Kent Academic Repository

Le Corre, Daniel, Mason, Nigel, Bernard-Salas, Jeronimo, Mary, David and Cox, Nick (2025) *New candidate cave entrances on the Moon found using deep learning.* Icarus, 441 . ISSN 0019-1035.

Downloaded from

https://kar.kent.ac.uk/110353/ The University of Kent's Academic Repository KAR

The version of record is available from

https://doi.org/10.1016/j.icarus.2025.116675

This document version

Publisher pdf

DOI for this version

Licence for this version

CC BY (Attribution)

Additional information

Versions of research works

Versions of Record

If this version is the version of record, it is the same as the published version available on the publisher's web site. Cite as the published version.

Author Accepted Manuscripts

If this document is identified as the Author Accepted Manuscript it is the version after peer review but before type setting, copy editing or publisher branding. Cite as Surname, Initial. (Year) 'Title of article'. To be published in ***Title of Journal*** , Volume and issue numbers [peer-reviewed accepted version]. Available at: DOI or URL (Accessed: date).

Enquiries

If you have questions about this document contact ResearchSupport@kent.ac.uk. Please include the URL of the record in KAR. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from https://www.kent.ac.uk/guides/kar-the-kent-academic-repository#policies).

Icarus 441 (2025) 116675

Contents lists available at ScienceDirect

Icarus

journal homepage: www.elsevier.com/locate/icarus

Research Paper

New candidate cave entrances on the Moon found using deep learning

Daniel Le Corre [a,b,*], Nigel Mason [a], Jeronimo Bernard-Salas [b], David Mary [c], Nick Cox [b]

[a] Centre for Astrophysics and Planetary Science (CAPS), University of Kent, Park Wood Road, Canterbury, CT2 7NH, Kent, United Kingdom
[b] Centre d'Etudes et de Recherche de Grasse (CERGA), ACRI-ST, Avenue Nicolas Copernic, Grasse, 06130, Provence-Alpes-Côte-d'Azur, France
[c] Lagrange UMR 7293, Université Côte d'Azur, Observatoire de la Côte d'Azur, Boulevard de l'Observatoire CS
34229, Nice, F-06304, Provence-Alpes-Côte-d'Azur, France

ARTICLE INFO

Keywords:
The Moon
Mars
Planetary surfaces
Remote sensing
Deep learning

ABSTRACT

Pits and skylights are circular to elliptical, rimless, steep-sided depressions on planetary surfaces formed through gravitational collapse, which are of interest for astrobiological investigation and future space exploration. This is due to their ability to signify the presence of, or allow access to, underground cave systems such as lava tubes. The Lunar Pit Atlas contains 16 such features situated within mare regions that were partly discovered via the automated PitScan tool, which was limited by searchable latitudes and data coverage. In order to search for pits and skylights within these unmapped regions, we have trained a series of Mask R-CNN (Region-based Convolutional Neural Network) models on various combinations of Lunar and Martian remote-sensing imagery to detect Lunar pits and skylights. The best-performing model, named ESSA (Entrances to Sub-Surface Areas), was trained upon all available training data with a ResNet50 backbone. During testing on imagery of the famous Mare Tranquillitatis Pit and self-produced mosaics of proposed lava tube collapses, ESSA achieved average F_1-scores of 82.4 and 93.7% for the bounding boxes and predicted masks, respectively. Despite only having surveyed ≈1.92% of the Lunar maria, ESSA has detected two previously uncatalogued skylights: the South Marius Hills and Bel'kovich A Pits (SMHP and BAP) - which are possible candidates for cave entrances on the Moon.

1. Introduction

Pits (also known as pit craters, collapse pits, or sinkholes) are circular-to-elliptical depressions on the surface of rocky planetary bodies that are generally formed through gravitational collapse (Ferrill, 2015). Unlike impact craters, which are formed through the collision between a planetary surface and a meteorite or ejecta material, pits will lack a raised rim or any visible ejecta rays. Pits can have a range of topographies and sizes, from having steep walls and relatively flat floors (Cushing et al., 2015; Wagner and Robinson, 2022) to being more conical or bowl-shaped (Wyrick et al., 2004) and from being on the order of tens of metres to several kilometres in diameter. Several genetic mechanisms have been suggested for the formation of pits, such as the subsidence of an underground evacuated magma chamber (Wagner and Robinson, 2022), sublimation of sub-surface volatiles (Spencer and Fanale, 1990), or dilational normal faulting (Wyrick et al., 2004). These mechanisms can also lead to pits occurring in linear or sinuous alignment — called pit chains (Hagen, 2015). Pits and pit chains can be found on Earth, such as those surrounding the Kilauea volcano site on the island of Hawai'i (Okubo and Martel, 1998), but have also been observed on the Moon (Greeley and Gault, 1979), Mars (Wyrick et al., 2004), Mercury (Gillis-Davis et al., 2009), Venus (Davey et al., 2013) and Phobos (Horstman and Melosh, 1989).

