Research Article

Overview of the CHILL-ICE 2021 Science Experiments and Research Campaign

David J. K. Smith, Charlotte R. Pouwels, Marc Heemskerk, Benedetta M. Cattani, Esther Konijnenberg, Robert Heemskerk, and Sebastian Ogalde

1ICEE Space, Netherlands
2EuroMoonMars, An Initiative of ILEWG (International Lunar Exploration Working Group, The Netherlands), Netherlands
3Vrije Universiteit Amsterdam, Netherlands
4Lunar Zebro, Delft University of Technology, Netherlands
5Heemskerk Innovative Technology, Netherlands

Correspondence should be addressed to David J. K. Smith; djksmith95@gmail.com

Received 17 November 2021; Accepted 29 September 2022; Published 12 November 2022

Copyright © 2022 David J. K. Smith et al. Exclusive Licensee Beijing Institute of Technology Press. Distributed under a Creative Commons Attribution License (CC BY 4.0).

The main objective of the CHILL-ICE (Construction of a Habitat Inside a Lunar-analogue Lava tube—Iceland; a campaign by ICEE Space and EuroMoonMars) prototype mission was to deploy an inflatable habitat and its systems inside a lunar analogue lava tube. This took place during an 8-hour extra vehicular activity (EVA) with three analogue astronauts as part of a three-day mission. CHILL-ICE 2021 was carried out in July/August 2021 and consisted of two missions and was accomplished through successful collaboration of nonprofit research organizations, agencies, companies, and universities across 16 nations. The pilot campaign successfully reached its main objective: the testing of emergency equipment designed to help astronauts survive when first arriving to a new celestial body and to perform experiments similar to those that would be carried out off-planet. This article is a review of the scientific research experiments carried out during and after the mission: SpotNet, an artificial intelligence (AI) astronaut detection vision system; training for studies of the geological surroundings examined during EVAs; astronaut vigilance experiments carried out before, during, and after the mission; and Lunar Zebro, a legged rover intended to assist the crew in traversing and exploring harsh terrain.

1. Introduction

Analogue missions are an established way to test equipment, operations, and procedures that are planned to be used in spaceflight missions, to reduce the risk of failures or problems that may result in mission failure or fatalities. The analogue setting is aimed at emulating certain aspects that characterize human space missions, which is why they are usually set in ICEE (isolated, confined, and extreme) environments, to better recreate the harsh extraterrestrial conditions. Crews of so-called analogue astronauts are often involved, to be able to include the human factor in the testing of technologies or practices.

1.1. Mission Overview. This paper is an extension of the “Mission Overview: Construction of a Habitat Inside a Lunar-analogue Lava tube—Iceland” [1]. The CHILL-ICE mission was carried out in Entrance 10 of the Stefánshellir lava tube, in Iceland from 26th July to 8th August, 2021. Within this framework, a team of international young professionals and students coming from 16 countries performed two lunar analogue missions. The main objective of the CHILL-ICE campaign was to deploy an inflatable habitat and its systems inside a lunar-analogue lava tube (Figure 1). To ensure that the mission idea was indeed redeployable and a portable analogue mission is not simply doable, but also repeatable, two short analogue missions to test basic operations and give a proof of concept were held during the campaign (Figure 1).

