 | 0.35 | 0.71 | 0.05 |

correct for the accumulation of relative localisation error between rover traverse activities.

Several methods have been used over the years to fulfil the task of rover global localisation, based on different techniques and various types of sensors. On-board celestial navigation methods, based on sun sensing [15] and star tracking [16] have been proposed. However, their performances are highly dependent on INS sensor resolutions as well as precise calibration and initialisation to be effective, often resulting in positional errors in the order of 100m and subject to sensor drifts.

Numerous strategies have been proposed using vision-based methods using conventional CV techniques. From the ground segment, Bundle Adjustment was successfully performed using descent imagery and collected rover images to construct an image network with manually defined tie-points. From on-board, Skyline-based methods [17] [18] estimate the rover's absolute position through comparing the observed skyline with a set of "simulated" skyline candidates rendered from DEMs. The matching of local and globally traceable characteristics, such as terrain topography directly [19] [20], or by matching specific salient visible features [21] [22] have also been proposed.

Recently, a number of Deep Learning approaches to map-based feature matching for global localisation have been developed. These have shown better generalisation properties in varied conditions, when compared with conventional CV techniques. Approaches include the successful demonstration of Siamese Neural Networks (SNNs) [23], with Generative Adversarial Networks (GANs) used to increase the dataset size for training. A CNN discriminator is trained to score the pose likelihood between a position rendered DEM image with a rover image to score the matching probability based on learned deep correspondences. Another model based on a Siamese Neural Network (SNN) model, called PLaNNet [24], learns deep associations between visible salient features of reprojected panoramic rover images and corresponding regional orbital image tiles, from which the rover position can be determined as the highest scoring region correspondence.

**Datasets**: For ViBEKO, the main interest was to develop a prototype in the AI-CV Framework, based on a Deep Learning approach that could run well on real image data. Therefore, three different datasets were selected: a synthetic dataset with a large number of images for training, and two real world datasets with more realistic and varied data that better represent an actual mission. A medium sized synthetic dataset was generated based on the work from [24]. This data was then pre-processed to generate the ground reprojection from four directional rover images, as shown in Figure 7 below. Additional filtering is then performed to remove images that happen to be captured in very dark areas that would lead to unusable reprojections. Each reprojection is then coupled with its corresponding regional orbital image tile.

To provide additional and more realistic samples for training, another real world dataset was explored, the ENAV dataset [25]. This dataset was collected at the Canadian Space Agency's Mars Emulation Terrain (MET) in Saint-Hubert, Quebec, Canada. The platform was fitted with eight cameras on a mast, a monocular camera, IMU, and ground truth data. It also provides an aerial DEM with a px resolution of 0.2px/m which is crucial to be able to associate rover and orbital views. With this new real-world data, more advanced filtering was required. This is because, as the sun would often overexpose some of the frames. A brightness balancing filter based on CLAHE was used to mitigate this effect and avoid having the sun-exposed side of the reprojection too dark.

The ESA Tenerife dataset (also known as Planetary Robotics Lab (PRL) Martian Terrain Dataset) is a collection of traverses at various lengths, totalling 13 km, in representative planetary landscape [26]. It contains multiple stereo camera streams for LocCam, HazCam and PanCam, and data from ToF camera, IMU, wheel and PTU odometry and LiDAR scans, Georeferenced DEMs and GPS ground truth. Given the interesting location and overall realism of the dataset, it was fully reserved for the final demonstration, while all the training and validation was performed on a combination of synthetic and ENAV data.

**Model Overview**: The chosen architecture was based on a Siamese Neural Network (SNN), as discussed above, with a ResNet backbone proposed by [24], and extended in a few ways:

- More datasets were considered to improve the generalisation and performances of the model, especially using real-world data.
- An evaluation of various backbones was performed, based on commonly used CNNs such as DenseNet, VGG, and ResNets of various sizes, and a few others.
- Various pre-processing steps were performed, customised for each dataset, in particular the two real-world datasets.
- Investigation and tuning of the training process and parameters, such as the depth of the retraining and

fine-tuning.

The two backbone CNNs receive the rover's reprojection and an orbital image time as input, both covering a 50x50m area at 128px resolution. The two CNNs generate an embedding of the two images which are then fed into the comparison layers, consisting of two fully connected layers that transform the 256 values of the embedding into a single similarity score between 0 and 1.
A brute-force approach to matching was performed, where given a single rover reprojection, an exhaustive search of the full orbital image space is performed, generating an orbital tile the same size as the reprojection at different locations, in order to find the best matches.