Pits are of particular focus for their potential to provide access to underground cavities such as lava tubes. Lava tubes are sub-surface conduits found sub-parallel to terrestrial surfaces that have been formed by the current or historic flowing of lava (Gadányi and van der Bogert, 2015). Intact lava tube systems can reveal themselves to the surface when their roofs become unstable and collapse to form circular, coalesced or elongated pits that can appear along the path of the tube. These pits are known as 'skylights' in cases where they exhibit steep walls and floors which are flat or sloping towards what may be a roof overhanging an underground cavity (van der Bogert and Ashley, 2015). Infrared images taken by Mars Odyssey's Thermal Emission Imaging System (THEMIS) have suggested that the thermal properties of seven Martian skylights in nighttime images appear to resemble what would be expected of sub-surface materials (Cushing et al., 2007). More recently, radar imagery taken by the Lunar Reconnaissance Orbiter (LRO) MiniRF instrument of the renowned Mare Tranquillitatis Pit

---

[*] Corresponding author at: Centre for Astrophysics and Planetary Science (CAPS), University of Kent, Park Wood Road, Canterbury, CT2 7NH, Kent, United Kingdom.

E-mail address: dl387@kent.ac.uk (D. Le Corre).

https://doi.org/10.1016/j.icarus.2025.116675
Received 11 April 2025; Received in revised form 29 May 2025; Accepted 29 May 2025

(MTP; Haruyama et al., 2010) has proved the existence of a cave conduit tens of metres long, potentially suggesting an accessible intact lava tube (Carrer et al., 2024).

Lava tubes are most commonly formed within smooth, low viscosity/velocity pahoehoe basaltic lava flows, but can also be found in aa flows, which are rougher and have higher viscosity and flow rates (Léveillé and Datta, 2010). Figure 2 in Sauro et al. (2020) displays the six main mechanisms (adapted from Kempe, 2012) for the formation of lava tubes, which are: (i) over-crusting of lava channels by coagulation of floating cooled slabs; (ii) lateral accretion and eventual closure of lava channel shelves; (iii) shallow inflation between pahoehoe sheets and downward erosion; (iv) inflation and draining of an aa flow nucleus; (v) deep inflation between older lava flow layers enlarged by thermal erosion and breakdown; (vi) thermal erosion along a sealed fracture.

Speleologists have been able to explore numerous lava tubes on Earth in locations known for volcanic activity, such as the Canary Islands (Sauro et al., 2019), Korea (Kim et al., 2022) and Sicily (Calvari and Pinkerton, 1999). The Ronald Greeley Centre for Planetary Studies hosts a digital database of 1084 terrestrial lava tubes from 34 countries across six continents (RGCPS, 2025). On the Moon, the total collapse of lava tubes has long been proposed as an explanation for the formation of sinuous trough-like structures called 'rilles' (Greeley, 1971). The same may also be true for the 'canali' observed on the surface of Venus (Gregg and Greeley, 1992). Lava tubes are also believed to emanate from the various shield volcanoes on Mars (Léveillé and Datta, 2010).

On the Moon and Mars, where physical exploration is not readily possible, the identification of the various collapse morphologies mentioned above relies upon remote-sensing data. Currently, there are two main catalogues for pits and skylights on the Moon and Mars. On the Moon, Wagner and Robinson (2021) have developed the Lunar Pit Atlas (LPA), which gives the positions, morphometric parameters, and descriptions of 281 Lunar pits. Despite the majority of LPA features being found within impact melt ponds and thus unrelated to lava tubes, there remain 20 pits found within the highlands (5) or maria (15). LPA mare pits, such as the MTP, have been the subject of 3D reconstruction efforts using mono and stereoscopic analysis (Wagner and Robinson, 2022) and photogrammetry (Zhou et al., 2024). The creation of the LPA was possible thanks to the high resolution (≈0.5 to 2.0 m/px) and good surface coverage of data from the LRO Narrow Angle Camera (LROC NAC; Robinson et al., 2010). The naming convention used by the LPA for non-impact melt pits typically combined the location where the feature is found and the suffix "Pit" irrespective of whether it is a pit or skylight by definition.

On Mars, Cushing (2015) has catalogued over 1000 Martian possible cave entrances to form the Mars Global Cave Candidate Catalog (MGC^3). MGC^3 subcategorises its feature based on morphology, but also by its location (either globally or relative to other features such as channels). Similarly to LPA, MGC^3 gives the coordinates, size estimates, and short descriptions for each of its features. Imagery taken by the Mars Reconnaissance Orbiter's (MRO) Context Camera (CTX) and High Resolution Imaging Science Experiment (HiRISE) sensors were used in the creation of MGC^3. The former has near-global coverage of the Martian surface at 6 m/px resolution (Malin et al., 2007), whilst HiRISE has a far smaller footprint but with much improved detail at 0.25 to 0.5 m/px (McEwen et al., 2007). HiRISE has also been acquiring multiband colour images during its lifetime, albeit at 1/5 of the swath width of the corresponding single-band red-wavelength versions (McEwen et al., 2007).

In terms of larger collapses, six of the 193 sinuous rilles catalogued by Hurwitz et al. (2013) using LRO Wide Angle Camera (LROC WAC) and SELENE Terrain Camera (TC) visual imagery were identified as being related to the partial collapse of lava tubes. In a similar study for Mars, Crown et al. (2022) mapped the western flank of Alba Mons for potential lava tube collapses using the MRO CTX mosaic (Robbins et al., 2023), resulting in 331 examples found to have a mean length of 36.2 km. Sauro et al. (2020) performed a comprehensive morphometric comparison between known lava tubes on Earth and proposed analogues on the Moon and Mars. In using high-resolution satellite imagery and Digital Elevation Models (DEMs), they found that Martian and Lunar lava tubes could be up to 1 to 3 times more voluminous than their terrestrial counterparts. The presence of intact Lunar lava tubes deep underneath the surface has also been inferred through mass-deficit (Chappaz et al., 2017) and radar data (Kaku et al., 2017).