1.2. Mission Location. For decades, it has been theorized that lava tubes, which are naturally formed volcanic “cave”
systems, could serve as an additional shelter for human exploration and house a potential settlement on the Moon and Mars [2]. They would ensure enhanced protection against meteorites, galactic cosmic rays (GCRs) and solar energetic particle (SEP) radiation, regolith, and temperature fluctuations [3]. Tubes with the same kind of geological origin are also present on Earth and, in particular, in Iceland. Concerning CHILL-ICE, this led to the selection of Entrance 10 to the Stefánshellir lava tube system in the Hallmundarhraun lava field, located near Reykholt in Vesturland, in the west of Iceland. The selection rationale can be explained in detail here [4]; however, generally, this section of the Stefánshellir lava tube offers a large, well-lit entrance, and upon entry has a wide, flat surface suitable for an inflatable habitat. The temperature inside the lava tube generally ranges from $7^\circ$ during the day, dropping down to $3^\circ$ at night—although this is dependent on the depth into the lava tube, getting closer to a stable $5^\circ$ the further away from the nearest skylight. The relative humidity inside the lava tube was measured to be stable around 69%, which produced a constant effect from the ceiling of the tube. This resulted in surface wetness inside the ECHO habitat [5], which proved to be uncomfortable for the astronauts and dangerous for some of the instruments and electronic devices. Even though this speaks against a proper analogue environment—the Moon and Mars are much drier—the geochemistry of the Hallmundarhraun lava field, in which Stefánshellir is located, is quite similar to the mare basalts of the Moon.

Besides being a fitting geochemical analogue, the mineralogy appears also similar to spectroscopic data from the lunar surface as well, and the whole Hallmundarhraun lava field is relatively young. With an estimated age of just over 1000 years, the insides and ceiling are only mildly weathered from natural processes, giving them a greater structural stability. Simultaneously, the lava tube is located quite close to a road, which makes that the most fragile parts of this cave have been damaged by tourists already, in turn making it easier for the analogue astronauts to roam around without damaging these unique and protected environments more. Lastly, the easy access to the field and the nearby located mountain Strutur with a 4G Internet antenna also allowed quicker communication and reaction times of emergency response teams, in case it would have been necessary.

## 2. Methodologies

The CHILL-ICE missions have been a platform for extensive interdisciplinary research, with the most relevant of these described in this paper.

### 2.1. SpotNet: Astronaut Detection Vision System

Maintaining accurate and up-to-date knowledge of astronaut location during an analogue simulation mission is essential information for both the analogue astronauts in the field and support staff at the mission control centre (MCC). In the current CHILL-ICE protocols, location information is relayed manually, via radio or Internet communication, which can be inaccurate and cause confusion due to the lack of orientation points or data/radio link quality.

Furthermore, MCC monitors and receives this information over a rolling 24-hour basis for the duration of the simulation. This can often be impractical, as MCC staff need to rotate shifts overnight, meaning sleeping patterns are disrupted which risks performance and vigilance degradation. This shows that the main weakness of the current protocol is reliability and the need of constant human support. There are two main improvements that can address these issues: on-site camera surveillance (hardware-oriented) and automated monitoring (software-oriented). Setting up cameras around the habitat and surrounding environment, or inside suit helmets with a video feed linking to MCC, would solve the hardware aspect of the problem. These hardware improvements are currently being developed for the next CHILL-ICE mission. SpotNet focuses on the software side: delivering an automated way of monitoring the captured footage. Recent advances in the field of machine learning have created computer vision programs well-suited to the task of identifying objects in images, which SpotNet uses to automate astronaut detection.

Machine learning algorithms are computer programs which are able to autonomously detect patterns in data [6]. A training dataset is required to “train” these algorithms to understand the patterns present in the data. For this work, images were collected from the wide array of footage captured during CHILL-ICE activities, including simulation test training, team training exercises, and the mission itself. A variety of footage was collected to have a diversity of depth, background, perspective, presence of nonastronauts and lighting conditions. The final collated dataset was an assortment of image and video data, which was split into individual frames and manually annotated with bounding boxes to highlight the location of astronauts within the scenes. A selection of these frames can be found in Figure 2.

This dataset was then fed into our object detection model. The YOLO (you only look once) is a state-of-the-art family of detection models, and a variation (YOLOv5) of its architecture is used within SpotNet [7]. The ultralytics Python implementation of YOLOv5 was used for this project [8]. YOLOv5 is a single-stage detector which uses a convolutional neural network (CNN) as a feature extractor backbone. The CNN backbone uses sequential layers of filters which extract low-level features and patterns from the image before expanding to larger features. These features are then stacked in feature pyramids and partitioned into boxes. These boxes are then classified with a probability of containing the target object, with the highest probability becoming the model prediction. A simplified version of YOLOv5 can be visualised in Figure 3.

