2024 International Conference on Space Robotics (iSpaRo)
June 24-27, 2024, Luxembourg

A Mask R-CNN Approach to Identify Lunar Landforms in Diverse Lighting Conditions

Aelyn Chong Castro[1], Sofia Coloma [1], Ernest Skrzypczyk[1] and Miguel A. Olivares-Mendez[1]

2024 International Conference on Space Robotics (iSpaRo) | 979-8-3503-6723-2/24/$31.00 ©2024 IEEE | DOI: 10.1109/iSpaRo60631.2024.10687946

***Abstract*— Planetary rovers have limited autonomous navigation capabilities. Delays in communication and terrain assessment significantly restrict the explored area and pose a risk to the mission lifespan. Enhancing autonomy is crucial for efficient exploration without constant human intervention. Scientists have explored various techniques to enable autonomous traversal across unfamiliar terrain. One crucial aspect is the detection and avoidance of obstacles and hazardous environments. Rock detection has been significantly challenging since rocks exist in different colors, shapes, sizes, and textures. This study uses transfer learning on Mask R-CNN to detect natural lunar features such as rocks, pebbles, and craters. The proposed model undergoes evaluation using three distinct cameras: the Ricoh Theta 360 for a panoramic view and the Mint Eye D and ZED 2 for stereo vision capabilities. Furthermore, two varied lighting conditions—full and partial illumination— are assessed, simulating a lunar analog environment. Finally, validation with Yutu-1 PCAM (Chang'e 3) imagery confirms its applicability on the Moon, achieving average detection confidence rates of 90.9% for rocks, 80.15% for pebbles, and 79.35% for craters.**

I. INTRODUCTION

Planetary exploration traces its origins back to 1970 with the teleoperation of the Lunokhod 1 rover from Earth, initiating surface exploration and soil testing on the Moon. In the following decades, more rovers ventured to explore the lunar and martian surfaces. Yutu-1 (Chang'e 3) inspected the lunar surface, and it became stranded. Likewise, Spirit and Curiosity Mars rovers also faced harsh conditions; Spirit got stuck in soft soil, and Curiosity got damaged as the sharp and hard surface of the rocks originated fatigue in the wheels' material, causing holes, dents, and scratches. Thereby, effective rock and pebble detection can extend the operational lifespan of the rover wheels, consequently prolonging mission durations.

Identification of environmental features (e.g., rocks, soil type, craters, etc.) is necessary to increase the reliability of autonomous rover navigation. It helps rover systems make quick decisions, avoiding risky situations that could endanger the mission, especially when there's a delay in human intervention due to long distances (e.g., between Earth and the Moon or Mars). Enhancing landscape recognition optimizes the rover's path planning for safe task completion and facilitates the detection of valuable resources such as mineral-rich rocks. However, developing methods that can effectively identify and track features in dark or low illumination in both structured and unstructured environments is challenging. Further, creating feature detection for celestial bodies is more complicated due to the limited access to data resources.

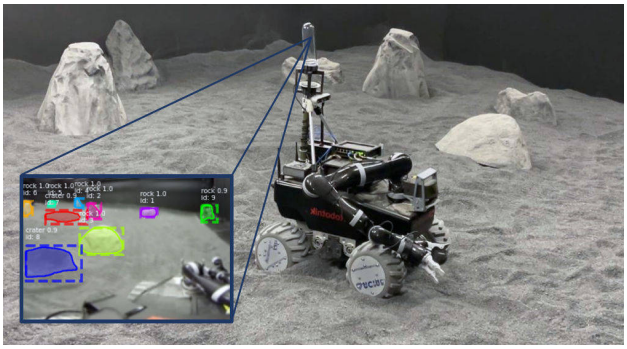

Fig. 1: Detection and segmentation of natural features on the Moon based on Artificial Intelligence techniques.

This paper proposes a method to detect natural features on the Moon based on the Mask-RCNN architecture and transfer learning techniques with a custom dataset that comprises images from an analog lunar environment (LunaLab [1]), and authentic lunar imagery. The model effectively identifies lunar features such as craters, rocks, and pebbles by extracting the object regions from the background at the pixel level, either in a low or high-lighting environment or in an unstructured environment. Validation tests used lunar images captured by Apollo missions (Apollo 11-15) and the Yutu PCAM rover (Chang'e 3) to assess the model's efficiency and adaptability for planetary exploration. Furthermore, the model performance is evaluated with three different cameras ((1) Ricoh Theta V, (2) ZED 2, and (3) Mynt Eye D) providing insights into the influence of rover camera specifications on detection accuracy.

II. RELATED WORK

Rock detection presents many difficulties since they exist in diverse colors, shapes, sizes, and textures. Moreover, in planetary environments, rocks can be covered in lunar or martian dust, embedded in the terrain, or clustered together, increasing the difficulty of differentiating them from the rest of the environment. To address these challenges, researchers have been studying different methods for rock detection. The main approaches are summarized below and compared in Table I.