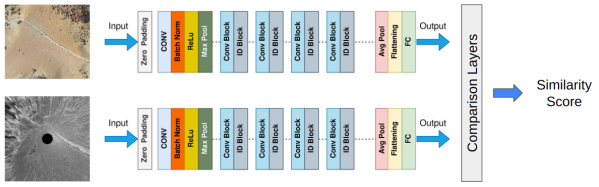


*Figure 7. Architecture of the Global Localisation Pipeline based on Siamese Networks.*

**Data Processing**: To further extend the dataset, data augmentation was performed. Typical augmentations, such as scaling, cropping, zooming, or flipping were not possible due to them breaking the spatial correspondence between the reprojections and orbital images. Ultimately, only the orientation could be altered. This method also made sense from an operational standpoint: in the problem of global localisation, it should not be assumed there will be a good absolute heading estimate to produce a reprojection aligned to the orbital images, which was the case in the synthetic dataset.

Lastly, a final post-processing phase is needed to generate a useful result with a position estimate, as the SNN only estimates the similarity between two images. The ENAV and PRL datasets have an orbital image containing both colour layers and a geotag layer that facilitates the mapping of pixels to GNSS coordinates. Therefore, a transformation between pixel and GNSS coordinates could be established, based on the highest location similarity identified by the SNN. The top N=5 matches were taken as location candidates and their pixel coordinates transformed to GNSS coordinates. The final results for the PRL-Tenerife dataset are shown in Figure 8.

**Results & Discussion**: A few different metrics were used to evaluate the model performances. The objective was to cover both traditional Deep Learning accuracy, and also capture the operational performances of the scenario of global localisation accuracy. Using an estimation score, the minimum distance between the N=5 best estimates and the GPS groundtruth are used and averaged across the batch. To maintain a consistent evaluation, the same batch of the testing dataset was used across all experiments.

Direct comparison with similar applications is not

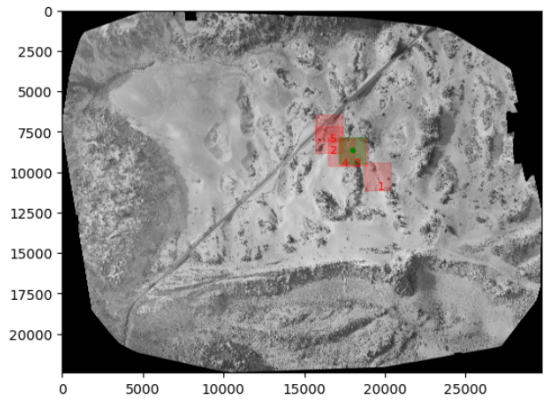


*Figure 8. Results of the Global Localisation Pipeline visualised on the orbital image of the PRL Tenerife dataset.*

straightforward, as this approach is not extensively used in literature. [24] showed similar performances as that proposed here, but while it used a similar architecture, it was on a purely synthetic dataset. In this project, the main interest was to be able to generalize to new and unseen environments coming from real data. [27] also uses a similar SNN, comparing the orbital image directly with the monocular frames, instead of the ground reprojection, as part of a bigger architecture with a Particle Filter and VO pose estimation. They were able to obtain much higher accuracy thanks to the VO refinement, large dataset and the progressive computation, whilst our approach was one-shot on a single sample of data and used a limited size dataset. For a direct comparison, our method used a similar approach as [24] by comparing its performances against fully random sampling, Sum of Absolute Differences (SAD), and Sum of Squared Differences (SSD). The results are reported in Table 3 and show a substantial improvement from random, SAD, and SSD thanks to the SNN architecture.

*Table 3. Evaluation of the Global Localisation Pipeline on the Tenerife Dataset.*

| SNN Evaluation with Random Sampling, SSD, and SAD, Position Score<br>Backbone: ResNet50 until last classifier. Epochs: 50. Steps: 500.<br>Dataset: PRL | | | |
|---|---|---|---|
| Random Sampling | Sum of Squared Differences (SSD) | Sum of Absolute Differences (SAD) | Position Estimation Score |
| 186.66956747 | 149.69381958 | 114.90633132 | 58.6318445 |

## 4. CONCLUSIONS

The ViBEKO activity has successfully demonstrated the potential for AI-CV pipelines to perform complex knowledge extraction tasks in the context of rover operations following an automated paradigm. In the activity, a new AI-CV Framework has been proposed, which builds upon existing ground systems infrastructure, namely the ESA AInabler platform and ONE-CC (integrated 3DROCS and EG(O)S-CC). Two representative rover operations

use cases; planetary terrain classification and global localisation, were implemented within the AI-CV Framework, validating the AI-CV concept. Overall, both software prototypes achieved impressive results in their own right.