The models were trained for 100 epochs on 356 images, with 233 used for training and 67 used for validation using a NVIDIA Tesla K80 GPU provided by Google Colab. The remaining 56 images were kept for testing the trained model. A batch size of 16 was used, and the learning rate was set at 0.01, with a stochastic gradient descent optimizer. The model did not use pretrained weights and was trained from scratch on the astronaut dataset, with the original YOLOv5 architecture used without modifications.
The model will be evaluated on the level of overlap between the predicted bounding box and the annotation, defined as the mean intersection of union (mIoU). An mIoU of 1.0 means an overlap of 100%. Generally, a threshold of 0.5 is used to determine whether the model has made a successful detection or not [10]. We grade our model both on mIoU with a threshold of 0.5 (mIoU@0.5) and over an average of increasing thresholds from 0.5 to 0.95 (mIoU@0.5:0.95). We also use precision and recall: model precision means the percentage of correct to total detections made by the model. Recall means the percentage of correct detections to total number of existing targets in the dataset. We also measure the prediction speed of the model based on the number of frames per second (FPS) processed during testing, which will be a critical factor when assessing feasibility of using the model with real-time footage.

2.2. EVA Geological Research. As part of the mission training, all analogue astronauts received a course in geological sampling and survey methods. A key aspect was the protection of the natural environment, important for future preservation of lunar and Martian lava tubes, but also for limiting the impact of the mission on the local environment of the protected lava field. The lava tubes in Iceland contain lava.
droplets and lava straws, which are stalactite-like and often ceiling-bound structures that solely form during or directly after the formation of the lava tube itself, making them highly unique and to be protected, being both scientifically and culturally significant.

The lava droplets are formed by still-liquid lava with a high viscosity that flows at rapid speeds through the lava tube, “splashing” to the ceiling, where they start dripping downwards and crystallize while doing so, contributing to the final shape and inner tephrostratigraphy of the lava tubes. Lava straws are formed by slow degassing shortly after the formation of the tube. These gasses penetrate through the still-liquid insides and can either grow downwards from the ceiling or act as a mud-volcano-like stalagmite formation. The two can be differentiated by their difference in size, abundance, and in the case of the lava straws, their hollow insides. Seeing that these formations contain information about the environmental conditions, geophysical parameters such as viscosity and with that, magma temperature, and the evolution of the magma reservoir, they can be highly interesting for geological and geochemical investigations. The key research aspect was not just the sampling and geological surveying itself but also the writing and setting up of protocols and an effective experiment training manual. The samples will be analyzed specifically on the major and minor elemental compositions of larger rock samples found in and around the Stefánshellir lava tube system, and mineralogical compositions will be based on microscopic observations of thin sections carried out postmission.

During the EVAs, the astronauts were instructed to find and picture small hand-sized samples that showed a distinct feature or contained a specific color that could directly lead back to a certain part of the lava tube. This way, the astronauts could map a specific location of the inside of the lava tube, without a directly destructive sampling method on the walls or ceiling. This can be visualised with some images captured during one of the EVAs as shown in Figure 4.

When a sample was depicted as being “potentially interesting,” a large range of photos were taken, as well as a description of the sample, its direct surroundings, and location in the cave. This would later be reconciled with an approximate GPS location. Like on the lunar surface, the insides of the lava tube are completely void of outside signals and electromagnetic radiation, meaning that no direct GPS location could be found. After returning to the ECHO habitat, the pictures and descriptions were sent to “Earth,” where the remote support (RS) and MCC geological support crew would advise on which samples to take. The RS team would then create a 3D approximated model from the pictures taken, both with a GoPro HERO9 Black and a Vuze XR 360° camera set to the stereo 180° picture mode.