Shadow/Region Contrast Methodology Gulick et al. [2] uses shadow and region contrast to detect rocks. They gathered data parameters from the Marsokhod rover related to azimuth, roll, elevation angles, and optics. By converting the image plane into a 3D map, prominent elevations

[1] Space Robotics Research Group (SpaceR), Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg, miguel.olivaresmendez@uni.lu

Method	Rock Detection											
	Size		Illumination		Quantity		Conditions			Mapping		
	Large	Small	Light	Dark	Individual	Cluster	Embedded in the terrain	Texture	Measurement	2D mapping	3D mapping	Accuracy
Shadow/ Region Contrast [2]	x		x							x	x	Not Specified
Shadow/ Region Contrast [3], [4], [5]	x	x	x		x	x			x	x	x	Not Specified
Edge Detector [6], [7], [8], [9]	x	x	x	x	x			x		x	x	94.72% - 73.35%
Edge detector and Image Segmentation [10]	x	x	x		x	x	x	x		x		90%
Superpixelization [11], [12]	x		x	x	x	x		x				81.95%
Superpixelization/ Region Contrast [13]	x		x		x			x				75%
SVM [14], [15]	x	x	x	x	x	x						99.41%
Intensity-based Range-based Segmentation [16]	x	x	x	x	x	x	x	x	x			Not Specified
Template-based Detection [17]	x	x	x		x	x						78%
CNN [18]	x	x	x	x	x	x						97.6% - 95.4%
CNN [19], [20]	x		x	x	x	x	x	x				97.10% - 100.0%
CNN & Transformer [21], [22], [23]	x		x		x		x	x				91.10% - 98.91%
INSTR [24]	x	x	x		x				x			96.08%

TABLE I: Comparison between different Lunar and Martian rock detection methods

that cast shadows are identified and classified as rocks. Golombek et al. [4], [5] use techniques previously developed by [25], [26], [27], [28], and [29]. Their method separates the shadow cast by rocks from the background by maximum entropy thresholding and applying gamma enhancement. The rock's shadows are used as a reference to measure the size, quantity, and distance from the rover or lander.

Edge-based Methodology Rockfinder [14], [6] utilizes morphology and size to detect rocks, distinguishing between small and large rocks by identifying closed-shape contours while reducing image resolution to mitigate noise from albedo or dust. CLARAty [7] employs stereo processing to create a point cloud of the environment, enabling the rover to identify and locate rocks within a coordinate frame using range data. Burl et al. [10]introduced Rockster and Rockster-MER, which segment rocks through edge regrouping, providing detailed shape characterizations alongside rock detection. Li et al. [30] approach uses gradient differences for rock and crater detection and abundance on the direction of the illumination in optical images.

Superpixelization Methodology Gong and Liu [11] used a Superpixelization for rock detection using entropy rate-based segmentation. The principle of Superpixelization is to over-segment an image into multiple small mosaics, resulting in homogeneous regions that closely align with objects in the image boundaries. Xiao et al. [13] continued to explore this approach. The algorithm was tested using panoramic cameras from Mars Exploration Rovers Spirit and Opportunity, as well as images from a five-degrees of freedom (DOF) planetary surface simulation platform. Xiao et al. [13] uses intensity and spatial layout to split homogeneous regions. [12] further explore this technique using two rock detection kernels (KPCA and KLRR) on Martian imagery.

SVM Thompson and Castano [14] employed a Support Vector Machine (SVM) classifier to characterize pixels likely to represent rocks, constructing feature vectors from region intensity values. Similarly, Kalra and Bhavsar [15] segmented images into four elements: sky, surface, large rocks, and small rocks, utilizing SVM to distinguish between large and small rock formations.

Intensity-based and Range-based Segmentation Gor et al. [16] integrated two distinct analysis techniques for rock detection, incorporating hierarchic and symmetric object recognition. Small rocks are identified through an intensity-based algorithm, which tracks color and texture changes in each image to generate an edge-flow vector. For larger rocks, range-based segmentation is employed, utilizing multiple cameras onboard the rover to calculate pixel displacement from various viewpoints. Di et al. [31] utilizes the mean shift method to segment homogeneous objects in the image, categorizing them as small rocks while employing a 3D point cloud to extract and characterize larger rocks, including their shape properties.

Deep Learning Suebe et al. [18] uses Convolutional Neural Networks (CNN) for terrain classification. The algorithm uses ResNet-50 as a feature extractor, along with fine-tuning and Elastic Weight Consolidation (EWC). Li et al. [19] utilize VGG-16 and deep transfer learning to classify dark-toned rocks, fine-grained rocks, and layered sedimentary rocks. Xu et al. [32] investigate Faster R-CNN and YOLOv4 for autonomous lithology assessment. Goh et al. [21] showcase CNN with a contrast learning approach on its SimCLR model to achieve terrain segmentation with fewer labels. Gasperini et al. [33] proposes CloverNet for real-time semantic segmentation. Lv et al. [23], Xiong et al. [22], and Liu et al. [20] implement CNN models with transformer variations for rock segmentation. Additionally, [20] implements a Hybrid Attention Semantic Segmentation (HASS) network, for instance, segmentation in panoramic Mars images. Coloma et al. [34] uses YOLOv5 for detecting rocks in 2D images and generating 3D meshes of the rover surroundings.

Geometry Based Durner et al. [24] introduces a rock instance segmentation method, termed INSTR, utilizing stereo images to generate pixel-wise masks while considering the scene's geometry. This approach enables testing in real-case scenarios for autonomous navigation, sample extraction, and geological analysis.

A. Contributions

The literature focuses mainly on detecting rocks or other terrain features on Mars rather than on the Moon. Our study targets Moon applications, aiming to improve lunar rover capabilities by using Mask-RCNN to detect, segment, and classify natural lunar features like rocks, giving us more details about their shape and size. Additionally, our approach works well in low-light conditions and with different types of cameras. In summary, the key contributions are:

- Agnostic method based on Mask-RCNN and transfer learning to identify and discern Lunar natural features (rock, pebble, and crater), allowing the rover to

recognize these features in unstructured and changing environments, even in varying lighting conditions.
- Novel dataset comprising images from past lunar missions and from a lunar analog facility that simulates lunar polar and equatorial illumination conditions.
- Experiments and evaluation of the transferability, operability, and adaptability across different camera types and environmental scenarios.

