Machine Vision and Applications (2024) 35:50
https://doi.org/10.1007/s00138-024-01533-3

RESEARCH

Lunar ground segmentation using a modified U-net neural network

Georgios Petrakis[1] · **Panagiotis Partsinevelos**[1]

Received: 17 September 2023 / Revised: 20 February 2024 / Accepted: 15 March 2024

Abstract
Semantic segmentation plays a significant role in unstructured and planetary scene understanding, offering to a robotic system or a planetary rover valuable knowledge about its surroundings. Several studies investigate rover-based scene recognition planetary-like environments but there is a lack of a semantic segmentation architecture, focused on computing systems with low resources and tested on the lunar surface. In this study, a lightweight encoder-decoder neural network (NN) architecture is proposed for rover-based ground segmentation on the lunar surface. The proposed architecture is composed by a modified MobilenetV2 as encoder and a lightweight U-net decoder while the training and evaluation process were conducted using a publicly available synthetic dataset with lunar landscape images. The proposed model provides robust segmentation results, allowing the lunar scene understanding focused on rocks and boulders. It achieves similar accuracy, compared with original U-net and U-net-based architectures which are 110–140 times larger than the proposed architecture. This study, aims to contribute in lunar landscape segmentation utilizing deep learning techniques, while it proves a great potential in autonomous lunar navigation ensuring a safer and smoother navigation on the moon. To the best of our knowledge, this is the first study which propose a lightweight semantic segmentation architecture for the lunar surface, aiming to reinforce the autonomous rover navigation.

Keywords Planetary environments · Deep learning · Semantic segmentation · Rover navigation

1 Introduction

Semantic segmentation plays a significant role in unstructured and planetary scene understanding, offering invaluable knowledge to a robotic system or a planetary rover about its surroundings [1]. Through terrain semantic segmentation, robotic systems are able to analyze images or videos and accurately detect and classify multiple features or regions within their environments, allowing superior comprehension and spatial awareness. More specifically, robotic systems are capable of identifying and differentiate various elements including boulders, craters, or even potential obstacles and hazards. This fact allows the use of semantic information in the path planning, enabling the robotic system to navigate in challenging landscapes with increased safety. Moreover, accurate semantic segmentation is able to recognize potential mineral deposits or geological formations, contributing to scientific research for planet exploration.

Several studies investigate semantic segmentation in unstructured and planetary scenes using traditional algorithms without learning-based processes including [2–5], and machine learning algorithms such as [6–9]. However the last five years, terrain semantic segmentation based on deep neural networks dominates the literature [10].

Regarding the earthy unstructured scenes, in [11] and [14] authors propose semantic segmentation methodologies based on a modified DeepLabV3 + [12, 13] and a U-net with EfficientNet [15] backbone respectively, aiming to improve the scene understanding of self-driving vehicles in unstructured environments. Both models were trained and evaluated with IDD (Indian Driving dataset) dataset due to its high diversity, achieving satisfactory results using mean IoU (mean Intersection over Union) metric. In [16], authors propose a lightweight neural network for terrain semantic segmentation focused on unstructured environments which is capable of merging multi-scale visual features, in order to efficiently group and classify different types of terrains while a

✉ Georgios Petrakis
gpetrakis2@tuc.gr

Panagiotis Partsinevelos
ppartsinevelos@tuc.gr

[1] Technical University of Crete, Chania, Greece

Published online: 09 April 2024

reinforcement learning algorithm, is able to utilize the predicted segmentation maps aiming to plan and guide a robot in paths with high safety. Similarly, in [17], a real-time terrain mapping method for autonomous excavators is presented, which is able to provide semantic and geometric information for the terrain using RGB images and 3D point cloud data, while a dataset which includes images from construction sites is designed and utilized. Regarding the datasets for earthy unstructured environments, in [18, 19], two publicly available datasets were developed for semantic segmentation deep learning models, focusing on self-driving in semi-unstructured or dense-vegetated environments. In [18], the dataset designed, for accurate comprehension in scenes with high coverage in grass, asphalt, soil and sand, while authors in [19], targeted more on dense-vegetated and rough terrain scenes for off-road self driving scenarios.

Concerning the planetary environments, several methodologies have been proposed for feature detection and terrain or scene segmentation aiming to reinforce and improve planet exploration tasks including landing, rover-based path planning, localization or planet surface investigation. In [20], a modified-U-net architecture [21] for rock segmentation on the martian surface is proposed, which was trained and tested with a Mars-like dataset [22] captured on Devon Island, achieving satisfactory accuracy. In [23], authors conduct a performance evaluation in rock detection for Mars-like environments using an original and modified versions of SSD (Single-Shot-Detector) [24] neural network, trained with the aforementioned dataset [22]. In [25], a modified Unet + + architecture [26] for rock segmentation in planetary-like environments is proposed where two rounds of training are performed for the learning process. In the pre-training stage, the proposed architecture is fed by a synthetic dataset, created by a proposed algorithm while in the fine-tuning stage, the architecture is trained using a limited part of the Katwijk beach planetary rover dataset [27]. In [28, 29], authors conduct a benchmark analysis in Hazard Detection (HD) for planetary landing using several state-of-the-art semantic segmentation models compared with a replicated HD algorithm from NASA's Autonomous Landing Hazard Avoidance Technology (ALHAT) project. The results proved that the segmentation architectures provide high efficiency on hazard detection outperforming the ALHAT algorithm in performance time and accuracy.

Several studies investigate the sky and ground segmentation in planetary environments, aiming to refine the scene understanding [30, 31]. In [30], an architecture for sky and ground segmentation in planetary scenes is proposed, inspired by U-net and NiN (Network In Network) [32] which was trained for two rounds with SkyFinder [33] and Katwijk beach planetery rover datasets [27] respectively. On the other hand in [31], a DeeplabV3 + neural network is utilized for skyline contour identification in martian environment, aiming to estimate the rover's global position.

A significant limitation of deep learning methods in planetary environments, is the lack of qualitative real or synthetic available datasets, compared with datasets for urban or indoor environments [34]. In [34], authors propose a simulator which is able to construct valuable synthetic scenes for planetary environments including rich metadata while furthermore it is capable of generating multi-level semantic labels based on pre-defined materials. On the other hand, in [35], authors propose a large-scale dataset called AI4MARS for terrain semantic segmentation of Mars, aiming to reinforce autonomous navigation on the martian surface. AI4Mars includes about 35K annotated images captured by Curiosity, Opportunity and Spirit rovers while the labeling conducted by experts with the aid of crowdsourcing using a web-based annotation tool.

A crucial use of terrain classification in planetary environments is the path planning optimization [36]. In [37], a terrain segmentation model is proposed using PSPNet [38], trained by real rover-based images from Mars and artificial images generated by the Unity3D software, aiming to automate a path planning algorithm on the Martian surface. In [39], authors propose a methodology for path rerouting using imagery data, depth maps and a CNN-based neural network trained with Katwijk beach planetery rover dataset, in order to detect and avoid obstacles such as rocks and boulders.

Although several studies investigate rover-based scene recognition in Martian surface or planetary environments in general, quite few investigate similar tasks for the lunar surface. Lunar topography includes several features including rocks, boulders and craters, while the terrain in many areas is quite uneven with mounds and valleys. Although several studies propose methodologies for crater [40–42] or hazard [43] detection and segmentation, they focus on safe landing using remote-sensing images while there is a deficiency in rock and boulder identification during the rover navigation; a quite important issue for the smooth and trouble-free navigation.

In this study, a lightweight encoder-decoder neural network (NN) architecture is proposed for rover-based ground segmentation on the lunar surface. The proposed architecture is based on U-net and MobilenetV2 [44] while the training and evaluation process were conducted using a synthetic dataset with lunar landscape images. The proposed model provides robust results, allowing the lunar scene understanding focused on rocks and boulders. The main contributions of the study can be described as follows:

- Development of a lightweight semantic segmentation model aiming to reinforce the autonomous rover navigation on the lunar surface

- Investigation of lunar scene understanding through deep learning, using synthetic dataset in training and a combination of real and synthetic datasets in evaluation
- Comparison of the model with U-net-based alternatives in different computing setups, proving the superiority of the proposed architecture in terms of accuracy and performance-time
- Lunar scene understanding based on semantic-segmentation through deep learning, proves a great potential in autonomous lunar navigation ensuring a safer and smoother navigation on the moon

2 Materials and methods

Semantic information in unstructured environments provides a contextual understanding of objects and their relationships within an image, enabling machines to recognize and categorize features semantically, reinforcing crucial tasks including autonomous navigation in unknown planetary scenes. Although the literature includes several studies focused on terrain segmentation in unstructured scenes, there are two main gaps, that this study attempts to fill:

- Semantic segmentation in the lunar surface using rover-based images, instead of the most studies that investigate scene understanding through semantic segmentation in earthy unstructured environments or in the martian surface
- A lightweight semantic segmentation model, capable of being used in systems with low computing resources, providing high efficiency after training with a limited size of dataset

In other words, the scope of this study, is to develop a lightweight and robust semantic segmentation model, aiming to be used in lunar surface exploration. Two challenges have to be encountered: The first one is the lack of valuable rover-based datasets for the lunar surface, compared with Mars where several datasets have been proposed. The second challenge is the size of the model, since most of the semantic segmentation architectures are computationally expensive.