2.3. Vigilance. To ensure crew safety and mission success, the astronauts have to be constantly vigilant during many activities, such as EVAs, equipment inspections, and monitoring habitat air quality. Before, during, and after the CHILL-ICE missions, the astronauts were asked to participate in a vigilance experiment: an area of research that aims to quantify the state of being alert to rare and hard-to-distinguish events. The vigilance research is a pilot study aiming to highlight the importance of it for long-term space missions.

The development of astronaut vigilance is tracked with a Mackworth clock test. This test consists of a participant monitoring a circular shape of dots, with one dot lighting up after another. The participant needs to respond when the light skips a dot, known as an event. The original clock test, as used by [11], took up to 2 hours to measure the deterioration in vigilance and the effectiveness of changing watch. For this research, deterioration of vigilance is measured over the 3 days, and therefore, the participants will only do shorter 15-minute tests, but in four days in a row, first on the evening before the start of the simulation mission, then on the first and second evening of the mission, and lastly in the evening after the simulation mission has ended, and they are back at mission control. The clock test used for this experiment was taken from the testing suite available in the Psychology Experiment Building Language (PEBL) application [12].

After instructions, the main researcher did two to three one-minute practice tests, which had a high number of events to ensure the participants understood how to respond. Next, a pretask Dundee Stress State Questionnaire (DSSQ) [13] was filled in. For the test itself, the 6 participants (3 females and 3 males, ages 24-38) performed the clock test with an event probability of 0.013 (i.e., for each iteration of the moving light, the probability of an event that will occur is 1.3%) and then filled in the posttask DSSQ. During the four days, the participant would do the pretask DSSQ, the 15-minute clock test, and the
posttask DSSQ, at the same time each day. The performance score is based on the hit rate and the false alarms during the clock test and is analyzed with the DSSQ data (scoring engagement, distress, and worry on a 0-32 scale) along with the observations of the experiment circumstances by the researchers. Due to the small test size and the expected differences in the course of the two missions, no statistical analysis can be performed. However, the results are examined for a trend to be investigated in the next CHILL-ICE missions.

2.4. Lunar Zebro Rover. While various rovers have already been sent to the lunar surface, no robotic or human exploration has ever been conducted in the Moon’s underground lava tubes. In such an unknown and challenging environment, smaller and cheaper rovers may provide a more robust and reliable alternative to the single, big, and expensive rover, typically used in space exploration missions. Moreover, a swarm of rovers would allow for faster exploration, could deploy a network of sensors, and would also be particularly reliable, due to its redundancy. This is why Lunar Zebro, a student-led team based at Delft University of Technology, has been working on a swarm of C-shaped legged miniaturized rovers. The CHILL-ICE missions provided the perfect opportunity to test the rover’s locomotion in the Icelandic lava tubes and to investigate if and how it is possible for astronauts to manually operate the rover, after receiving a basic training.

While Lunar Zebro has been designing and developing multiple rovers, the model that was featured in the CHILL-ICE mission is the one described in [14]. Its main features are six C-shaped legs and an embedded camera. The rover weights around 1 kg and its dimensions are 40 × 30 × 20 cm (Figure 5).

The rover testing was performed by the astronauts during EVAs. The test plan included two phases: operating the rover via direct visual monitoring and operating it via smartphone app using footage live-streamed from a GoPro HERO 9 mounted on the rover. During the testing, the rover had to climb both uphill, downhill, forward, and sideways, to estimate the maximum steepness it can overcome without falling. It also had to overcome obstacles of different size, to assess its capabilities in this regard. The test sequence was noted by the astronauts, which were also asked to record it using the embedded camera of the rover. The test results and the feedback of the astronauts were gathered through a questionnaire.

3. Results

In this section, the results can be found of the CHILL-ICE missions regarding SpotNet, geology study, vigilance, and the Lunar Zebro rover study.