III. METHODOLOGY

A. Mask R-CNN

Mask R-CNN [35] was developed for CNN-based feature extraction due to its accurate object detection and instance-level segmentation, which provides a good balance between speed and accuracy. First, it extracts the features from the image using convolutional layers to create a feature map. Subsequently, the Region Proposal Network (RPN) identifies Regions of Interest (RoIs) within this feature map. Mask R-CNN uses RoI Align instead of RoI Pool, used in its predecessor [36]. RoI Align fixes a problem where the features didn't match up perfectly with the original image. Instead of dividing the features into boxes, it smooths things out using continuous coordinates instead of discrete boxes. Following RoI Align, the mask head, composed of two convolutional layers, generates masks for each RoI, enabling precise segmentation of images down to each pixel. Although Mask R-CNN has been employed on the Earth for different use cases on rock detection [37], [38], and [39], its performance in a space mission context, particularly in lunar and martian environments, remains untested.

1) Hyperparameters: The hyperparameters for the proposed Mask-RCNN model were meticulously chosen to optimize its performance during training. The model employs the Stochastic Gradient Descent (SGD) optimizer due to its efficiency in memory allocation. This is a crucial consideration, especially in resource-constrained scenarios like space applications, where a less computationally expensive model will be advantageous. The model's initial learning rate is set at 0.001, with a step decay schedule gradually reducing its training progress, in hand with a learning momentum of 0.9. With a batch size of 5, the model balances training speed and computational feasibility in the hardware. The training was conducted over 100 epochs to allow the model the necessary time to learn from the data without overfitting, all while having a confidence level threshold of 0.9. In addition, both loss weights for mask loss and bounding box loss were set to (1.0), seeking to enhance the quality of the segmentation.

B. Dataset

The implemented model is based on Mask R-CNN by Matterport Inc. [40]. The model was initially pre-trained with Microsoft's COCO dataset, compromising ∼300k images across 91 classes [41]. The COCO dataset wasn't specifically trained to recognize rocks, pebbles, and craters. Therefore, a custom dataset was developed, in hand with transfer learning techniques to fine-tune the COCO model for our specific needs.

The proposed model was trained and tested using a generated dataset captured with the Ricoh Theta V, a 360-degree camera. The images, from the LunaLab facility, University of Luxembourg, have a resolution of 3840 x 1920 px and were taken under different lighting conditions. Three hundred fifty of the five hundred training images were intentionally taken under harsh lighting to mimic lunar polar regions, while the rest represent a fully illuminated scenario. Isabelle Guyon's partitioning method [42] was followed to prevent overfitting during training. Guyon's scaling law for validation-set training ratio (Equation 1) focuses on the unique features to be extracted instead of the number of observations. The fraction of patterns reserved for validation ($g*$) and training ($f*$) is inversely proportional to the square root of the number of families of recognizers (N) over the largest complexity of families or considered recognizers (h_{max}).

$$\frac{f^*}{g^*} = \sqrt{\frac{InN}{h_{max}}} \tag{1}$$

To validate the model, Chang'e 3 Observation 0008 [43] imagery was employed. These images were captured during the Yutu-1 mission on the lunar surface, utilizing the onboard Panoramic Camera (PCAM), which consists of a stereo pair of color cameras. Each snapshot has dimensions of 2352 x 1728 px and covers a Field of View (FoV) of approximately 19.7° (H) x 14.5° (V). Additionally, imagery from the Apollo 11 to 15 missions, taken from the lunar surface, was included for comprehensive validation.

C. Image Pre-processing

Panoramic images taken from the Ricoh Theta V camera were used to generate the dataset. The captured images were resized to dimensions 512 x 256 px [44] and divided into three sub-images, each measuring 256 x 256 px. This process involved first cropping a 256 x 256 px section from the original image's position. Then, moving horizontally to the right 128 px and cropping another 256 x 256 px section. Each resulting images were manually labeled with a polygon shape using VGG Image Annotator (VIA) open-source software [45]. For labeling, three classes were considered:

- **Rocks:** Rocks (19 cm to 65 cm) pose a threat to the rover's path and can potentially cause damage upon collision.
- **Pebbles:** Small rocks (5 cm to 18 cm) that the rover can still go over, but with different speeds and torque.
- **Craters:** Depression on the surface (50 cm to 100 cm in diameter) presents hazards such as potential damage to the rover or disruption of communication if the rover falls into them.

Data augmentation was used on each annotated image to expand the dataset. Each image was flipped horizontally and vertically, rotated 90°, 180° and 270°, and filters like Gaussian Blur were applied. As a result of these augmentation methods, the dataset increased from 500 to 3500 images of 256 x 256 px.