To address these challenges above, a U-net based architecture is proposed, since U-net is an efficient and accurate neural network in terms of accuracy which doesn't require large datasets [45, 46].

More specifically, the proposed architecture is composed by an encoder-decoder architecture where a modified version of MobileNetV2 neural network is used as an encoder and a lighter decoder of U-net is utilized for the segmentation stage. To speed up, the learning process, the MobileNetV2 has been trained with ImageNet, a well-known image dataset which includes millions of general-purpose photographs. Thus, during the training process, the pre-trained network "transfers" its earned "experience" to the model, encountering the issue of the limited size of lunar surface dataset.

2.1 Modified U-net architecture

As referred above, the proposed architecture is based on U-net, a well-known architecture for semantic segmentation which was initially proposed for medical applications.

The U-shaped model of U-net can be separated in two main components: (a) the encoder, which reduces the image dimensions, increasing the feature maps while learning to classify the desired features, and (b) the decoder, which reconstructs the image dimensions, decreasing the feature maps and performing precise segmentation of the detected features. The U-net decoder, is able to segment the detected features retrieving the topology of the image content through four skip connections among different levels of the encoder. These connections transfer information to the decoder in order to maintain the spatial details of images with the aim to reconstruct them (Fig. 1).

U-net is mainly composed by convolutional (Conv2D) and "BatchNormalization" layers. Regarding the encoder-decoder functionality, the encoder downsamples the image through the "MaxPooling2D" layer, and the decoder upsamples the image using the UpSampling2D layer while the "Concatenate" layer generates the skip connections between the encoder and decoder part. At the end, "softmax" (Eq. 1) which is the activation function is utilized in order to export the segmentation map for each input image.

$$\sigma(\overrightarrow{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \quad (1)$$

where $\overrightarrow{z}$ is the input vector and z_i presents the elements of the input vector. $\sum_{j=1}^{K} e^{z_j}$ is a normalization term with K classes which ensures that the output of the function will sum to one and each output value will be in a range of (0, 1). In this study, the classes that are represented by K are rocks / boulders, sky and ground (background) (see Sect. 2.2).

Although U-net is an accurate semantic segmentation architecture, it provides increased performance-time while it requires a time-consuming training process with much experimentation in fine-tuning, since it includes about 31,000,000 trainable parameters. In order to accelerate the training process, "transfer learning" technique is utilized, using a pre-trained (with ImageNet dataset) MobileNetV2 as the encoder (Fig. 2).

MobileNetV2 is a CNN-based architecture designed for providing high efficiency in mobile devices while it has been utilized in multiple tasks of computer vision including classification, semantic segmentation, object detection, etc. The

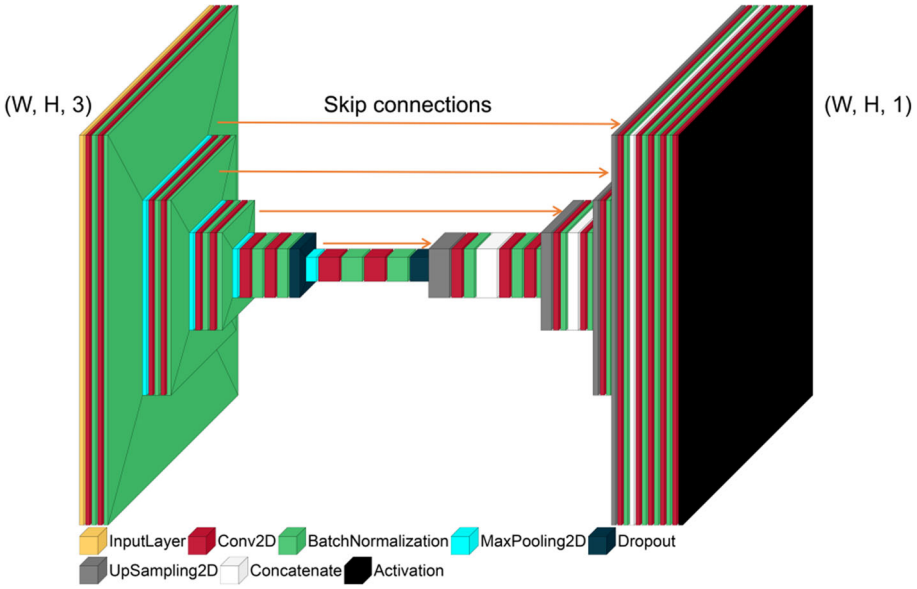

Fig. 1 U-net architecture

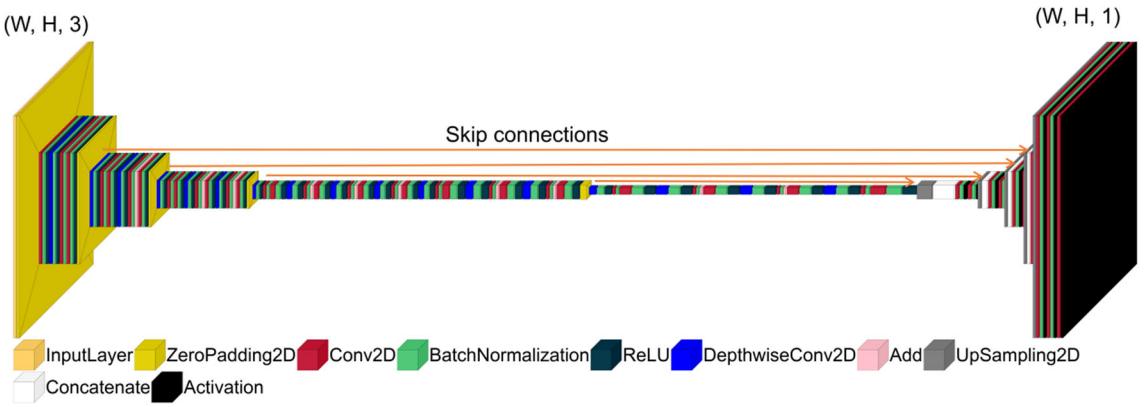

Fig. 2 Architecture of U-net with MobilenetV2 as encoder

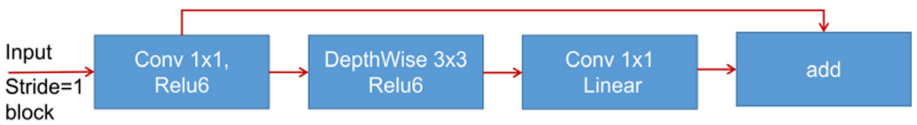

Fig. 3 Inverted residual block architecture

main MobilenetV2 architecture is composed by 19 residual bottleneck layers where each bottleneck is based on inverted residual block. The inverted residual block is based on a narrow-wide-narrow approach using a point-wise convolution with ReLU6, followed by a depth-wise convolution with ReLU6, followed by a linear point-wise convolution. Moreover, a skip connection, merges the input of the block with the output through the "Add" layer (Fig. 3). ReLU6, a modification of the well-known activation function ReLU (Rectified Linear Unit), performs the non-linear transformation aiming the model to learn more complex tasks while outperforms the traditional ReLU in accuracy and execution-time [47].

The approach of inverted residual blocks reduces the extracted parameters and computation compared with conventional convolution layers. According to Sandler et al. 2018 [44] when the kernel k = 3 for 3 × 3 depth-wise convolution, the computational cost is about 9 times smaller compared with traditional convolution without significant reduction in accuracy.

More specifically, if the input of a traditional convolution is $h_i \times w_i \times d_i$ where h and w, the image dimensions and d, the depth or channels while the output is $h_i \times w_i \times d_j$, then the computational cost is calculated as $h_i \times w_i \times d_i \times d_j \times k \times k$, where k, the kernel size, while the corresponding computational cost of an inverted residual block will be: $h_i \times w_i \times d_i(k^2 + d_j)$.