3.1. SpotNet: Astronaut Detection. We present the training results for a range of YOLOv5 architectures here, containing metrics of mIoU, precision (P), recall (R), detection frames per second (FPS), and model size.

<table>
<thead>
<tr>
<th>YOLOv5s</th>
<th>mIoU@0.5</th>
<th>mIoU@0.5:0.95</th>
<th>P</th>
<th>R</th>
<th>FPS</th>
<th>Model size</th>
</tr>
</thead>
<tbody>
<tr>
<td>YOLOv5s</td>
<td>0.98</td>
<td>0.65</td>
<td>0.933</td>
<td>0.97</td>
<td>66</td>
<td>14.1 MB</td>
</tr>
<tr>
<td>YOLOv5m</td>
<td>0.995</td>
<td>0.668</td>
<td>1</td>
<td>0.97</td>
<td>37</td>
<td>40.1 MB</td>
</tr>
<tr>
<td>YOLOv5l</td>
<td>0.97</td>
<td>0.633</td>
<td>0.98</td>
<td>0.93</td>
<td>21</td>
<td>90.2 MB</td>
</tr>
</tbody>
</table>

Figure 5: Lunar Zebro rover in the Stefánshellir lava tube and being inspected by one of the CHILL-ICE analog astronauts. Photographed by Jamal Ageli (http://www.jamal-ageli.com/).
Hallmundarhraun lava field, was taken. Two on-site geologists explained the formation of lava tubes, geomorphological features from the field, inside the lava tubes, and on the inner walls of the skylights, and a general overview of the geological history was given. Lava tube safety procedures and sampling methods were presented to the analogue astronauts. As for some of the astronauts, this was their first experience with speleological exploration; the first descent was performed with supervision from the Icelandic search and rescue team. During the afternoon and evening hikes, pop-quiz style questions were asked to the analogue astronauts to test their knowledge. Although many of the safety procedures were clear, some of the sampling methods—specifically regarding the potential location of the samples—were less clear.

During the analogue missions, a total of 15 geological samples, combining a total weight of 1.875 kg, were taken by the analogue astronauts, from 11 different locations (see Table 2). They selected these samples based on how the rocks looked from the outside, by intersections made by hammer, and digital analyses by the geologists on MCC and RS. The rock samples were then exported to the Netherlands with a permit from the Icelandic Institute of Natural History after the mission and to be analyzed at the Vrije Universiteit in Amsterdam. X-ray fluorescence and optical microscopy of thin sections will be performed to get a closer understanding of the elemental and mineralogical compositions of the rocks.

As seen above, 11 sample bags were collected with a total mass of 1.875 kg. 11 of these were acquired inside Stefánshellir, one from outside, and 2 from inside Surtshellir. The majority required size reduction postmission.

3.3. Vigilance. Due to difficulties in planning and the extreme environmental conditions, two participants were not able to complete the experiment on all four days. Their data is therefore removed from analysis. Despite the small test size ($n=4$, two in each mission), the results show a notable difference in performance between the two missions. In Figure 7, the participants’ hit rates in the clock test are plotted over the course of their missions and show that the participants in the first mission have little to no decrease of vigilance over the four experiment days. However, the participants in the second mission show a large deterioration of vigilance during their mission, before partially recovering postmission.

An analysis of the DSSQ data shows that the participant 1 clock test results of mission 1 matched the trend of their DSSQ scores, scoring lower on engagement and elevated on distress and worry the 2nd and 4th days. However, participant 2 did not score with a similar trend as their DSSQ was more irregular. For mission 2, participant 1 scored the clock test with a similar trend as their DSSQ data: simultaneous shifts in both pre- and posttask engagement (-3), distress (+2), and worry (+2) scores during the second and third tests compared to the other two days. Participant 2 however rated only days 2 and 4 significantly less engaged and more worried and distressed than days 1 and 3. Regarding the circumstances during the second mission—the habitat was leaking which required long EVA’s and pushed the
time of the experiment to midnight—the lower clock test scores on days 2 and 3 would not be totally unexpected. However, their level of vigilance cannot be estimated from their DSSQ’s as for 2 out of 4 participants this did not match. There was also a strong contrast between the pre- and post-task questionnaires. Although some scores shifted upwards while others shifted downwards, it appears that the task was quite demanding during the mission. On average, the engagement lowered with 5/32, and distress and worry elevated with 2/32 and 6/32, respectively, these days in only 15 minutes, significantly higher shifts than the shifts found between the days.