Pits and skylights, due to their propensity to suggest the presence of, or allow access to, underground cave systems, will be relevant for future crewed and robotic space exploration on the Moon (Coombs and Hawke, 1992). Sub-surface caves, such as lava tubes, have long been identified as a primary habitat for future astronauts on the surface of bodies without significant atmospheres (Boston et al., 2004) due to the natural shelter from ionising radiation, micrometeorite impacts and dust storms. The risk of Acute Radiation Syndrome, cancers, and nervous system damage to future astronauts on Mars has been modelled by Atri et al. (2022), who concluded that constructing habitats inside lava tubes is one of the "best options" for mitigating the dangerous levels of incident radiation. Martian caves have also been simulated by Williams et al. (2010) to sustain stable reserves of water ice (i.e. no net loss after one Martian year) across much of the surface.

With regard to the Moon, potential caves have already been the subject of proposed space missions, such as Moon Diver (Nesnas et al., 2023). Moon Diver was a mission concept in which a rotor-craft tethered to a lander would land near the MTP, abseil down its wall and determine the composition and mineralogy of the Lunar mare bedrock. Three context cameras would also provide information on whether the void beneath the MTP is, in fact, an evacuated intact lava tube that could become a future Lunar base. As such, the structural stability of lava tubes (Blair et al., 2017; Theinat et al., 2020) and the possibility of pressurising a small example with breathable air (Martin and Benaroya, 2023) have both been modelled for a range of roof thicknesses and materials. Lunar caves have also been simulated to maintain near-constant temperatures of around 290 K all year round (Horvath et al., 2022).

Perhaps due to the abundance of high-quality satellite data, there have been multiple studies employing automated techniques to detect and analyse planetary surface features related to the collapse of underground cavities. Firstly, several LPA features were detected using the automated PitScan tool (Wagner and Robinson, 2014). PitScan operates by applying a threshold to LROC NAC Calibrated Data Record (CDR) images in order to extract shaded regions. Then, it considers how the pixel brightness changes along the Sun's line of sight across the feature to determine if they are positive or negative elevation features. However, in order to restrict the types of elevations that can cast significant shadows, PitScan was only applied to images which had a solar incidence angle (α) of less than ≈50°. Therefore, PitScan could not search for Lunar pits or skylights far outside of the ±50°latitude range, and the coverage of mid-latitude regions was marginally reduced.

As opposed to detecting pits/skylights outright, Machine Learning methods have been used by the Pit Topography from Shadows (PITS) tool to automatically calculate apparent depth (h) profiles of catalogued Lunar and Martian pits (Le Corre et al., 2023). h is the relative depth between the pit's rim and the shadow's edge along the Sun's line of sight. PITS does this by segmenting between shadow and non-shadow pixels with k-means clustering and silhouette analysis and then measuring the shadow's width (parallel to the direction of illumination) at each pixel along its entire length. This is an improvement upon previous work that infers pit depths from shadows, as they typically only provide single values (Wyrick et al., 2004; Cushing et al., 2015). Since it only requires one MRO HiRISE or LROC NAC image to operate, PITS allows for depth calculation to continue where corresponding altimetry data may be of insufficient resolution or stereo imagery is unavailable. Another benefit to PITS is that it is able to automatically correct its profiles for any protruding boulders, as well as for non-nadir observations for all solar and spacecraft azimuth angles. Across 19 HiRISE images of MGC^3 pits, PITS' automated shadow extraction achieved recall (R) and precision (P) rates of 99.6% and 94.8%, respectively.

DeepLandforms is a platform for training Deep Learning (DL) models to detect planetary landforms in satellite imagery, whose first use case was detecting Martian pits in MRO HiRISE imagery (Nodjoumi et al., 2023). When training on two target classes of only skylights and impact craters, DeepLandforms achieved a testing F_1-score (F_1) of 81.1%. Alternatively, Watson and Baldini (2024) has trained a simple Convolutional Neural Network (CNN) on screenshots of MGC^3 features within the MRO CTX Global Mosaic (Robbins et al., 2023) in JMARS (Christensen et al., 2009). 61 new and 24 known (out of a total of 341 previously catalogued) potential cave entrances (PCEs) were found out of a total of 10,834 detections when applied to five volcanic regions, resulting in a P and R of 0.78% and 7.03%, respectively. Whilst there has been no work yet on automatic detection of Lunar pits or skylights, Zhang et al. (2024) have used a combination of elevation (LOLA-Kaguya co-registered DEM (Barker et al., 2016) at 59 m/px), slope (derived from DEM) and visual imagery (LROC WAC upscaled to DEM resolution) data to train a DL model to detect Lunar sinuous rilles. This multi-modal approach achieved an F_1 of 93.67% during testing, with significant improvement compared to the model that was trained on a single channel of the above sources.