The combination of the original pre-trained MobileNetV2 as an encoder with U-net decoder, provides a more

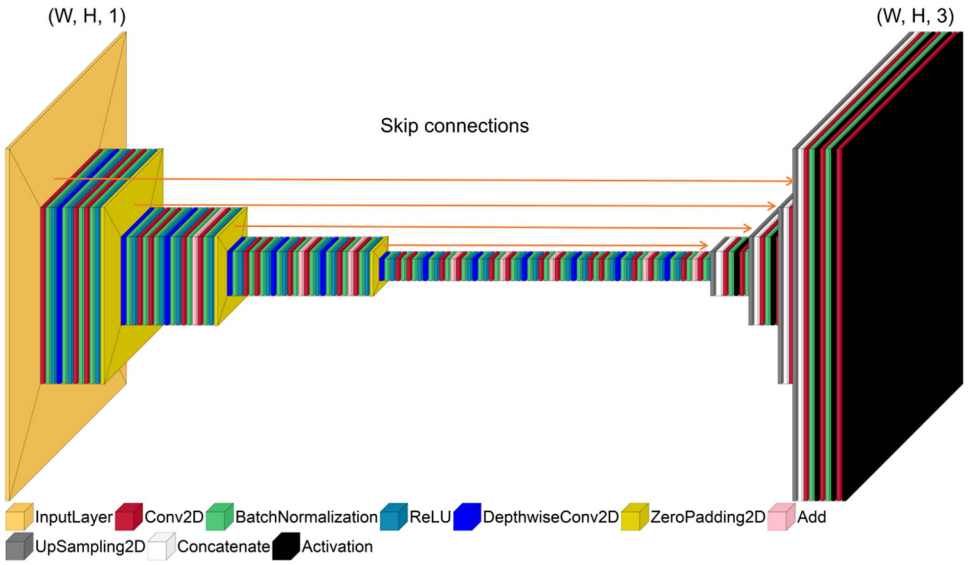

Fig. 4 Proposed architecture

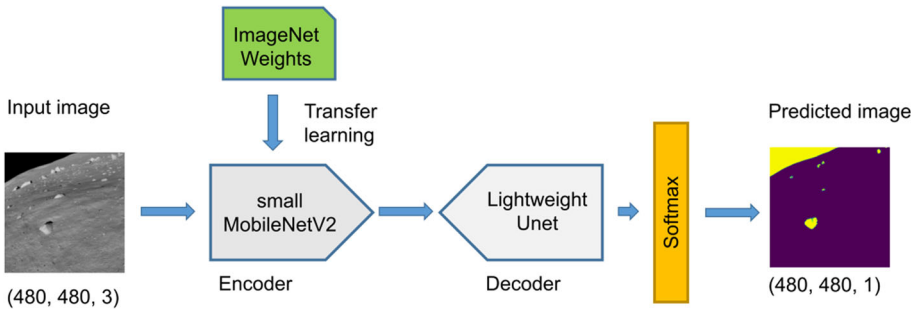

Fig. 5 Proposed architecture for lunar terrain segmentation

lightweight architecture including about 8,000,000 trainable parameters instead of U-net which includes about 31,000,000, while it is able to accelerate the training process. However, this architecture remains unsuitable for applications which require high efficiency in terms of inference-time especially for real-time tasks.

To deal with low-performance time without reducing the accuracy, an architecture based on a modified MobileNetV2 encoder and a lightweight U-net decoder, is proposed.

Regarding the modified MobileNetV2, is composed by an initial fully convolution layer followed by 13 residual bottleneck layers, instead of the original MobileNetV2 which includes 19, since right after the block 13, the parameters are highly increased in the original architecture from about 92,000 to 155,000. Moreover, to further reduce the computational cost, the depth-multiplier which is a positive factor that multiplies the channels through the depth-wise convolution, was defined with a value of 0.35 instead of 1.0 aiming to decrease the output channels of the depth-wise convolution layers. It's worth mentioning that for depth-multiplier values less than 1.0, the depth-multiplier is applied to all layers except the last convolution layer.

Concerning the U-net decoder, all the filters of the convolution layers were divided by the factor of 2 aiming to accelerate the segmentation stage while the four skip connections connects the input image, the block 1, block 3 and block 6 of the encoder respectively.

The proposed architecture includes about 220,000 trainable parameters which are far fewer than the 31,000,000 and 8,000,000 trainable parameters of U-net and original MobileNetV2/U-net respectively.

The proposed architecture with detailed representation of the layers is presented in Fig. 4 while a more abstract representation is depicted in Fig. 5.

2.2 Dataset

As referred above, there is a lack in datasets for lunar surface segmentation while to the best of author's knowledge, there is not rover-based image dataset with real lunar landscapes. Instead, several datasets for the martian surface have been proposed.

Thus, a dataset with artificial rover-based images which depicts lunar landscapes was utilized for training and validation of the proposed architecture,. The dataset was created by the Space Robotics Group of Keio University in Japan, using Planetside Software's Terragen and a DEM (Digital Elevation Model), based on Lunar Orbiter Laser Altimeter on

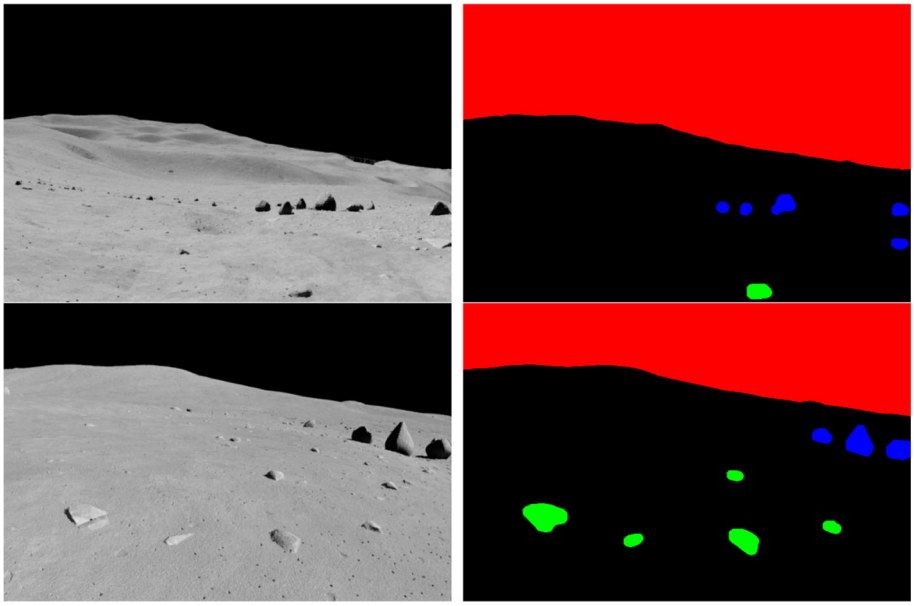

Fig. 6 Dataset of lunar surface for semantic segmentation bySpace Robotics Group of Keio University in Japan. The artificial images are presented in the left column while the corresponding masks in the right column

NASA [48]. It includes about 9,700 artificial images and the corresponding annotated masks taking into account the following four classes: large rocks, small rocks, sky and ground (background) (Fig. 6):

Several drawbacks are included in the dataset, such as the decreased accuracy in feature segmentation and the lack of balance between the classes of large rocks and small rocks. To deal with the imbalanced classes, the two classes of rocks were merged in one class. Thus, the new dataset includes the following classes: rocks, sky and ground (background).

Nevertheless, since this is the only publicly available dataset for the lunar surface focused on semantic segmentation, it was utilized in order to train and validate the proposed architecture, aiming to provide a lightweight model for potential use in systems with low computing resources during the rover navigation, on the lunar surface.

3 Implementation and results

In this section, the implementation of the proposed modified U-net architecture is described while afterwards, the evaluation and results of the model for lunar ground semantic segmentation, are presented.

3.1 Training process of modified U-net

The proposed architecture was implemented using Python and Keras / TensorFlow deep learning library [49] while several Python libraries including NumPy [50], Matplotlib [51] and Scikit-learn [52] were utilized.