The developments achieved within ViBEKO, related with the AI-CV Framework as part of the MCS also has applicability in relation to spacecraft operations. Similar use cases to those identified in ViBEKO can be foreseen where data products may be automatically analysed using trained models, deployed within the integrated AI Platform and MCS solutions, switching out the frontend for a more suitable visualisation tool. This would be applicable for both deep space missions (e.g., JUICE) and also missions targeting LEO (e.g., IOSM/ADR style missions, or Earth Observation platforms).

## ACKNOWLEDGEMENTS

This work was funded by the European Space Agency under Contract No 4000137729/22/NL/AT "Vision Based Knowledge Extraction using Artificial Intelligence". The activity was coordinated by GMV Innovating Solutions SP. Z O. O. and includes GMV NSL Limited (GMV-UK) and the University of Surrey.

## REFERENCES

[1] L. Joudrier, K. Kapellos, and K. Wormnes. 3d based rover operations control system. In *ASTRA*, 2013.

[2] M. Cardone et al. The meteron operations environment and robotic services, a plug-and-play system infrastructure for robotic experiments. 05 2016.

[3] E.V. Ntagiou et al. Towards an ai-enhanced robotic digital twin for space exploration assets. In *SpaceOps*, 03 2023.

[4] S. Beltrami et al. Enabling ai applications for space operations through a multi-mission devops platform. *SpaceOps*, 2023.

[5] A. M. Barrett et al. Noah-h, a deep-learning, terrain classification system for mars: Results for the exomars rover candidate landing sites. *Icarus*, 371:114701, 2022.

[6] T. Claudet, K. Tomita, and K. Ho. Benchmark analysis of semantic segmentation algorithms for safe planetary landing site selection. *IEEE Access*, 10:41766–41775, 2022.

[7] H. Liu et al. A hybrid attention semantic segmentation network for unstructured terrain on mars. *Acta Astronautica*, 204:492–499, 2023.

[8] E. Goh et al. Mars terrain segmentation with less labels. In *IEEE Aerospace Conference*, pages 1–10. IEEE, 2022.

[9] R.M. Swan et al. Ai4mars: A dataset for terrain-aware autonomous driving on mars. *IEEE/CVF CVPR Workshops*, page 1982–1991, 2021.

[10] Ai4mars dataset. https://data.nasa.gov/Space-Science/AI4MARS-A-Dataset-for-Terrain-Aware-Autonomous-Dri/cykx-2qix.

[11] S.P. Schwenzer et al. Labelmars: Creating an extremely large martian image dataset through machine learning. *LPSC*, (2132):1970, 2019.

[12] O. Ronneberger et al. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer, 2015.

[13] L-C. Chen et al. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *ECCV*, pages 801–818, 2018.

[14] Y. Shi et al. Semanticstylegan: Learning compositional generative priors for controllable image synthesis and editing. In *IEEE/CVF CVPR*, pages 11254–11264, 2022.

[15] R. Volpe. Mars rover navigation results using sun sensor heading determination. In *IEEE/RSJ IROS*, 1999.

[16] H. Zhu, X.L. Wang, and J.C. Fang. An innovative highprecision sins/cns deep integrated navigation scheme for the mars rover. In *Aerosp. Sci. Technol.*, 2011.

[17] P. Furgale, P. Carle, and T.D. Barfoot. A comparison of global localization algorithms for planetary exploration. In *IEEE/RSJ IROS*, page 4964–69, 2010.

[18] F. Cozman et al. Outdoor visual position estimation for planetary rovers. In *Autonomous Robots*, 2000.

[19] B. Van Pham, A. Maligo, and S. Lacroix. Absolute map-based localization for a planetary rovers. In *ASTRA*, 2003.

[20] D. Geromichalos et al. Slam for autonomous planetary rovers with global localization. In *Journal of Field Robotics*, page 830–47, 2020.

[21] J.W. Hwangbo et al. Integration of orbital and ground image networks for the automation of rover localization. In *ASPRS*, 2009.

[22] E. Boukas et al. Introducing a globally con-sistent orbital-based localization system. In *Journal of Field Robotics*, page 275–98, 2018.

[23] A. Naguib et al. Planetary long-range deep 2d global localization using generative adversarial network. In *Journal of Korea Robotics Society*, page 26–30, 2018.

[24] A. Chung et al. Localization: Merging orbital maps with surface perspective imagery. In *NASA Frontier Development Lab*, 2018.

[25] O. Lamarre et al. The canadian planetary emulation terrain energy-aware rover navigation dataset. In *The International Journal of Robotics Research*, 2006.

[26] M. Azkarate et al. Heavy duty planetary rover tenerife field test. In *ASTRA*, 2017.

[27] V. Franchi and E. Ntagiou. Planetary rover localisation via surface and orbital image matching. In *IEEE Aerospace Conference (AERO)*, pages 1–14, 2022.

View publication stats