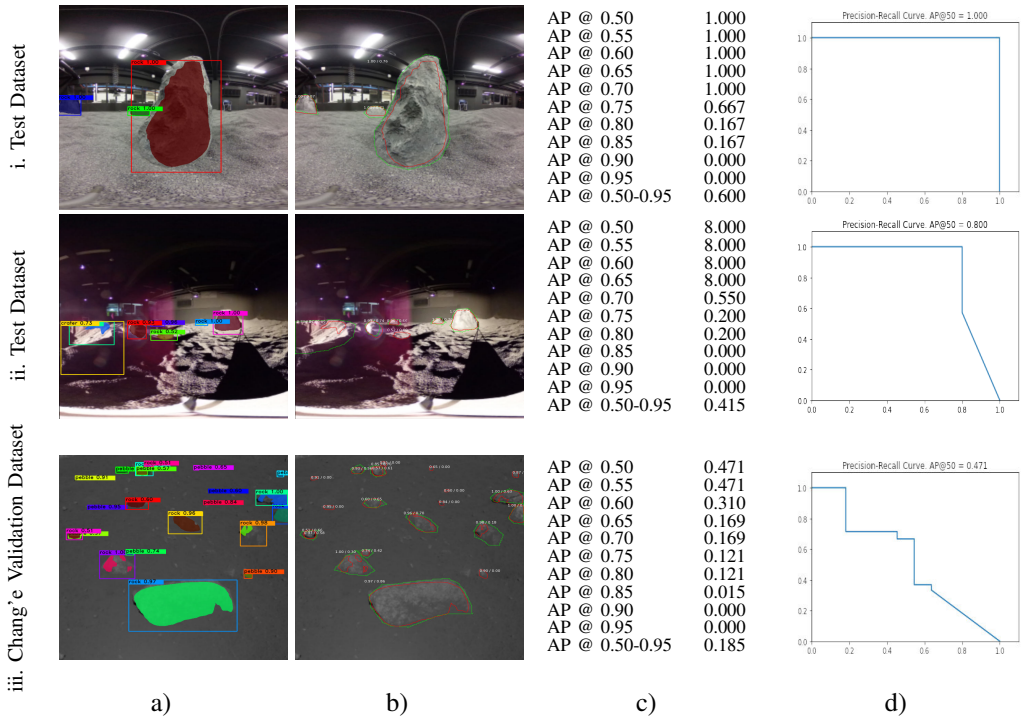

TABLE II: Model performance in a i. fully illuminated image, ii. image with harsh illumination, and iii. image from Chang'e 3. Here we present the model a) Prediction, b) Ground truth in comparison with the detection, c) the average precision ranging from 50 to 95 Intersection over Union threshold, and d) Precision-Recall curve.

D. Training

The model was trained over frozen COCO layers [41] by transfer learning, using NVIDIA Tesla K80 Accelerator. Breaking up Mask R-CNN architecture, the backbone network of choice Feature Pyramid Network (FPN) derived from Residual Network depth 50 layers (ResNet-50-FPN). ResNet 50 has a short training time and delivers good gains in accuracy and speed in feature extraction to generate the feature maps.

Subsequently, the Region Proposal Network (RPN) employs 256 anchors per image to generate several RoI, identifying areas within the image likely to contain the target object and refining their spatial and size parameters. With 256 anchors, the goal is to seek overlapping regions and extract the highest foreground score (Non-max Suppression), discarding the rest.

Afterwards, the algorithm seeks 200 RoI proposed by the RPN and feeds them into a fully connected layer to generate two outputs for each RoI: classes and bounding boxes. Finally, each positive region from the RoI classifier gets converted into a mask through a Mask classifier consisting of two CNNs.

E. Evaluation Metrics

To measure the performance of the model, Average Precision (AP) was utilized, a metric ranging from 0.0 to 1.0 and indicates when the Intersection over Union (IoU) (Equation 2) surpasses a threshold of 50%. If this is the case, the model has achieved a True Positive detection. Average Precision (AP) verifies the model's efficacy in the detection and localization of the classes by consolidating the precision-recall curve into a single scalar value (Equation 3). Precision quantifies the accuracy of detections, while recall gauges the model's ability to capture all TPs among the ground truths. An optimal performance is achieved when the model exhibits high precision and recall simultaneously.

$$IoU = \frac{area(\text{ground truth} \cap \text{prediction})}{area(\text{ground truth} \cup \text{predicion})} \tag{2}$$

$$AP@\alpha = \int_{r=0}^{1} p(r)dr \tag{3}$$

p(r) = measured precision (p) at recall (r).

During the experiment, each detected frame in the output video is evaluated using three parameters:

Accuracy measures the correctness of predictions across all classes. It considers True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) values, as expressed in Equation 4.

$$Accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{4}$$

Average confidence, the collective confidence levels associated with detected features. The sum of each confidence value for each detected class is divided by the number of detected features.

Average elapsed time quantifies the average time taken by the model to process instances within each frame.

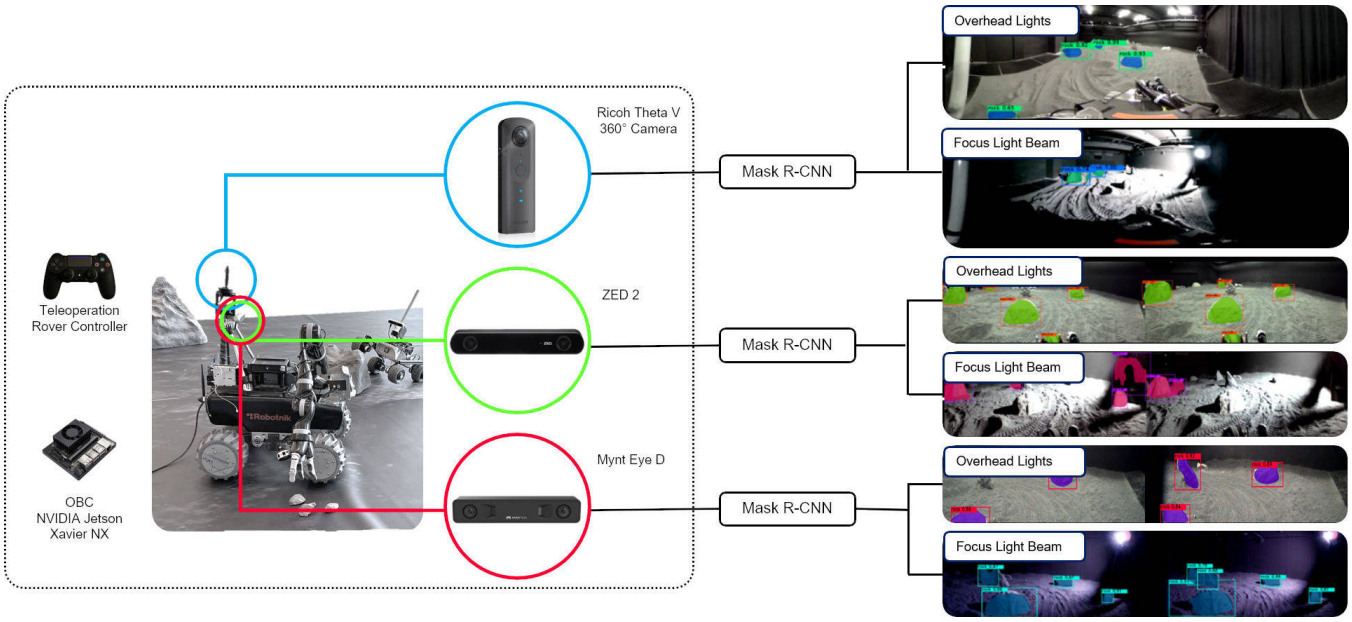

Fig. 2: Experimental Setup

F. Experiments conducted

The experiments are categorized into two groups:

- The model is trained and tested using the generated dataset. Its performance is then validated against real lunar imagery (Section IV. A).
- The model's real-time performance is evaluated with three different cameras and under two light conditions within the LunaLab (Section IV. B).

IV. EXPERIMENTS

A. Test and Validation

The model reliability was evaluated using 127 LunaLab images under full and low light, plus 127 lunar images from different space missions. Each image was previously labeled to obtain quantitative data by comparing the model's predictions with the ground truth. Table II a shows an example of the model prediction when detecting rocks from the test and validation dataset. It provides details like bounding boxes, segmented masks, class IDs, and confidence percentages.

Comparing the predictions' data (red contour) against the labeled data (green contour) on Table II b, the achieved precision-recall curve in a fully illuminated environment is AP@50 of 100% (Table II d). For low illumination, where the harsh lighting environment simulates the lunar polar regions (Table II), the precision-recall curve drops to AP@50 of 80%. On the other hand, the precision-recall curve for lunar images has an AP@50 of 47.1%. Although lower AP is observed in Table II a-iii compared to the ground truth in Table II b-iii, it's noteworthy that the model successfully identified pebbles that were not previously labeled during ground truth annotation. This suggests that, despite the lower AP score, the model's performance surpassed human detection capabilities.

B. Model's Camera and Illumination Agnosticism

1) Experimental environment:

To demonstrate the camera agnosticism of the model, this section evaluates its performance using RGB input images from three different cameras: (1) Ricoh Theta V, (2) ZED 2, and (3) Mynt Eye D. While the Ricoh Theta V is a 360-degree camera, the ZED 2 and Mynt Eye D are stereo cameras. Key distinctions include resolution, FoV, and frame rate.

In the analog lunar environment, eleven rocks (30-65 cm in height, 26-57 cm in width) and six pebbles (5-10 cm) were randomly distributed and kept in the same place throughout the experiment. Additionally, two craters (approx. 50 cm and 1 meter in diameter) were excavated on the surface. The objective was for the model to identify these features using three different cameras and under two lighting conditions while traversing the LunaLab, as shown in Figure 2.

The first lighting condition simulates uniform illumination mimicking the Moon's equatorial regions, achieved through ceiling lamps in the Lunalab. In contrast, the second condition replicates lunar polar regions' lighting, achieved with an LED spotlight tilted at -6 degrees relative to the horizon. This setup casts harsh shadows, challenging the model's detection capabilities.

2) Implementation Setup:

The cameras and the model operate on an NVIDIA Jetson Xavier NX. Utilizing the live video feed from the cameras, the Summit-XL Rover is teleoperated to navigate the environment. A confidence threshold of 65% is set for the model detection in full and partial light conditions, enhancing sensitivity to detect features even in challenging lighting situations. Additionally, to optimize performance, the input frame undergoes cropping to eliminate ceiling elements, minimizing unnecessary noise and simulating a darker horizon for improved image clarity during model execution.

3) Experimental Results:

The qualitative results from sample frames using Mask R-

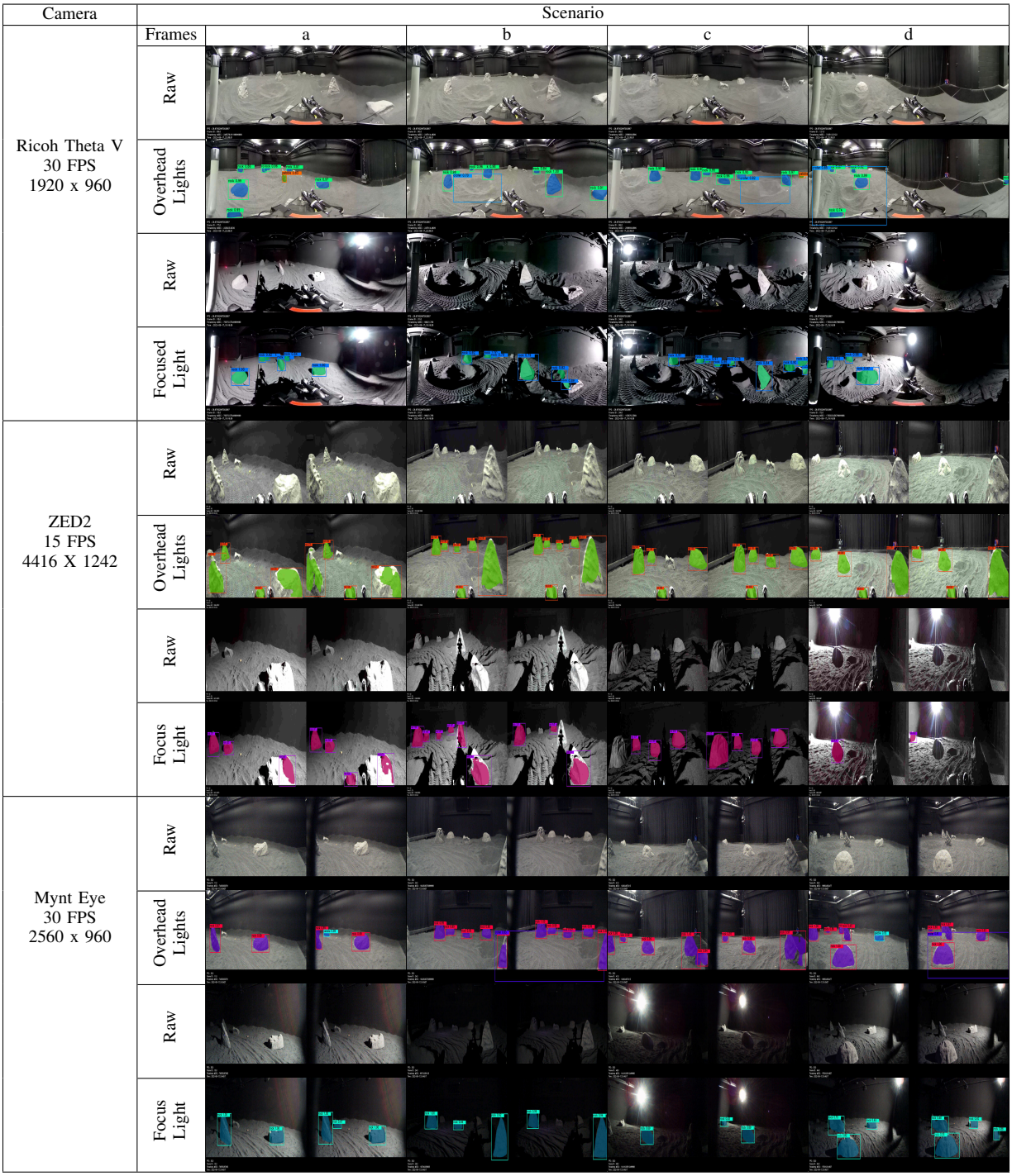

TABLE III: Ricoh Theta V, ZED 2, and Mynt Eye D frames comparison with raw input frame, and frames with overhead, and focus light in the Lunar Analog environment

CNN with our model are presented in Table III, whereas the quantitative analysis is summarized in Table IV. For each camera, two sets of sample frames are included: one captured under fully illuminated conditions and the other illuminated solely by the focus light beam. Within Table III, each of these set are presented in two sample frames showcasing the input and output of the model. The output images feature bounding box, segmented mask, class labels, and the accuracy of instance detection.

Ricoh Theta V

a) Overhead Lights: On average, feature detection accuracy reached 61.0%. Rocks exhibited a notably high mean accuracy of 93.15%, although distant rocks were occasionally misidentified as pebbles. Craters, however, achieved an average accuracy of 77.5%. A noteworthy limitation was the exclusive use of bounding boxes for crater detection,

completely omitting the mask. Moreover, the detection time between frames varied depending on the number of instances detected. Across this scenario, elapsed times ranged from 2.0 to 7.0 seconds, with an average of 3.0 seconds.

b) Focus Lights: The average feature prediction accuracy decreased to 39.46%, primarily due to reduced scene visibility. Notably, neither pebbles nor craters were detected. However, rocks still exhibited an average detection accuracy of 90.31%. The mean elapsed detection time was 2.0 seconds, with a minimum of 1.0 second and a maximum of 2.0 seconds between output frames.

ZED 2

c) Overhead Lights: The average detection of features was 64.58%. Within a one-minute video of the rover traversing the analog lunar environment, only one frame identified a pebble with 68.0% confidence. On the other hand, craters had a mean confidence of 83.8%, while rocks exhibited an average confidence of 92.13%. The average time taken to detect features in each frame was 4.0 seconds. The lowest is 2.0 seconds, and the highest is 6.0 seconds.

d) Focus Lights: The average feature detection rate was 72.40%, the highest among the three cameras and the two light conditions. Rocks were detected with an average confidence of 86.73%, followed by pebbles at 84.0%. No craters were detected. Despite the model's accuracy in detection, the elapsed time between each detected frame was also the highest, averaging 4.0 seconds. The shortest time between frames was 2.0 seconds, while the longest was 17.0 seconds, particularly when no instances were identified.

Mynt Eye D

e) Overhead Lights: On average, the model achieved a feature prediction accuracy of 69.64% under the overhead light condition. This setup yielded the highest confidence in detection across all cameras. The confidence levels for rocks, pebbles, and craters were 97%, 85% and 88%, respectively. The detection time ranged from 1.0 to 4.0 seconds per instance, with an average of 2.0 seconds per frame.

f) Focus Lights: The average feature detection rate was 41.8%. This scene had the lowest confidence levels among the three cameras. Rocks were detected with a confidence of 86.0%, while craters had a confidence of 66.0%. No pebbles were identified. The average detection time per frame was 2.0 seconds, with a minimum of 1.0 second and a maximum of 4.0 seconds.

			Accuracy [%]	Confidence Accuracy [%]			Detection Time per Frame [seconds]		
				rock	pebble	crater	min	max	average
Ricoh Theta V	30 FPS	FI	60.84	93.08	83	79	2	7	2
	1920 *x* 960 px	LED	39.46	90.08			1	2	2
ZED 2	15 FPS	FI	64.58	92.13	68	83.8	2	6	4
	4416 *x* 124 px	LED	72.40	86.73	84		2	17	4
Mynt Eye D	30 FPS	FI	72.41	86.73	84		1	4	2
	256 *x* 960 px	LED	41.80	86.08		66	1	4	2
Average			58.12	90.90	80.15	79.35	1.50	6.67	2.67

TABLE IV: Experimental results

V. DISCUSSION

This paper suggests that the proposed model is camera agnostic. However, it faces challenges in accurately detecting lunar features under specific lighting conditions, particularly when simulating the Moon's polar regions. This issue could be alleviated by employing artificial lights. Additionally, distant rocks may be incorrectly identified as pebbles. The creation of the segmented mask on craters is limited, although it can localize the feature with a bounding box. Overall, rock detection has an average confidence of 90.0% in both lighting conditions, able to distinguish between rocks in non-complex overlapping environments. Pebbles pose a greater challenge, with an average detection confidence of 80.15%. Craters present even greater difficulty in identification; in such cases, only a bounding box is generated without the mask, achieving a mean detection confidence of 79.35%. Despite these challenges, the model maintains an average processing time of 2.0 seconds per frame for feature detection, segmentation, and classification.