The main goal of the architecture is to detect and localize rocks and boulders while in order to segment the whole scene, three classes are taken into account: rocks, sky and background. The training data which constitute the 70% of the lunar landscape dataset feeds the modified U-net while the remaining 30% of the dataset is used for the validation and testing. The model was trained for 15 epochs using the early stopping technique while the batch size was defined equal to 16. The categorical cross entropy loss function and Adam optimizer with a learning rate of 5×10^{-5} were utilized. Regarding the input size, the dimensions of 480×480 pixels was used, since it was observed that a larger image size was provided more refined results than the widely used size of 256×256 pixels.

The training and validation process were conducted in a machine with Intel i7- 3.50 GHz $\times$ 8 cores of CPU, 16 Gb of RAM and NVIDIA GTX 1080 Ti of GPU with CUDA version 11.2 enabled.

3.2 Evaluation and results of modified U-net

The proposed architecture was trained and validated using dice-coefficient, recall, Io (Intersection over Union) and precision metrics which are defined with the following formulas:

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{2}$$

$$\text{Dice} = \frac{\text{TP}}{2\text{TP} + \text{FN} + \text{FP}} \tag{3}$$

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FN} + \text{FP}} \tag{4}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{5}$$

Table 1 Loss function, dice-coefficient, recall, IoU (Intersection over Union) and precision after the training process

	Loss	Dice-coef	Recall	IoU	Precision
Training	0.07	0.79	0.98	0.70	0.77
Validation	0.06	0.78	0.98	0.69	0.76

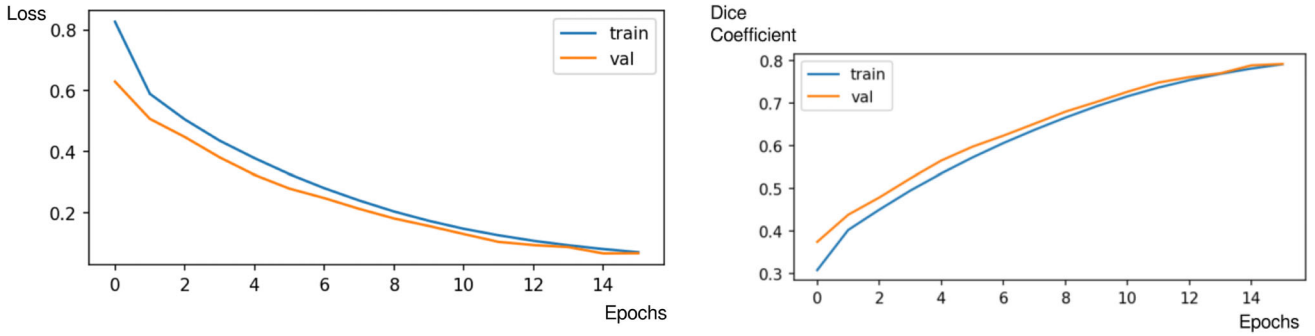

Fig. 7 Loss and dice coefficient curves during training and validation

where, TP stands for true positive while FN and FP stands for false negative and false positive. The results after the training process are presented in Table 1 while the learning curves of loss function and dice coefficient are depicted in Fig. 7.

As observed in Table 1, the value of loss function is below 0.1, the dice-coefficient is in a level of 0.80, the recall is close to 1.0 while the IoU and precision are in a level of 0.70 and 0.75 respectively, indicating that the model is able to provide satisfactory results. Moreover, in Fig. 7 the learning curves of the training and validation process for loss function and dice-coefficient are quite close after the sixth epoch without fluctuations proving that the model doesn't overfit.

After the training process, the proposed architecture was validated in testing data which are completely unknown for the model including images from the synthetic dataset and from real lunar landscape images while the corresponding qualitative results are presented in the Figs. 8 and 9.

As observed in Fig. 8, the proposed architecture provides satisfactory results in testing data with synthetic images, achieving IoU (Intersection over Union) in a level of 0.85 or above. It is able to differentiate the sky from the ground region defining the horizon line with high accuracy while it precisely predicts the location of the small rocks and boulders on the lunar surface. It is not affected from the number of rocks that exist in the scene, since it is able to provide robust results in a scene without any or one rock (Fig. 8d, e) or with multiple small rocks and boulders (Fig. 8c).

Moreover, the proposed architecture achieves respectable results in real rover-based images (Fig. 9a-d) which are quite different in terms of color and illumination compared with the training data. The model is not affected by the camera tilt, being capable of identifying rocks, either the camera targets on the horizon (Fig. 9a, b) or on the ground (Fig. 9c, d).

Regarding size of the model, it includes only 220,000 trainable parameters while the weights file size of the model is about 3.5 MB which is considered quite small for semantic segmentation models. The model was tested in terms of inference time for a set of images with a size of × 480 pixels using three different computing setup: (a) a GPU-enabled conventional desktop machine, (b) CPU-only conventional desktop machine and (c) a CPU-only embedded system with quite low resources. The results are presented in Table 2.

As observed in the Table 2, the model provides quite satisfactory inference time in the GPU-enabled machine achieving 40 ms inference time per image and 25 FPS (Frames per second). The model performs sufficiently without GPU (CPU-only) in the same machine, providing a performance time in a level of 100 ms per image and 10 FPS. The model was also tested on a Raspberry Pi 4 with 4 GB of RAM which is a CPU-only embedded system with quite low resources, providing inference time equal to 1080 ms and 0.92 FPS. Overall, the results are considered respectable taking into account that the image segmentation tasks require high-end GPU-enabled machines and prove that the model can be be used in GPU-enabled or CPU-only conventional machines and embedded systems with low computing resources.

4 Discussion

In this study, a deep learning architecture for semantic segmentation is proposed, which is able to understand semantically a lunar scene, focused on detecting and classifying rocks and boulders. The main goal of this study is the implementation of a lightweight deep learning model with

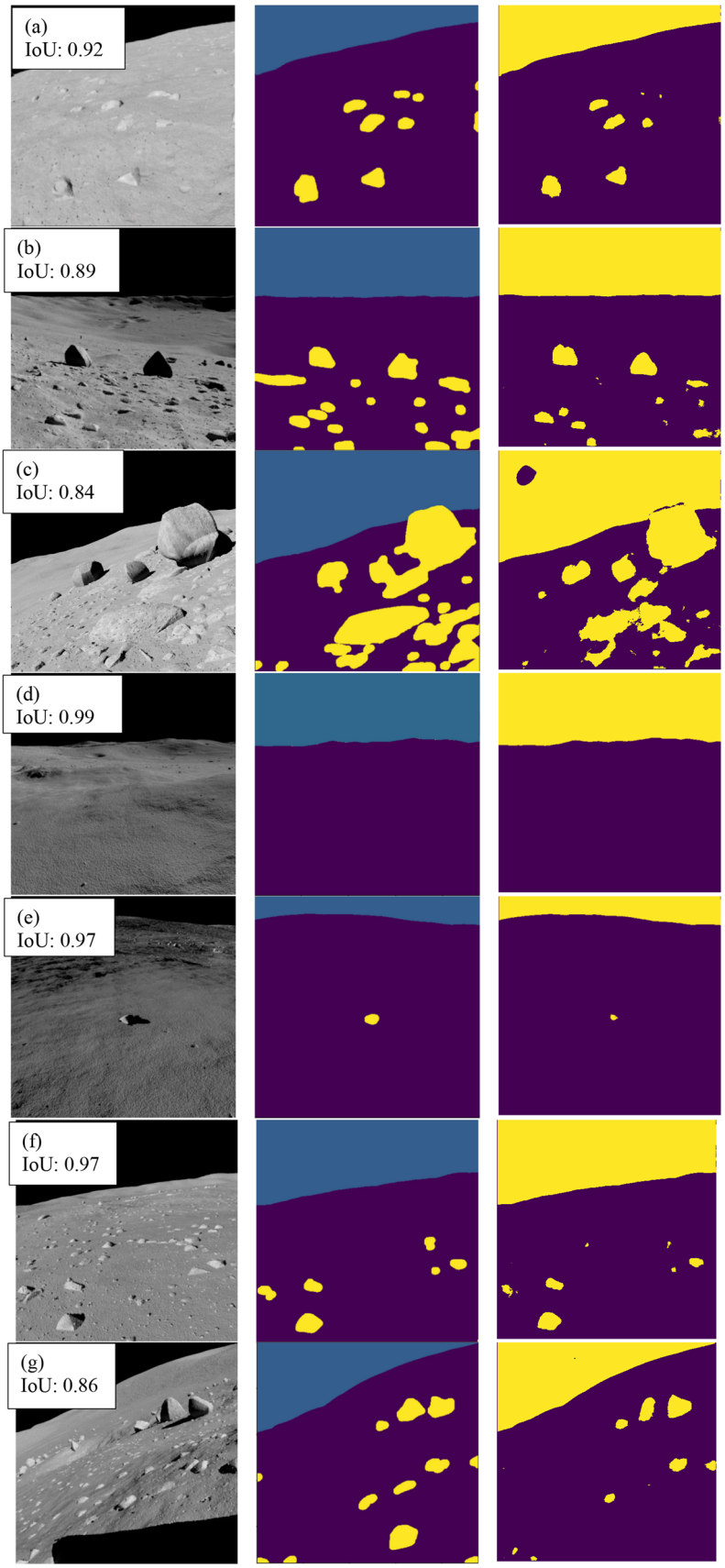

Fig. 8 Left column: Original images from the synthetic lunar surface, (middle column) The corresponding annotated masks, (right column) Predictions of the proposed architecture. In each prediction (row) the IoU (Intersection over Union) metric is presented