Insomuch as there have been several efforts at manually or automatically mapping pits, skylights and other lava tube-related collapses, there is yet to be a study which uses DL in the search for pits and skylights on the Moon that may play a role as cave entrances. Given the global availability of high-resolution LROC NAC images and the aforementioned mapping exercises, there is high-quality imagery for creating a training dataset. As such, this work aimed to train DL models upon a combination of Lunar and Martian imagery to detect individual Lunar pits and skylights. To maximise the chance of detecting previously uncatalogued features, which could provide access to subsurface cavities, these models were applied to select mare regions which satisfied one or more of the following conditions: (i) the region had not been completely mapped with PitScan — either due to high-latitudes or the unavailability of data at the time; (ii) it exhibited sinuous or linear depressions within the global LROC WAC mosaic; (iii) it had large sub-surface mass-deficit according to Chappaz et al. (2017). We have also used the PITS tool, as described above, to derive h profiles where there were appropriate LROC NAC observations of any prospective detections.

2. Dataset

2.1. Training and validation data sources

A combination of several manually and automatically produced catalogues was used in this study as a source of 'ground truth' for pits and skylights found on the Moon and Mars. A full breakdown of the numbers of LROC NAC/HiRISE image products and labelled features amassed from each of the following sources of Lunar and Martian training, validation and testing data is given in Table 1.

Table 1
Sources of training, validation and testing data of pits and skylights for the Moon and Mars — along with the numbers of acquired LROC NAC/HiRISE images and labelled features. The LPA mare quantities include the 5 LROC NAC images containing the Mare Tranquillitatis Pit (MTP), which were used for testing. The number of images given for the proposed lava tube collapses in Sauro et al. (2020) are the amount used to produce the mosaics used for testing data. All other data was used for training and validation.

Body	Source	Feature description	Images	Labels
Moon	Lunar Pit Atlas (LPA; Wagner and Robinson, 2021)	Impact melt	42	68
		Mare	47	53
		Highland	15	15
	Hurwitz et al. (2013)	Lava tube-related rilles	11	37
	Klem et al. (2014)	Apollo 11–16 sites	11	10
	Haase et al. (2019)	Apollo 17 site	1	0
	Sauro et al. (2020)	Lava tube collapses	27	25
Mars	Mars Global Cave Candidate Catalog (MGC^3; Cushing, 2015)	Atypical pit crater	126	235
		Generic pit	4	37
		Skylight	20	66
		Small, rimless pit	15	23
	Watson and Baldini (2024)	Potential cave entrance	2	17

2.1.1. Moon

The LPA documentation gives the details of the 2 to 3 LROC NAC images which were used in cataloguing and measuring each impact melt, mare and highland feature. As such, they are of sufficient resolution and have appropriate sensing conditions to be useful for model training. The raw Experimental Data Record (EDR) versions of these LROC NAC images (Robinson, 2009) were obtained from NASA's Planetary Data System (PDS). Although, as is explained further in Section 3.1, the small size of impact melt features meant that only the highest resolution image available for each feature was utilised.

To train models on more elongated pits, LROC NAC EDR images were also acquired for all six lava tube-related sinuous rilles (#24, 37, 39, 75, 83 and 175) mapped by Hurwitz et al. (2013). Since there has been repeated coverage of these regions, two LROC NAC EDR products were selected for each feature — one where the Sun was high above the horizon ($\alpha \gtrsim 50°$) and one where it was low ($\alpha \lesssim 40°$).

All EDR images were then processed from raw data to high-level map-projected images using the ISIS (Laura et al., 2023) and GDAL (GDAL/OGR contributors, 2025) software packages. EDR products were first converted to ISIS cubes, then echo and radiometrically corrected, map-projected, and converted to GeoTiff format.

Due to a user-defined shape model being used when map-projecting, there was a misalignment between the processed images and the LROC NAC shapefile footprints that are available through the PDS Lunar Orbital Data Explorer (ODE). An accurate footprint of the image as it appears on the surface is necessary for ensuring that large no-data regions are not seen during training. Therefore, new footprint shapefiles were created with GDAL by taking the processed image and setting all pixels with intensities greater than 0 to a value of 1 to produce a new binary image. This binary image was then polygonised to convert from raster to vector data. By also using these footprint polygons to remove any detections made within no-data regions, the final model developed in this work can be applied to tiles at the edges of individual image products without fear of erroneous predictions.

Controlled LROC NAC mosaics of the Apollo program landing sites (Klem et al., 2014), as well as the Apollo 17 controlled 0.5 m/px ortho-mosaic (Haase et al., 2019), have been utilised as a source of 'empty' (i.e. no target features present) images of the Lunar maria. Training on such data aims to reduce the rate of false positive detections. However, the Hadley rille seen in the Apollo 15 mosaics was labelled as an example of an elongated pit (see Fig. 2).

2.1.2. Mars

Knowledge of the locations of Martian pits and skylights was also required to be able to deduce whether there was any improvement when training with imagery of a different planetary surface. The primary source of 'ground truth' for Mars was MGC^3 features with relevance to the collapse of sub-surface cavities (i.e. no cracks, fractures or karst pits). Therefore, MRO HiRISE Reduced Data Record Version 1.1 (RDRV11) images, which contained MGC^3 'Atypical pit craters', 'generic pits', 'skylights' and 'small rimless pits' were acquired from PDS (McEwen, 2007).

The same cross-referencing was performed for the 61 potential cave entrances (PCEs) identified by Watson and Baldini (2024), although just 3 were found to have corresponding HiRISE RDRV11 imagery.

Furthermore, only 2 of the 3 features appeared clearly as either a pit or a skylight under the increased resolution of HiRISE. Unfortunately, no useful HiRISE RDRV11 images could be found for the lava tubes mapped by Crown et al. (2022). This was likely because only the start, end and mid-points of the lava tube path are available (Crown et al., 2022), but also as many of the features were heavily modified by aeolian processes.