3.4. Lunar Zebro Rover. Within CHILL-ICE, the Zebro rover was operated by the analogue astronauts for three half an hour shifts, during the first of the two missions that took place. This allowed to assess that the estimated steepness that rover can climb, both forward and sideways, is between 40° and 50°. On a scale from 1 to 5, the rover locomotion was ranked 3 by almost all the astronauts, signifying that the rover can successfully move, but not on all kinds of terrains. In particular, due to its small size, it struggled with steep climbs and larger obstacles. Rocks higher than 5 cm and holes around 25 cm wide and 10 cm deep posed a challenge to the rover locomotion. Operating the rover through the live-stream yielded a successful result. The maximum distance at which the live-stream would work was also tested: this resulted to be approximately 3 m. Overall, it was always possible to continuously operate the rover, unless one of its legs would get stuck due to the movement on uneven terrain. In that case, it was necessary to manually turn it off and on again.

Some drawbacks of operating the rover through the camera were that the field of view in both directions was reduced and that the vision was slightly occluded by the rover’s body. The astronauts’ feedback included possible enhancement of the rover chassis, legs, and operational interface.

The main comparison found from the rover experimentation that can be made to lunar operation is the similarity of

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Weight (g)</th>
<th>Image ID</th>
<th>Latitude/longitude</th>
<th>Reduced</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>200</td>
<td>2</td>
<td>64°47'15.86&quot;N, 20°42'43.88&quot;W</td>
<td>Y</td>
</tr>
<tr>
<td>2</td>
<td>350</td>
<td>2</td>
<td>64°47'15.41&quot;N, 20°42'48.33&quot;W</td>
<td>Y</td>
</tr>
<tr>
<td>3</td>
<td>300</td>
<td>2</td>
<td>64°47'13.80&quot;N, 20°42'51.24&quot;W</td>
<td>Y</td>
</tr>
<tr>
<td>4</td>
<td>200</td>
<td>1</td>
<td>64°47'13.72&quot;N, 20°42'51.12&quot;W</td>
<td>Y</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>1</td>
<td>64°47'13.72&quot;N, 20°42'51.12&quot;W</td>
<td>Y</td>
</tr>
<tr>
<td>4</td>
<td>50</td>
<td>2</td>
<td>64°47'13.72&quot;N, 20°42'51.12&quot;W</td>
<td>Y</td>
</tr>
<tr>
<td>5</td>
<td>175</td>
<td>2</td>
<td>64°47'13.84&quot;N, 20°42'52.83&quot;W</td>
<td>N</td>
</tr>
<tr>
<td>6</td>
<td>150</td>
<td>2</td>
<td>64°47'12.76&quot;N, 20°42'56.29&quot;W</td>
<td>Y</td>
</tr>
<tr>
<td>7</td>
<td>150</td>
<td>2</td>
<td>64°47'12.76&quot;N, 20°42'56.29&quot;W</td>
<td>N</td>
</tr>
<tr>
<td>8</td>
<td>25</td>
<td>1</td>
<td>64°47'13.78&quot;N, 20°42'53.31&quot;W</td>
<td>Y</td>
</tr>
<tr>
<td>8</td>
<td>125</td>
<td>3</td>
<td>64°47'13.78&quot;N, 20°42'53.31&quot;W</td>
<td>Y</td>
</tr>
<tr>
<td>9</td>
<td>25</td>
<td>2</td>
<td>Unknown, close to location of 3</td>
<td>Y</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>1</td>
<td>64°47'13.01&quot;N, 20°42'47.67&quot;W</td>
<td>N</td>
</tr>
<tr>
<td>10</td>
<td>25</td>
<td>2</td>
<td>64°47'1.79&quot;N, 20°43'28.28&quot;W</td>
<td>Y</td>
</tr>
<tr>
<td>11</td>
<td>25</td>
<td>1</td>
<td>64°47'1.79&quot;N, 20°43'28.28&quot;W</td>
<td>Y</td>
</tr>
</tbody>
</table>