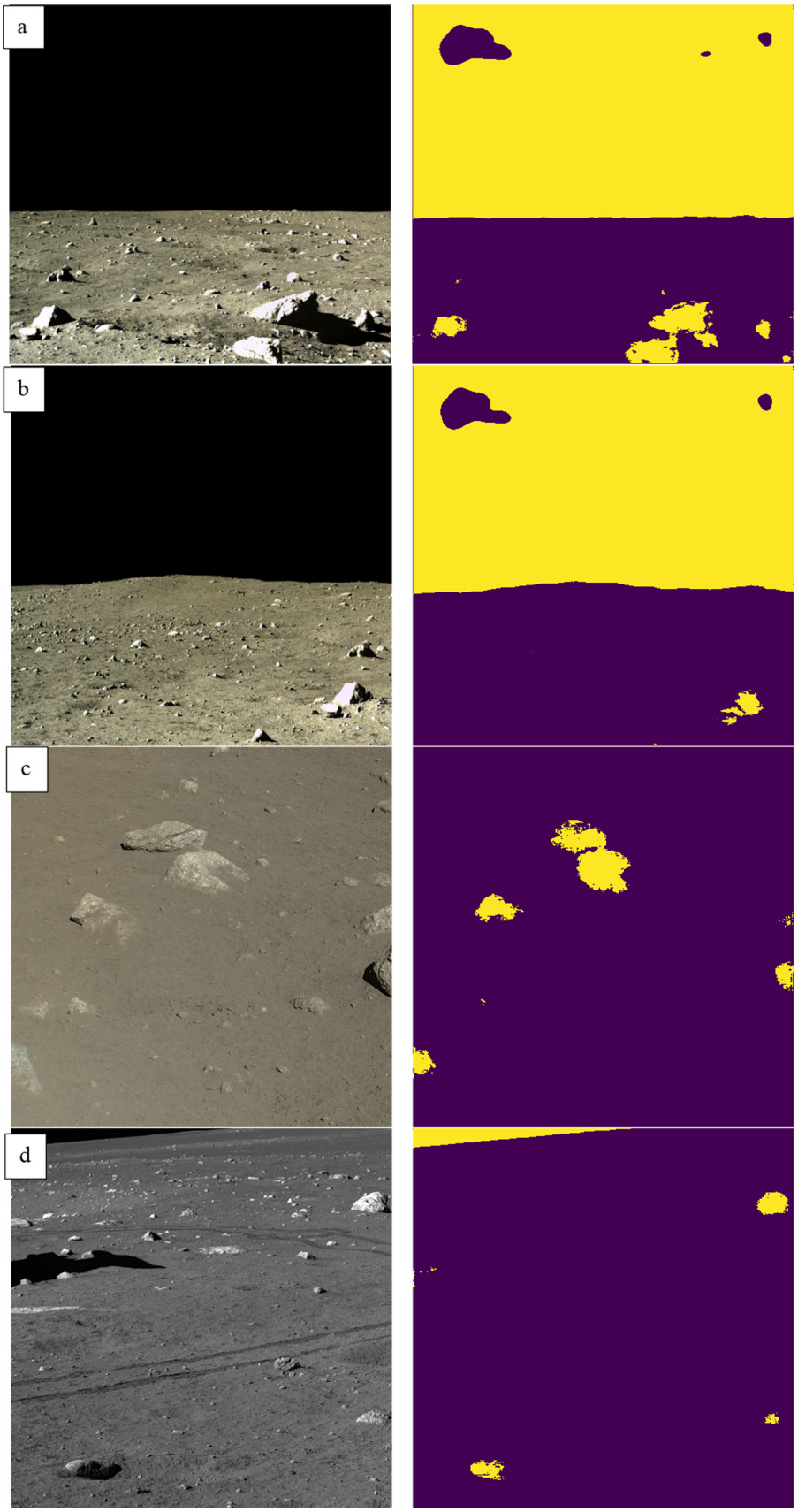

Fig. 9 Left column: Real images from the lunar surface, (right column) Predictions of the proposed architecture. In each prediction (row) the IoU (Intersection over Union) metric is presented

Table 2 Inference time (in milliseconds and FPS) of the proposed model in a desktop GPU-enabled and CPU-only conventional desktop computer and in a CPU-only embedded system with low resources

Inference time	Desktop machine /GPU-enabled		Desktop machine /CPU-only		Embedded system Rasp. Pi 4	
	ms	FPS	ms	FPS	ms	FPS
Proposed model	40	25	100	10	1080	0.92

Table 3 Parameters and model size of the U-net, VGG16/U-net, MobV2/U-net and the proposed architecture

Architecture Encoder/Decoder	Total params	Trainable params	Non-trainable params	Model file size (MB)	Units
U-net	31,061,416	31,047,712	13,704	373.1	17,179
VGG16/U-net	23,752,708	23,748,676	4,032	285.4	12,805
MobV2 / U-net	8,047,876	8,011,780	36,096	97.3	47,423
Proposed architecture	228,588	221,724	6,864	3.5	16,202

a potential use in real-time, in order to increase the safety of rover navigation during a mission on the moon.

Thus, an encoder-decoder architecture was developed which is composed by a modified MobileNetV2 neural network as encoder and a lightweight U-net decoder. Regarding the MobileNetV2 architecture, it includes a fully convolution layer followed by 13 residual bottleneck layers while the depth-multiplier factor was defined in a value of 0.35 instead of the original MobileNetV2 which includes 19 residual bottleneck layers and the default depth-multiplier factor is equal to 1.0. Concerning the segmentation stage, all the filters of U-net decoder were divided by the factor of 2 while the skip connections transfer information related with the spatial content of each image from several layers (the initial input, the block 1, the block 3 and the block 6) of the encoder part.

As presented in Sect. 3.2, the proposed architecture provides robust results achieving IoU in a level of 0.80 or above, detecting and classifying rocks and boulders with satisfactory accuracy in both synthetic and real rover-based images from the lunar surface. To further validate the proposed architecture, it was compared with three similar and widely used encoder-decoder architectures based on U-net:

- The original U-net
- The U-net with VGG16 as encoder
- The U-net with the original MobileNetV2 as encoder

The architectures above, were trained and tested under the same parametrization so as a fair and proper evaluation to be conducted.

The trainable parameters of the proposed architecture is about 220,000 while the corresponding trainable parameters of U-net, VGG16/U-net and MobileNetV2/U-net are about 31,000,000, 24,000,000 and 8,000,000 respectively. The weights file sizes are about 370 MB for U-net, 285 MB for VGG16/U-net and 97 MB for MobileNetV2/U-net while the corresponding weights file size of the proposed architecture is about 3.5 MB (Table 3).

In Fig. 10, qualitative results from the alternative and the proposed architectures are depicted while in Table 4 the corresponding IoU score is presented. It's worth noting that original MobileNetV2/U-net could not converge with this specific parametrization, thus in the results below the proposed architecture is compared with original U-net and VGG16/U-net.

As observed in Fig. 10, all the models produce respectable segmentation results. In Fig. 10a, b, c and d the proposed model provides similar accuracy in rocks segmentation compared with the original U-net and VGG16/U-net, predicting all the important rocks and boulders that could harm a rover during navigation. On the other hand, in Fig. 10e and f which are depicted real images from lunar surface, the proposed architecture provides refined segmentation results compared with the alternative models. For instance, in Fig. 10e, the proposed model precisely segments the two main rocks on the ground instead of original U-net which fails to predict them while VGG16/U-net falsely unifies them in a bigger rock. Similarly, in Fig. 10f, the proposed model and VGG16/U-net produce quite close results while original-U-net falsely predicts a large shadow as a rock.