VI. CONCLUSION & FUTURE WORK

Developing a high-accuracy method to detect natural features on celestial bodies is crucial for planetary exploration. Although various studies have been conducted on this matter, it has been challenging to find one method that balances high performance and low computational cost. This paper presents a general framework for transfer learning applied to Mask R-CNN to automate rock, pebble, and crater detection on the Moon and Mars, aiding future path-planning efforts.

The proposed model has been implemented onboard the Summit-XL Robotnik rover in a lunar analog environment (LunaLab) under two lighting conditions: (1) the overhead lights of the facility and (2) a focus light beam simulating polar illumination on the Moon. The model effectiveness is evaluated on images from the Chang'e 3 Yutu-1 rover and Apollo 11-15 missions, as well as live stream videos from three onboard cameras during traversability tests in LunaLab.

The described method could enable future lunar and martian rovers to assess their environment and navigate obstacle-free paths autonomously. One major advantage presented in this paper is the model's versatility across different cameras with varying resolutions, FoV, and frame rates, maintaining high accuracy and validation with authentic Moon images.

For future work, the dataset is planned to include rover and lander images obtained from the Moon, as well as incorporating Mars imagery with diverse lighting conditions. It can also be augmented with synthetic data or upcoming open-source datasets. Running complementary algorithms might aid in performance as well as accuracy. Additionally, object tracking can be implemented in the model so that the rover can estimate its position relative to detected objects.

ACKNOWLEDGMENTS

We express our gratitude to Professor Kazuya Yoshida from Tohoku University's SRL for his valuable insights, Assistant Professor Mickael Laine for his constructive critiques, Andrej Orsula from SpaceR, Luxembourg, for his insightful

feedback, and Johan Bertrand for his invaluable guidance and constant support throughout this endeavor.

REFERENCES

[1] S. Coloma, C. Martinez *et al.*, "Enhancing rover teleoperation on the moon with proprioceptive sensors and machine learning techniques," *IEEE Robotics and Automation Letters*, vol. 7, no. 4, 2022.

[2] V. C. Gulick, R. L. Morris *et al.*, "Autonomous image analyses during the 1999 marsokhod rover field test," *Journal of Geophysical Research: Planets*, vol. 106, no. E4, pp. 7745–7763, 2001.

[3] M. Golombek, J. A. Grant *et al.*, "Selection of the mars exploration rover landing sites," *Journal of Geophysical Research: Planets*, vol. 108, no. E12, 2003.

[4] M. P. Golombek, A. Huertas *et al.*, "Size-frequency distributions of rocks on the northern plains of Mars with special reference to Phoenix landing surfaces," *Journal of Geophysical Research E: Planets*, vol. 114, no. 3, pp. 0–09, 3 2008.

[5] M. Golombek, A. Huertas *et al.*, "Detection and characterization of rocks and rock size-frequency distributions at the final four mars science laboratory landing sites," *International Journal of Mars Science and Exploration*, vol. 7, pp. 1–22, 2012.

[6] A. Castano, R. C. Anderson *et al.*, "Intensity-based Rock Detection for Acquiring Onboard Rover Science," in *Lunar and Planetary Science Conference*, 2004.

[7] R. Castano, M. Judd *et al.*, "Current results from a rover science data analysis system," in *2005 IEEE Aerospace Conference*, vol. 2005, 2005, pp. 356–365.

[8] R. Castano, T. Estlin *et al.*, "Onboard autonomous rover science," in *IEEE Aerospace Conference Proceedings*, 2007.

[9] R. Castano, T. Estlin, and R. Anderson, "Oasis: Onboard autonomous science investigation system for opportunistic rover science," *Journal of Field Robotics*, vol. 24, no. 5, pp. 379–397, 2007.

[10] M. C. Burl, D. R. Thompson *et al.*, "ROCKSTER: Onboard rock segmentation through edge regrouping," *Journal of Aerospace Information Systems*, vol. 13, no. 8, pp. 329–342, 3 2016.

[11] X. Gong and J. Liu, "Rock detection via superpixel graph cuts," in *Proceedings - International Conference on Image Processing, ICIP*, 2012, pp. 2149–2152.

[12] X. Xiao, M. Yao *et al.*, "A Kernel-Based Multi-Featured Rock Modeling and Detection Framework for a Mars Rover," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 7, pp. 3335–3344, 7 2023.

[13] X. Xiao, H. Cui *et al.*, "Autonomous rock detection on mars through region contrast," *Advances in Space Research*, vol. 60, no. 3, pp. 626–635, 8 2017.

[14] D. R. Thompson and R. Castaño, "Performance comparison of rock detection algorithms for autonomous planetary geology," in *IEEE Aerospace Conference Proceedings*. IEEE, 2007.

[15] A. Kalra and A. Bhavsar, "Lunar Rock Classification Using Machine Learning," pp. 1–5, 7 2020.

[16] V. Gor, E. Mjolsness *et al.*, "Autonomous rock detection for Mars terrain," in *AIAA Space 2001 Conference and Exposition*. Reston, Virigina: American Institute of Aeronautics and Astronautics, 8 2001.

[17] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1. IEEE, 2001, pp. 1–9.