As the surveys above have focused mostly on the features which could provide access to caves, they were often found within chains of, or in proximity to, other uncatalogued pits. Therefore, the total number of labelled Martian pits and skylights was significantly larger than the number of corresponding HiRISE RDRV11 image products.

Whilst RDRV11 products are map-projected, there is an occasional error when viewing and analysing them in GIS software, whereby an offset in the latitude direction is observed. This misplacement was corrected by replacing the values of the 'standard parallel' and 'centre latitude' parameters in the Proj4 string and reprojecting the image using GDAL. This meant the labelling could be carried out in GIS software, as is discussed further in Section 3.1, with the knowledge that the labels were in the correct location on the surface.

2.2. Testing data sources

In addition to separate training and validation datasets, it is also vital to have an additional dataset upon which the final, trained model can be tested as a means of determining its real-world performance. This testing dataset should be representative of the features that the model is tasked with detecting, which in this case are Lunar pits and skylights, which may be related to the collapse of lava tubes.

As a result, LROC NAC EDR images that covered the sinuous pit chains identified by Sauro et al. (2020) as potential lava tube collapses were also collected. These include the Gruithuisen, Marius Hills A and Marius Hills B sinuous collapse chains. It was made certain that none of the lava-tube-related rilles that were seen during training were also included in this testing dataset. The EDR products containing these lava tube candidates were processed using the same procedure as those described in Section 2.1.1, except with the additional step of photometric correction. This was done to reduce the visibility of seams when these individual processed images were stitched together to form three separate uncontrolled mosaics. These self-produced uncontrolled mosaics are shown later in Fig. 7.

Since there were only two skylights (the Marius Hills and West Marius Hills Pits, MHP and WMHP; Haruyama et al., 2009) found within the candidate lava tube collapses mentioned above, additional testing data of this target class was required. Therefore, five LROC NAC images containing the MTP were acquired and processed as above. The five images which were selected were near-nadir, high-resolution observations, which exhibited a good range of illumination conditions and shadow sizes.

2.3. Inference data sources

As mentioned in Section 1, eight Regions of Interest (RoIs) in the search for previously uncatalogued Lunar pits and skylights were selected from the various mare units on the Lunar surface as they satisfied a series of conditions. RoIs were chosen if they had not been completely mapped by PitScan due to being situated at latitudes beyond ±50° or insufficient coverage by LROC NAC at the time. These RoIs may have also exhibited linear collapses in the global LROC WAC mosaic or are regions with high degrees of sub-surface mass deficit according to Chappaz et al. (2017).

The chosen RoIs were the following: the Imbrian mare deposits inside the (a) Antoniadi, (b) Bel'kovich A, (c) Endymion, and (d) Lyot impact craters; (e) the Eratosthenian mare deposit inside Plato crater; (f) the wider Marius Hills region (between 10 to 15°N and 55 to 60°W) which contains the Marius A and B proposed lava tube collapses, as well as high degrees of sub-surface mass deficit (Chappaz et al., 2017); (g) the high-latitude Mare Humboldtianum and (h) the Imbrian mare deposits within Poincaré Basin. In terms of area, these RoIs (labelled with the same letters in Fig. 1) constitute ≈1.92% of the Lunar maria and just ≈0.23% of the entire surface.

The LROC NAC images (with resolutions better than 1.5 m/px, α lower than 70°, and satellite emission angles, ϵ, of less than 5°) which overlapped each RoI were found by importing the geo-referenced shapefile of the corresponding mare unit(s) from the global geologic map by Fortezzo et al. (2020) into JMARS (Christensen et al., 2009). Using the 'Mosaic Select' tool within JMARS, the fewest number of LROC NAC images which provided the maximum degree of coverage of the RoI were found. The raw products were then acquired through PDS and processed as in Section 2.1. In order to preserve disk space, however, the ISIS cubes were clipped to the wider latitude–longitude bounds of the RoI. Considering the relatively small size of these RoIs, a total of 2157 raw LROC NAC images were downloaded and processed using the above procedure for these eight sites at a rate of around 25 to 30 images per hour (using one of 4 CPU nodes sharing a HPE ProLiant DL380 and 512 GB of RAM).

3. Methodology

3.1. Training dataset creation

The first step in creating a labelled dataset of LROC NAC and HiRISE images was to address the difference in spatial resolutions between the two sensors. Between September 2009 and December 2011, while LRO was still in its original 50 km circular polar orbit, the majority of LROC NAC images had a resolution of ≈0.5 m/pix (Wagner et al., 2024). Since then, LRO has been in a fuel-conserving elliptical orbit with its periapsis near the Lunar South Pole, meaning that images taken in the northern and southern hemispheres range in resolutions from 1.0 to 2.0 and 0.4 to 1.0 m/px, respectively (Wagner et al., 2024). Given that HiRISE has a best resolution of 0.25 m/pix, there is a maximum of 8 times more detail in its images when compared to the worst available with LROC NAC. As such, all processed LROC NAC EDR images and projection-corrected MRO HiRISE RDRV11 images were downscaled to a common resolution of 1.5 m/px using cubic spline interpolation. This value was selected to maximise detail and minimise the number of instances where the LROC NAC resolution was any poorer. In the few cases where this was true, it meant that they were upscaled, with the poorest resolution being ≈1.83 m/px. Downscaling to a consistent resolution was aimed at enabling the model to learn a sense of the scales of Lunar and Martian pits. However, this did mean that some of the catalogued features mentioned in Section 2 (particularly LPA impact melt pits) could not be resolved at this scale, nor could they be confidently labelled.