Regarding the evaluation of the models on testing data in terms of intersection over union (IoU), the proposed architecture provides an IoU score of 0.84 (Table 4) outperforming the VGG16/U-net while is close to IoU of U-net which is equal to 0.86. The results above, determines the superiority of the proposed architecture since, it is about 110 times and about 140 times smaller than the VGG16/U-net and the original U-net respectively while provides similar segmentation predictions in both alternative architectures.

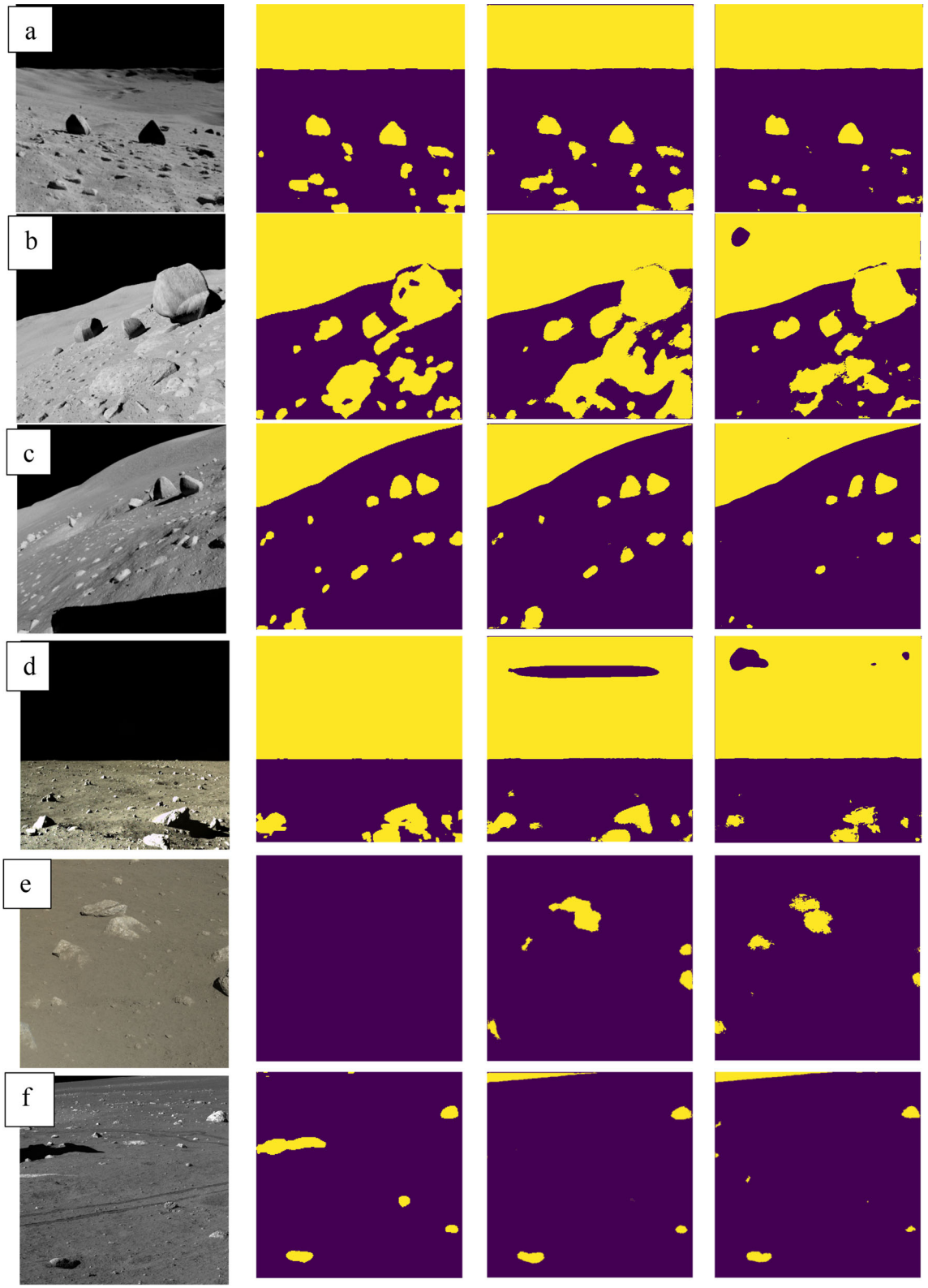

Fig. 10 First column: original synthetic (a, b, c) and real (d, e, f) lunar images, second column: original U-net model predictions, third column: VGG16/U-net model predictions, fourth column: proposed architecture

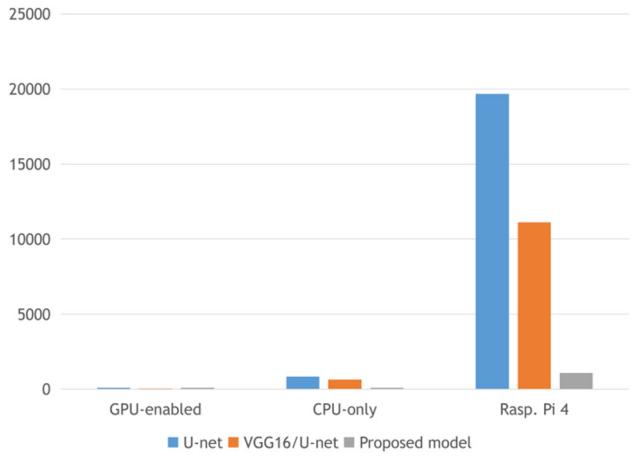

Fig. 11 Inference time in millisecond (ms) of the U-net, VGG16 / U-net and the proposed model for the GPU-enabled machine, the CPU-only machine, and the Raspberry Pi 4 embedded system

Table 4 IoU score in testing data of the original U-net, VGG16/U-net and the proposed model, trained with the same dataset and parametrization

Architecture Encoder/Decoder	IoU
U-net	0.86
VGG16 / Unet	0.82
Proposed model	0.84

It's worth noting that, although all the models provide robust results in sky segmentation defining the horizon line, they are unable to classify the sky as separate class. This is due to dataset's lack of color variety on the ground features and the large black shadows presented on the ground. Thus, because the images are synthetic, there is no a meaningful difference between the sky and the large black areas on the ground. Nevertheless, a refined synthetic rover-based dataset or a dataset with real lunar landscape images, would solve this issue, improving the classification results of all the models.

Regarding the inference-time, the models were tested on a large set of images with a size of 480 × 480 in three different computing setups: (a) a GPU-enabled conventional desktop machine, (b) CPU-only conventional desktop machine and (c) a CPU-only embedded system with quite low resources. The corresponding results are presented in the Table 5 and Fig. 11.

As observed in Table 5 and Fig. 11, the proposed model achieves quite less inference time compared with the U-net and VGG16 / U-net while the difference in performance-time is increased among the models when the computing resources are reduced. Regarding the GPU-enabled machine, the proposed model achieves 43 ms and 23.25 FPS, while the VGG16/U-net provides 52 ms (19.23 FPS) of inference-time and U-net about 100 ms (10 FPS) which is twice the time compared with the proposed model. In the CPU-only machine, the proposed model provides inference-time in a level of 100 ms (10 FPS) while the VGG16 / U-net and U-net models perform predictions with 640 ms (1.56 FPS) and 850 ms (1.17 FPS) inference-time respectively, six and nine times more than the proposed model. Concerning the Raspberry Pi 4 with 4GB of RAM embedded system, the proposed model achieves an inference-time about 1080 ms (0.92 FPS) which is quite satisfactory since to the best of our knowledge, this embedded system provides the lowest computing resources on the market, especially in deep learning. Instead, the VGG16/U-net and U-net models provide 11,120 ms (0.09 FPS) and 19,680 ms (0.05 FPS) inference-time, proving that the proposed model is about 11 and 20 times faster in the Raspberry Pi 4 embedded system compared with the VGG16/U-net and U-net models respectively.

5 Conclusions

In summary, an encoder-decoder architecture for semantic segmentation was developed, aiming to reinforce the safety of rover navigation on the lunar surface. The main goal of this study was the implementation of a semantic segmentation model for the lunar surface, capable of being utilized by embedded systems with low computational resources.