Figure 7: Clock test results of the CHILL-ICE missions. Credits: Robert Heemskerk.
the rockiness and steepness of the terrain type. The wetness in the lava tube and the composition of Earth’s atmosphere and gravity field were limited any further comparison.

4. Discussion

This section contains a summary of our results and a quick discussion on their limitations and implications for future extensions of the research in analogue environments.

4.1. SpotNet: Astronaut Detection. The results in Table 1 show we can achieve accurate detection of analogue astronauts. This is shown by the near 100% detection rates when mIoU is thresholded at 0.5, the widely accepted level for a successful detection [10]. There is also good performance when the threshold is incrementally increased to 0.95. Similarly, precision and recall are close to 100% for all models, meaning they are able to be precise in the predictions they make, as well as capturing the majority of correct examples. Interestingly, we see a drop off in accuracy between YOLOv5m and YOLOv5l, the largest model, which is likely due to overfitting. The model with the highest mIoU is YOLOv5m. It also demonstrates a modest storage requirement and adequate FPS levels. A key feature of the final SpotNet system will be real-time astronaut detection, meaning the FPS capability of the model is important. FPS decreases as the model gets larger and requires more computational power. This is a standard trade-off for object detection models, as discussed here [15]. YOLOv5m exhibits real-time detection speeds while maintaining the highest accuracy level. Therefore, YOLOv5m was chosen as the optimal architecture for SpotNet.

We still see some variation in YOLOv5m’s performance, as shown in Figure 6, which drops to 55% for one of the examples. This is likely due to the available training data, which greatly influences the generalisation capabilities of the models. For instance, there was a low number of training images with a rocky background, meaning the model was less confident when predicting over testing images with a similar appearance. Similarly, all the models were partially trained on video data split into frames. Some frames in the test subset, i.e., images meant to be new to the model, were taken from the same video as frames in the training subset, meaning they looked highly similar. This means that the model was exposed to data almost identical to that in the testing set, making its confidence less reliable for these examples. Therefore, a key next step of the project is to improve the quality and quantity of the training data, to make a more robust and generalisable model.

4.2. Geological Training and Sampling. Due to limited time on each EVA, small amounts of training, and the focus on having the least destructive impact on the environments, a very small set of samples were taken. Due to the PI retreating from this study, thus far no XRF or microscopic thin section analysis has been done but should be completed by CHILL-ICE mission researchers by the end of 2022. The location of almost all of the samples was taken properly, but not according to a standard procedure. For future analogue missions, longer trainings and more focus on sampling methods should be included in the premission protocols. In future geological research, having an infrared (IR) spectrometer on-site would be recommended. This can help with on-site geochemical analyses and determining which rocks are distinct from each other and of interest for further analysis. By using an IR spectrometer, different lava flows can be distinguished in the lava tube system by their geochemical content. This will also provide the analogue astronauts and MCC with information on the history of the lava field and how the volcanic system evolved through time.