[18] Y. Suebe, K. Nagaoka, and K. Yoshida, "Terrain Classification With An Omni-Directional Camera Using Convolutional Neural Networks," in *International Symposium on Artificial Intelligence, Robotics and Automation in Space (i-SAIRAS 2018)*, 2018, pp. 1–7.

[19] J. Li, L. Zhang *et al.*, "Autonomous Martian rock image classification based on transfer deep learning methods," *Earth Science Informatics*, vol. 13, no. 3, pp. 951–963, 9 2020.

[20] H. Liu, M. Yao *et al.*, "A hybrid attention semantic segmentation network for unstructured terrain on Mars," *Acta Astronautica*, vol. 204, pp. 492–499, 3 2023.

[21] E. Goh, J. Chen, and B. Wilson, "Mars Terrain Segmentation with Less Labels," *IEEE Aerospace Conference Proceedings*, vol. 2022-March, 2 2022.

[22] Y. Xiong, X. Xiao *et al.*, "MarsFormer: Martian Rock Semantic Segmentation with Transformer," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, 2023.

[23] W. Lv, L. Wei *et al.*, "MarsNet: Automated Rock Segmentation With Transformers for Tianwen-1 Mission," *IEEE Geoscience and Remote Sensing Letters*, vol. 20, 2023.

[24] M. Durner, W. Boerdijk *et al.*, "Autonomous Rock Instance Segmentation for Extra-Terrestrial Robotic Missions," *IEEE Aerospace Conference Proceedings*, vol. 2023-March, 2023.

[25] Y. Cheng, A. E. Johnson *et al.*, "Passive Imaging Based Hazard Avoidance for Spacecraft Safe Landing," in *Proceeding of the 6th International Symposium on Artificial Intelligence and Robotics & Automation in Space*. St-Hubert, 2001.

[26] Y. Cheng, Y. Cheng *et al.*, "Optical landmark detection for spacecraft navigation," *In Proceedings of the 13th Annual AAS/AIAA Space Mechanics Meeting*, 2003.

[27] A. Huertas, Y. Cheng, and R. Madison, "Passive imaging based multicue hazard detection for spacecraft safe landing," in *IEEE Aerospace Conference Proceedings*, vol. 2006, 2006.

[28] A. Huertas, Y. Cheng, and L. H. Matthies, "Real-time Hazard Detection for Landers," in *Proc. NASA Science Technology Conf.*, Adelphi, MD, 5 2007.

[29] L. Matthies, A. Huertas *et al.*, "Landing hazard detection with stereo vision and shadow analysis," in *Collection of Technical Papers - 2007 AIAA InfoTech at Aerospace Conference*, vol. 2. American Institute of Aeronautics and Astronautics Inc., 2007, pp. 1284–1293.

[30] Y. Li and B. Wu, "Analysis of Rock Abundance on Lunar Surface From Orbital and Descent Images Using Automatic Rock Detection," *Journal of Geophysical Research: Planets*, vol. 123, no. 5, pp. 1061–1088, 5 2018.

[31] K. Di, Z. Yue *et al.*, "Automated rock detection and shape analysis from mars rover imagery and 3D point cloud data," *Journal of Earth Science*, vol. 24, no. 1, pp. 125–135, 2 2013.

[32] Z. Xu, W. Ma *et al.*, "Deep learning of rock images for intelligent lithology identification," *Computers & Geosciences*, vol. 154, p. 104799, 9 2021.

[33] D. Gasperini, W. Masawat *et al.*, "CloverNet: A Real-Time Network for Semantic Segmentation Onboard Edge Devices Towards Planetary Exploration," *9th 2023 International Conference on Control, Decision and Information Technologies, CoDIT 2023*, pp. 333–338, 2023.

[34] S. Coloma, A. Frantz *et al.*, "Immersive rover control and obstacle detection based on extended reality and artificial intelligence," *arXiv preprint arXiv:2404.14095*, 2024.

[35] K. He, G. Gkioxari *et al.*, "Mask R-CNN," *Computing Research Repository (CoRR)*, pp. 1–12, 1 2018.

[36] S. Ren, K. He *et al.*, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 1–14, 6 2015.

[37] Z. Chen, T. R. Scott *et al.*, "Geomorphological Analysis Using Unpiloted Aircraft Systems, Structure from Motion, and Deep Learning," 9 2019.

[38] P. Loncomilla, P. Samtani, and J. Ruiz-del Solar, "Detecting rocks in challenging mining environments using convolutional neural networks and ellipses as an alternative to bounding boxes," *Expert Systems with Applications*, vol. 194, p. 116537, 5 2022.

[39] D. Yang, X. Wang *et al.*, "A Mask R-CNN based particle identification for quantitative shape evaluation of granular materials," *Powder Technology*, vol. 392, pp. 296–305, 11 2021.

[40] W. Abdulla, "GitHub - matterport/Mask_RCNN: Mask R-CNN for object detection and instance segmentation on Keras and TensorFlow," 2017. [Online]. Available: https://github.com/matterport/Mask_RCNN

[41] T. Y. Lin, M. Maire *et al.*, "Microsoft COCO: Common objects in context," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 8693 LNCS, no. PART 5. Springer Verlag, 2014.

[42] I. Guyon, "A scaling law for the validation-set training-set size ratio," AT&T Bell Laboratories, Tech. Rep., 1997.

[43] The Planetary Society, "Chang'e 3 data: Rover Panoramic Camera (PCAM)," 1 2016. [Online]. Available: https://planetary.s3.amazonaws.com/data/change3/pcam.html

[44] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[45] A. Dutta and A. Zisserman, "The VIA Annotation Software for Images, Audio and Video," in *Proceedings of the 27th ACM International Conference on Multimedia*. New York, NY, USA: ACM, 10 2019, pp. 2276–2279.