To achieve this goal, a deep learning architecture based on U-net neural network was developed, since U-net is able to provide respectable results, trained with limited size of datasets [21]. To reduce the computational cost of U-net, a modified MobileNetV2 neural network was used as the encoder, while a lighter version of U-net decoder were implemented in order to accelerate the segmentation stage. The proposed architecture was fed with a publicly available dataset which includes rover-based synthetic images from the lunar surface. Although it contains several limitations and drawbacks, including lack of color variations, and low accuracy in labeling of features, to the best of author's knowledge, it is the only available dataset of the lunar surface focused on deep learning models' training.

As a result, the proposed model achieves satisfactory accuracy in scene segmentation, in synthetic images and in real rover-based images of the lunar surface while it includes significantly less trainable parameters than U-net based alternatives. The proposed architecture was evaluated compared with the original U-net, with VGG16/U-net and with the original MobileNetV2/Unet neural networks which were trained under the same parametrization. The trainable parameters and weights file size of the models proved that the proposed architecture is about 140 times smaller than the original U-net, 110 times than the VGG16/U-net and 36 times smaller

Table 5 Comparison in terms of inference time (in milliseconds and FPS) of the original U-net, VGG16/U-net and the proposed model in a desktop GPU-enabled and CPU-only conventional desktop computer and in a CPU-only embedded system with low resources

Inference time per image	Conventional machine /GPU-enabled		Conventional machine / CPU-only		Embedded system / Rasp. Pi 4	
	ms	FPS	ms	FPS	ms	FPS
U-net	100	10	850	1.17	19,680	0.05
VGG16/U-net	52	19.23	640	1.56	11,120	0.09
Proposed model	43	23.25	100	10	1080	0.92

than the original MobilenetV2/U-net while it provides quite close accuracy in terms of IoU with the original U-net and outperforms the U-net based alternatives. Moreover, the models were tested in three different computing setups, two conventional machines (GPU-enabled and CPU-only) and an embedded system with low computing resources, proving that the proposed model is quite faster than U-net and VGG6/U-net in all computing systems and especially in the embedded system.

However, due to the aforementioned drawbacks of the dataset, the proposed model could further be improved especially in classification but also in segmentation task adding more classes such as, sandy regions, bedrocks, craters, etc., using a more qualitative dataset with synthetic or even better with real rover-based lunar images. Given that a qualitative dataset from the lunar surface will be available in the near future due to the planned missions of NASA's Artemis program, the proposed architecture is able to provide a significant potential in lunar scene understanding, ensuring safe and precise navigation, and to contribute in groundbreaking discoveries, expanding the scientific understanding of the Moon.

Acknowledgements The implementation of the doctoral thesis was co-financed by Greece and the European Union (European Social Fund-ESF) through the Operational Programme «Human Resources Development, Education and Lifelong Learning» in the context of the Act "Enhancing Human Resources Research Potential by undertaking a Doctoral Research" Sub-action 2: IKY Scholarship Programme for PhD candidates in the Greek Universities.

Author contributions Conceptualization: GP, PP; Methodology: GP; Formal analysis and investigation: GP; Software: GP; Writing—original draft preparation: GP; Writing—review and editing: PP, GP; Funding acquisition: PP; Resources: GP; Supervision: PP.

Funding Open access funding provided by HEAL-Link Greece. This research received no external funding.

Data availability The dataset that is used in this study is publicly available by Keio University of Japan.

Code availability There is no available code.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

1. Swan, R.M., Atha D., Leopold, H.A., Gildner, M, Oij, S., Chiu. C., Ono M., AI4MARS: A Dataset for Terrain-Aware Autonomous Driving on Mars. In: Proceedings of the 2021 IEEE/CVF, CVPRW, 2021, pp. 1982–1991
2. George, D.A., Privitera, C.M., Blackmon, T.T., Zbinden, E., Stark, L.W.: Segmentation of stereo terrain images. Proc. Human Vision Electr. Imaging V, Bellingham, WA, USA **3959**, 669–679 (2000). https://doi.org/10.1117/12.387204
3. Howard, A., Seraji, H.: An intelligent terrain-based navigation system for planetary rovers. IEEE Robot. Autom. Mag. **8**(4), 9–17 (2001). https://doi.org/10.1109/100.973242
4. Gong, X., and Liu, J. (2012). Rock detection via superpixel graph cuts. In: 2012 19th IEEE international conference on image processing (pp. 2149-2152). IEEE.https://doi.org/10.1109/ICIP.2012.6467318
5. Di, K., Yue, Z., Liu, Z., Wang, S.: Automated rock detection and shape analysis from mars rover imagery and 3D point cloud data. J. Earth Sci. **24**, 125–135 (2013). https://doi.org/10.1007/s12583-013-0316-3
6. Song, Y., and Shan, J. (2006). A framework for automated rock segmentation from the Mars Exploration rover imagery. In Proceedings of ASPRS 2006 Annual Conference, Reno, Nevada, USA.
7. Dunlop, H., Thompson, D. R., & Wettergreen, D. (2007). Multi-scale features for detection and segmentation of rocks in mars images. In: 2007 IEEE conference on computer vision and pattern recognition (pp. 1-7). IEEE.https://doi.org/10.1109/CVPR.2007.383257.
8. Fujita, K., and Ichimura, N. (2011). A terrain classification method for planetary rover utilizing dynamic texture. In: AIAA Guidance, Navigation, and Control Conference (p. 6580). https://doi.org/10.2514/6.2011-6580