4.3. Vigilance. With only four participants not much can be statistically proven. However, this pilot study does show an alarming trend in the deterioration of vigilance of the analogue astronauts. To prove this trend, future studies should aim to collect more data, especially over longer periods of time. Another result found was the inconsistency to their DSSQ scores that aims to link the participants’ feelings of engagement, distress, and worry to their state of vigilance. As the pre- and posttask questionnaires showed these feelings responded to the 15-minute experiment much more than between the experiment days, it could be concluded that these questionnaires are not suitable for this study. Other metrics need to be researched that might correlate to the state of vigilance. These metrics, for example, physical and mental energy levels, could be tested in future studies, such as CHILL-ICE II (the CHILL-ICE campaign planned for 2022). With more knowledge on how and when vigilance deteriorates, such research could help increase safety for future space missions.

4.4. Lunar Zebro Rover. The potential of the current Zebro rover was thoroughly assessed. The results that were obtained during the CHILL-ICE campaign pointed out that the rover would not be able to enter or exit such lava tubes on its own, due to the steepness of the skylight openings and to the big rocks that characterize the entrance, which overcome the rover capabilities. Instead, the rover would prove to be particularly efficient in crawling in the inside of the caves, especially at the back, which is typically narrower and darker than closer to the entrance. It could aid the astronauts by scouting caves before the astronauts, to let them know if it is a place that could host a human settlement or if it has any particularities that are worth a human exploration. On top of this, the live-stream could be used to maneuver the rover from inside the habitat, to assess the conditions of the outside, such as tubes, holes, valves, or the wiring to the power and communications, without having to perform an EVA, saving precious time and oxygen.

Some payload that could be taken into account for future missions is a LIDAR, a temperature sensor and an infrared camera. Another interesting enhancement that was proposed was the inclusion of an obstacle detection model powered by machine learning.

To conclude, the locomotion capabilities of the Zebro rover were tested and proved to be sufficient for such a terrain. All the astronauts agreed about the strong potential of a swarm of Zebro rover to assist humans in the exploration of...
lunar caves, and the valuable feedback that was gathered will be used to enhance future versions of the rover.

**Data Availability**

For access to the data used in these experiments, requests will be considered at the corresponding author’s discretion. The data used in the SpotNet machine learning model is currently unavailable for distribution due to proprietary reasons, but for specific requests, please email David Smith. For data requests relating to the Lunar Zebro experiment, please contact Benedetta Cattani. For data requests relating to the geological experiments, please contact Marc Heemskerk or Estívar Konijnemberg. For data requests relating to the vigilance experiment, please contact Robert Heemskerk.

**Conflicts of Interest**

All authors declare there is no conflict of interest related to this paper or its publication.

**Acknowledgments**

First, we would like to thank the Kwantlen Polytechnic University, Astroland Agency, Blinkinglights, Space Iceland, and TVG Zimens for making this year’s mission possible, especially given the circumstances of the Coronavirus disease 2019 (COVID-19). They have been the key factor in supporting the CHILL-ICE core team before, during, and after this mission. Second, we would like to thank Bernard Foing and EuroMoonMars for their supportive collaboration in the analogue space sector. Thirdly, our research partners Lunar Zebro, TU Delft, Reykjavik University, and 4th Planet Logistics for their contribution in establishing a research habitat inside a lunar-analogue lava-tube: Iceland campaign of Euro-MoonMars, Published in the European Planetary Science Congress, 2020.

For the CHILL-ICE core team, we would like to thank Bernard Foing and EuroMoonMars for their supportive collaboration in the analogue space sector. Thirdly, our research partners Lunar Zebro, TU Delft, Reykjavik University, and 4th Planet Logistics for their contribution in establishing a research base for this mission. And a special gratitude to the landowners of Surthshellir and Stefánshellir for letting us use this magnificent and unique environment for our out-of-this-world-research and to our collaborator Árni B. Stefánsson, Fjörgyn ses., and Punktaský, for their help in the 3D modelling of the cave, research ideas, and local contacts. Lastly, we would like to thank our sponsors LYO Food, Hague University of Applied Sciences, GoPro, Scarpa, 66North, Extreme Design Lab, and Heemskerk Innovative Technology.

**References**