Thanks to their map projection, these downscaled image products could then be loaded into QGIS (QGIS Development Team, 2025) and labelled by drawing polygons around the target feature so as to include the funnel, rim and floor regions of the feature. These polygons were then saved to a separate shapefile layer for each image, meaning that image products with multiple pits or skylights will have as many polygons in their shapefile. Each polygon was assigned an integer according to whether they appeared to be a (1) skylight (no raised rim; steep and/or overhanging walls; deep with a pronounced shadow) or a (2) pit (no raised rim; often larger, coalesced, or channel-like; all walls are visible; and may not have a pronounced shadow). Examples of the labelling procedure for a series of pits and skylights on the Moon and Mars are given in Fig. 2.

Since LROC NAC and HiRISE image products are far too large to pass through a DL model (even after downscaling), they both require cropping into smaller tiles. This was done with the use of a Python script that would loop through all features in the image and randomly generate coordinates for a 2048 × 2048 px tile which fully encapsulates

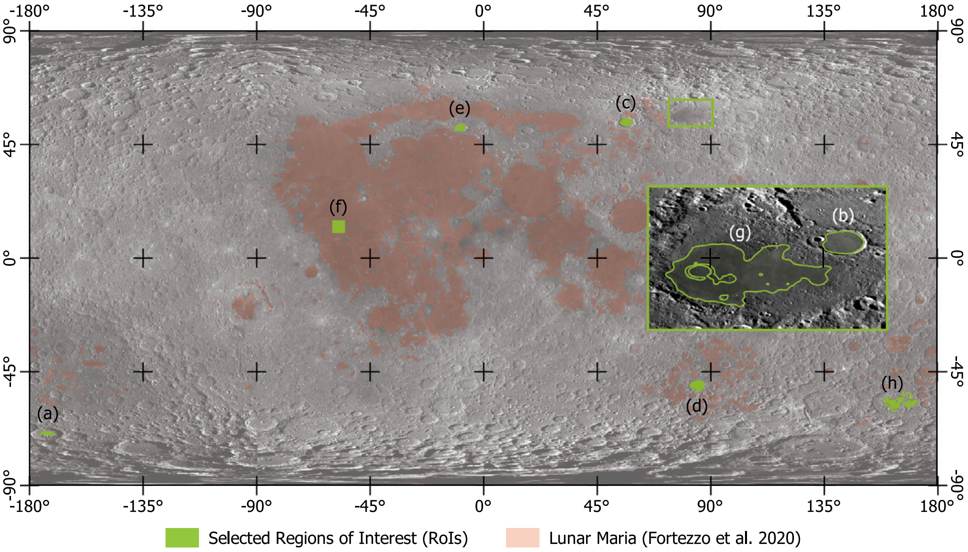

Fig. 1. Global map of the Regions of Interest (RoIs) in the search for previously uncatalogued Lunar pits and skylights. (a, c, d, e) Antoniadi, Endymion, Lyot and Plato craters; (f) Marius Hills and (h) Poincaré Basin. The inset map highlights (b) Bel'kovich A crater adjacent to (g) Mare Humboldtianum. The Lunar maria, from Fortezzo et al. (2020), yet to be surveyed are given in red. Base-map is the 100 m/px LROC WAC Global Morphology Mosaic (Speyerer et al., 2011).