9. Lu, S., Oij, S. L. (2017). Horizon detection for mars surface operations. In: 2017 IEEE Aerospace Conference (pp. 1-8). IEEE. https://doi.org/10.1109/AERO.2017.7943975
10. Kuang, B., Gu, C., Rana, Z.A., Zhao, Y., Sun, S., Nnabuife, S.G.: Semantic terrain segmentation in the navigation vision of planetary rovers—A systematic literature review. Sensors. **22**(21), 8393 (2022). https://doi.org/10.3390/s22218393
11. Baheti, B., Innani, S., Gajre, S., Talbar, S.: Semantic scene segmentation in unstructured environment with modified DeepLabV3+. Pattern Recogn. Lett. **138**, 223–229 (2020). https://doi.org/10.1016/j.patrec.2020.07.029
12. Chen, L.C., Papandreou, G., Kokkinos, I., Murphy, K., Yuille, A.L.: DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. **40**(4), 834–848 (2018). https://doi.org/10.1109/TPAMI.2017.2699184
13. Chollet F, Xception: Deep Learning with Depthwise Separable Convolutions, ArXiv, 2016
14. Baheti, B., Innani, S., Gajre, S., & Talbar, S. (2020). Eff-unet: A novel architecture for semantic segmentation in unstructured environment. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (pp. 358-359).https://doi.org/10.1109/CVPRW50498.2020.00187
15. Tan, M., & Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning (pp. 6105-6114). PMLR.
16. Guan, T., Kothandaraman, D., Chandra, R., Sathyamoorthy, A.J., Weerakoon, K., Manocha, D.: GA-Nav: efficient terrain segmentation for robot navigation in unstructured outdoor environments. IEEE Robot. Automat. Lett. **7**(3), 8138–8145 (2022). https://doi.org/10.1109/LRA.2022.3187278
17. Guan, T., He, Z., Song, R., Manocha, D., & Zhang, L. (2021). Tns: Terrain traversability mapping and navigation system for autonomous excavators. arXiv preprint arXiv:2109.06250.
18. Metzger, K., Mortimer, P., Wuensche, J.H., A Fine-Grained Dataset and its Efficient Semantic Segmentation for Unstructured Driving Scenarios, ArXiv 2021
19. Wigness, M., Eum, S., Rogers, J. G., Han, D., & Kwon, H. (2019). A rugd dataset for autonomous navigation and visual perception in unstructured outdoor environments. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 5000-5007). IEEE.https://doi.org/10.1109/IROS40897.2019.8968283
20. Furlán, F., Rubio, E., Sossa, H., & Ponce, V. (2019). Rock detection in a Mars-like environment using a CNN. In: Pattern Recognition: 11th Mexican Conference, MCPR 2019, Querétaro, Mexico, June 26–29, 2019, Proceedings 11 (pp. 149-158). Springer International Publishing.https://doi.org/10.1007/978-3-030-21077-9_14
21. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation, medical image computing and computer-assisted intervention – MICCAI 2015. MICCAI (2015). https://doi.org/10.1007/978-3-319-24574-4_28
22. Furgale P T, Carle P, Enright J, and Barfoot T D, The Devon Island Rover Navigation Dataset, International Journal of Robotics Research, 2012
23. Furlán, F., Rubio, E., Sossa, H., Ponce, V.: CNN based detectors on planetary environments: a performance evaluation. Front. Neurorobot. **14**, 590371 (2020). https://doi.org/10.3389/fnbot.2020.590371
24. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC, SSD: Single Shot MultiBox Detector, In: Proceedings of ECCV 2016, Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-46448-0_2
25. Kuang, B., Wisniewski, M., Rana, Z.A., Zhao, Y.: Rock segmentation in the navigation vision of the planetary rovers. Mathematics **9**(23), 3048 (2021). https://doi.org/10.3390/math9233048
26. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J., UNet++: A Nested U-Net Architecture for Medical Image Segmentation. In Proceedings of DLMIA, 2018, Springer, https://doi.org/10.1007/978-3-030-00889-5_1
27. Hewitt, R., Boukas, E., Azkarate, M., Pagnamenta, M., Marshall, J., Gasteratos, A., Visentin, G.: The Katwijk beach planetary rover-dataset. Int. J. Robot. Res. (2018). https://doi.org/10.1177/0278364917737153
28. Tomita, K., Skinner, K., Iiyama, K., Jagatia, B., Nakagawa, T., Ho, K.: Hazard detection algorithm for planetary landing using semantic segmentation, AIAA 2020–4150. ASCEND (2020). https://doi.org/10.2514/6.2020-4150
29. Claudet, T., Tomita, K., Ho, K.: Benchmark analysis of semantic segmentation algorithms for safe planetary landing site selection. IEEE Access **10**, 41766–41775 (2022). https://doi.org/10.1109/ACCESS.2022.3167763
30. Kuang, B., Rana, Z.A., Zhao, Y.: Sky and ground segmentation in the navigation visions of the planetary rovers. Sensors **21**(21), 6996 (2021). https://doi.org/10.3390/s21216996
31. Ebadi K., Coble K., Atha D., Schwartz R., Padgett C., Hook J.V., Semantic mapping in unstructured environments: Toward autonomous localization of planetary robotic explorers. In: IEEE Aerospace Conference, 2022
32. Lin, M., Chen, Q., Yan, S., Network in Network, arXiv 2013
33. Mihail R.P., Workman S., Bessinger Z., Jacobs N., Sky segmentation in the wild: An empirical study. In: Proceedings of WACV, Lake Placid, NY, USA, 7–10 March 2016
34. Müller, M.G., Durner, M., Gawel, A., Stürzl, W., Triebel, R., Siegwart, R., A Photorealistic Terrain Simulation Pipeline for Unstructured Outdoor Environments. In: Proceedings of IROS, Prague, Czech Republic, 2021, pp. 9765–9772, DOI: https://doi.org/10.1109/IROS51168.2021.9636644
35. Swan R.M., Atha D., Leopold H.A., Gildner M., Oij S., Chiu C., Ono M., AI4MARS: A Dataset for Terrain-Aware Autonomous Driving on Mars. In: Proceedings of CVPRW, Nashville, TN, USA, 2021, pp. 1982–1991, DOI: https://doi.org/10.1109/CVPRW53098.2021.00226
36. Chiodini S., Torresin L., Pertile M., Debei S., Evaluation of 3D CNN Semantic Mapping for Rover Navigation, ArXiv 2020
37. Huang, G., Yang, L., Cai, Y., Zhang, D.: Terrain classification-based rover traverse planner with kinematic constraints for Mars exploration. Planet. Space Sci. **209**, 105371 (2021). https://doi.org/10.1016/j.pss.2021.105371
38. Zhao H., Shi J., Qi X., Wang X., Jia J., (2017), Pyramid scene parsing network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, https://doi.org/10.1109/CVPR.2017.660
39. Chiodini S., Pertile M., Debei A., Occupancy grid mapping for rover navigation based on semantic segmentation, ACTA IMEKO, 2021, https://doi.org/10.21014/acta_imeko.v10i4.1144
40. Jia Y., Wan G., Liu L., Wu Y., Zhang C., (2020) Automated Detection of Lunar Craters Using Deep Learning. In: Proceedings of ITAIC, Chongqing, China, https://doi.org/10.1109/ITAIC49862.2020.9339179.
41. Hashimoto, S., & Mori, K. (2019). Lunar crater detection based on grid partition using deep learning. In: 2019 IEEE 13th International Symposium on Applied Computational Intelligence and Informatics (SACI) (pp. 75-80). IEEE.https://doi.org/10.1109/SACI46893.2019.9111474
42. Hu, Y., Xiao, J., Liu, L., Zhang, L., Wang, Y.: Detection of small impact craters via semantic segmenting lunar point clouds using deep learning network. Remote Sens. **13**(9), 1826 (2021). https://doi.org/10.3390/rs13091826
43. Moghe, R., Zanetti, R.: A deep learning approach to hazard detection for autonomous lunar landing. J. Astronaut. Sci. **67**(4), 1811–1830 (2020). https://doi.org/10.1007/s40295-020-00239-8

44. Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 4510-4520).
45. Jeon, E.I., Kim, S., Park, S., Kwak, J., Choi, I.: Semantic segmentation of seagrass habitat from drone imagery based on deep learning: a comparative study. Ecol. Inform. **66**, 101430 (2021). https://doi.org/10.1016/j.ecoinf.2021.101430
46. Chhabra, S., Rohilla, R.: A comparative study on semantic segmentation algorithms for autonomous driving vehicles. Ijraset J. Res. Appl. Sci. Eng. Technol. (2022). https://doi.org/10.22214/ijraset.2022.44511
47. Nworu, C.C., Ekpenyong, J.E., Chisimkwuo, J., Okwara, G., Agwu, O.J., Onyeukwu, N.C.: the effects of modified ReLU activation functions in image classification. J Biomed. Eng. Med. Dev. **7**, 237 (2022)
48. Smith, E., Zuber, T., Jackson, B., et al.: The lunar orbiter laser altimeter investigation on the lunar reconnaissance orbiter mission. Space Sci. Rev. **150**, 209–241 (2010). https://doi.org/10.1007/s11214-009-9512-y
49. Chollet F. et al, Keras, 2015 GitHub. Retrieved from https://github.com/fchollet/keras
50. Harris, C.R., Millman, K.J., van der Walt, S.J., et al.: Array programming with NumPy. Nature **585**, 357–362 (2020). https://doi.org/10.1038/s41586-020-2649-2
51. Hunter, J.D.: Matplotlib: a 2D graphics environment. Comput. Sci. Eng. **9**(3), 90–95 (2007)
52. Pedregosa, M., et al.: Scikit-learn: machine learning in python. JMLR **12**, 2825–2830 (2011)

Georgios Petrakis is a geospatial data scientist and engineer specializing in geographic information systems and artificial intelligence. He completed his undergraduate studies at Rural, Surveying & Geoinformatics engineering from the National Technical University of Athens and received his master's degree from the school of Mineral Resources Engineering at the Technical University of Crete. Subsequently, he pursued PhD studies in the same department where he focused on improving autonomous navigation and scene understanding in planetary environments with reduced visual information and illumination. He has participated in several national and European research projects, while his main research interests include topics from satellite remote sensing, semantic segmentation and real-time object detection.

Panagiotis Partsinevelos is an Associate Professor in the area of Space Informatics including Uncrewed Aerial Systems, GIS, Remote Sensing and GNSS from their computer science perspective. He received his PhD in Spatial Information Science & Engineering from the University of Maine, part of the National Center for Geographic Information and Analysis (NCGIA) in USA and NASA Center of Excellence in Remote Sensing Applications. Prof. Partsinevelos directs SenseLab Research group, a leading interdisciplinary research entity developing novel solutions for three-dimensional processing, tangible GIS, gestural interfaces, visualization platforms, location-based services, ML algorithms, smart cities, etc. SenseLab has been recognized by a series of prestigious international awards and distinctions including 2019 Airbus Global Earth Observation challenge, 1st globally, 2018 RMIT drones for refugees, Amman, 2017 Space Oscars, Tallinn, 1st globally, 2016 European GNSS Service (GSA), 1st place globally, 2016 ESNC Satellite masters, 2nd overall winners, (400 teams), Madrid, 2016 UAE Drones for Good (1st in EU and 3rd internationally between 1017 teams from 165 countries), Dubai, 2016 DJI drones Developer Challenge (short-listed), USA, 2015 Copernicus Masters NCMA, 1st winner in Remote Sensing visualization, Berlin. After many years of teaching in many versatile and demanding environments, countries, continents, cultures and levels, Dr. Partsinevelos is privileged with long-term collaborations, and through a genuine inclination towards philosophy, humanities and cognition, he is always there to creatively share an provoke classroom experiences.