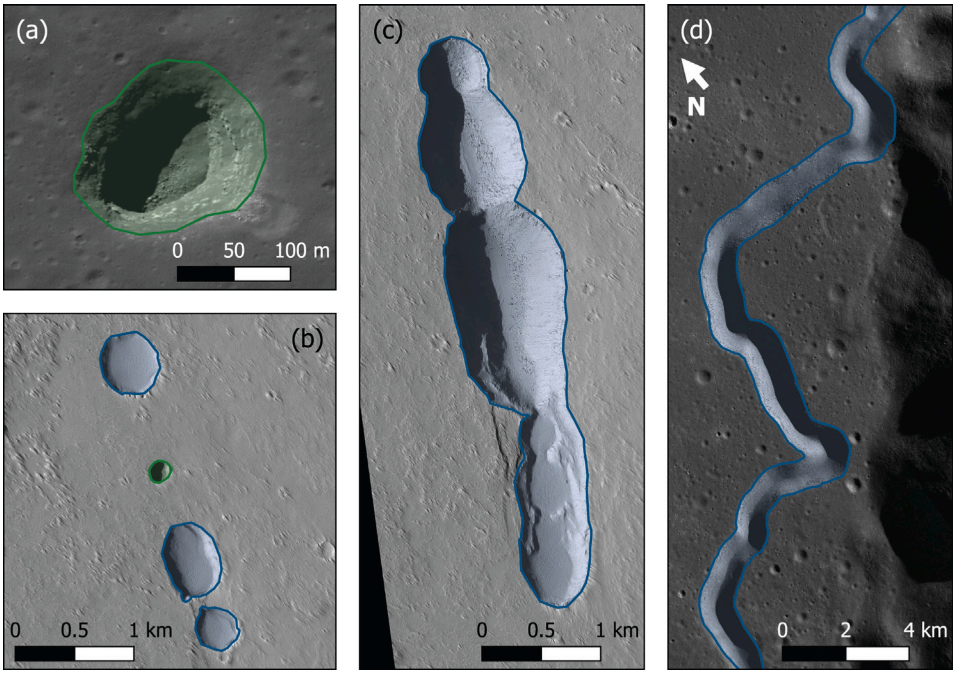

Fig. 2. Examples of Lunar and Martian pits (in blue) and skylights (in green), along with their corresponding labels produced in this work. (a) LROC NAC image M138819477R of the LPA feature known as the 'Mare Ingenii Pit'; (b) a Martian pit chain, as seen in HiRISE image ESP_011677_1655, containing the MGC^3 feature 'APC079'; (c) coalesced pits in the same chain and image as (c); (d) A section of Hadley rille close to the Apollo 15 landing site within the mosaic NAC_ROI_APOLLO15LOB_E259N0038.

it. Furthermore, these random tiles could not contain any other features which already had corresponding tiles. As such, this was repeated for a maximum of 10 iterations in case the coordinates were not appropriate. If a randomly generated tile fully encapsulated more than one feature which did not already have corresponding tiles, no further tiles were generated for all those contained. This ensured that features were not contained within multiple tiles sourced from a single LROC NAC or HiRISE observation.

There were features, however, such as Lunar rilles, whose extents were larger than this tile size in either or both of the latitude and longitude directions. In these instances, a coordinate grid of 2048 × 2048 tiles with a 75% overlap was created. By iterating through this grid, by row by row, tiles were created if they met at least one of the following two conditions: (i) 90% of the feature's area was encapsulated within the tile, or (ii) the feature occupied more than 10% of the tile's area. This ensured that elongated pits or rilles could be seen during training, and in sufficient amounts that they would not be confused for the edges of large craters. Additionally, for a tile to be created via this iterative process, it could not overlap any of the already created tiles with an Intersection over Union (IoU, see Section 3.3) of more than 20%. As mentioned in Section 2, when tiling the controlled LROC NAC mosaics of the Apollo landing sites, tiles which did not overlap any features whatsoever were allowed to be exported.

Tiling was also necessary for the imagery used for model testing and inference upon the eight RoIs. For the uncontrolled testing mosaics and five LROC NAC images containing the MTP, a coordinate grid with an overlap of 25% was utilised for tiling. This was to allow the Mask R-CNN model the maximum opportunity to detect some of the larger collapses while not misrepresenting the amount that it would see when being applied to the images covering the RoIs. With regard to the RoIs, the processed images covering each site were tiled with no overlap whatsoever. This was because there was already a large degree of overlap between the LROC NAC image products, meaning that tiles could be removed from the final inference dataset if they did not improve the coverage of that particular RoI. This was done due to concerns regarding storage limits since the total surface area of the RoIs was significantly larger than that of the testing regions. Therefore, as shown later in Section 4.3, some detections made within the Lunar RoIs have been made in multiple tiles originating from different LROC NAC image products.

As per the requirements for the DL model of choice (see Section 3.3), the polygon labels produced in QGIS were rasterised separately for each target class to create up to two images (also with a x–y size of 2048 px) which contain pixel-level annotations of the target features found within the tile. In these annotation images, the background was set to 0, and the pixels corresponding to each instance of the given class were assigned a sequentially increasing integer value starting from 1.

Given the limited number of catalogued features, some data augmentation was necessary to overcome training dataset deficiency, but also to avoid over-fitting. As such, rotated versions of each image (clockwise by 90, 180 and 270°) were also included in training, thus increasing the number of images by a factor of 4. Images which were earmarked for training as opposed to validation also had the possibility of undergoing three further steps of data augmentation, each with a probability of 0.5 and no mutual exclusivity. Firstly, images could undergo Contrast-limited Adaptive Histogram Equalisation (CLAHE), which is a variant of Adaptive Histogram Equalisation which clips an image's histogram to a chosen value to minimise the amplification of noise in uniform regions (Pizer et al., 1987). Secondly, Gaussian noise could also be applied to the images with a σ randomly selected from a uniform distribution between 0.0075 to 0.0125. Lastly, the perspective of the image could also be transformed to simulate different viewing angles. This was done using the `perspective` function within PyTorch and a distortion scale randomly selected from a uniform distribution between 0.5 to 1.5.

After downscaling, tiling and data augmentation, the resulting LROC NAC and HiRISE tiles had to be split between those used for model training and validation. A split of 4:1 was chosen, with the proportion of annotated and unannotated tiles also being divided equally across the training and validation sets. Seeing as nearly twice as many unannotated tiles were exported, unannotated tiles were randomly removed from the final training/validation datasets such that the ratio of annotated to unannotated tiles was approximately equal. The resulting training and validation datasets are summarised in Table 2, with breakdowns of the number of image tiles which contained target features. These datasets are comprised of repeat observations under different sensing conditions and augmented data containing approximately 80 and 250 unique Lunar and Martian features, respectively. These quantities have been estimated since the features were not labelled with their catalogued name (e.g. 'MTP'), solely whether they were a pit or a skylight.

3.2. Creating synthetic moon data

Up to this point, efforts have been made to make up for the relative lack of catalogued pits and skylights on the Moon when compared to Mars for use as training data for DL models. These include data augmentation and utilising images taken of the same features under different illumination conditions. However, the suspicion was that this would be insufficient. Therefore, the possibility of creating synthetic labelled training data for Lunar pits and skylights was explored to counteract this.

Table 2
Breakdown of the resulting training and validation dataset sizes. The 'Moon + Synth.' dataset consists of genuine LROC NAC data plus synthetic Moon samples (see Section 3.2). 'Moon + Mars' contains only genuine LROC NAC and HiRISE data. Whereas 'Moon, Mars + Synth.' contains all available training data. The numbers of labelled or empty samples have been given separately.

Dataset	Purpose	No. tiles	
		Labelled	Empty
Moon + Synth.	Training	1152	1149
	Validation	284	287
Moon + Mars	Training	1156	1152
	Validation	284	288
Moon, Mars + Synth.	Training	1808	1802
	Validation	444	450

It could be feasible to generate a series of artificial DEMs containing Lunar pits by drawing from real elevation data and illuminating it with a given solar position. Although, this would require significant computational effort to generate enough DEMs from stereo imagery and would likely require training an additional generative DL model. Instead, synthetic Moon training samples were generated by implanting augmented Martian target features into LROC NAC tiles known not to include any.

This process involved extracting tiles which did not contain any target features (i.e. could be considered 'empty') from the larger LROC NAC products containing LPA (only mare and highland) features and lava-tube-related rilles. Empty tiles were not retrieved from LROC NAC images containing LPA impact melt pits since the surrounding surface was often too heavily fractured. A Python script was then written, which would loop through all empty LROC NAC tiles and randomly select a HiRISE tile from the training dataset. If the HiRISE tile contained multiple target features, then only one was randomly selected at a time for implantation.

The HiRISE tile corresponding to the chosen feature was then cropped to the extent of the target feature using the training annotations described in Section 3.1. The cropped tile was then rotated by any angle (drawn from a uniform distribution) and resized in its x and y directions by separate factors (both drawn from a normal distribution with a mean and σ of 1 and 0.15, respectively). This allowed for circular Martian features to become more elliptical and vice versa. These same transformations were applied to the training annotation so that the pixel-level labels were placed correctly for the rotated and resized feature. Using the training annotation as a binary mask meant that the pixels solely corresponding to the target feature could be extracted. The transformed Martian feature was then implanted into the empty LROC NAC tile, thereby replacing the original pixel values at a random position such that the entirety of the feature fell within it. By using the result of passing the feature's binary mask through a Gaussian filter (with a kernel size of 5 and a σ of 3) as weights, the implanted HiRISE pixels and the surrounding empty LROC NAC image were averaged to blend the two regions.

Whilst it would be possible to rotate the cropped Martian feature to match the solar azimuth angle of the empty LROC NAC image, as well as implant features with a comparable α, the expectation is that this would only make the synthetic tiles more convincing to a human, rather than a DL model. Typical DL models tasked with feature recognition, such as the one used in this work (see Section 3.3), subdivide images into regions before attempting to classify them. Therefore, the presence of shadows being cast from different directions is unlikely to affect the model's performance. However, Section 5.3 will describe a way in which azimuth-matching may be beneficial in future work.

Before the HiRISE tile was cropped to the extent of the selected feature, histogram-matching was conducted in order to better assimilate with the brightness and contrast of the Lunar surface. Histogram-matching is the process of adjusting the pixel values of an input image (HiRISE tile) such that the Cumulative Distribution Function

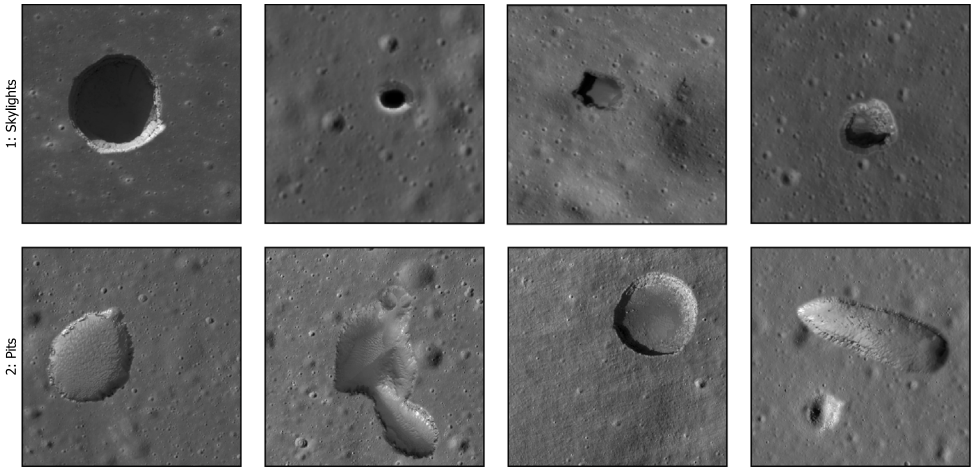

Fig. 3. Examples of synthetic Moon training samples produced by implanting transformed and histogram-matched Martian pits (bottom row) and skylights (top row) into 'empty' LROC NAC tiles. The samples shown here have been cropped to show the implanted Mars feature in detail within the context of the Moon's surface and, as a result, are not to scale.

(CDF) of its histogram is equal to that of a reference image. The Python script then displayed five options on-screen according to how the histogram-matching was conducted. These options were: (i) no histogram-matching; (ii) histogram-matched to the randomly selected empty LROC NAC tile; (iii) average between (i) and (ii); (iv) histogram-matched to the average histogram of all annotated LROC NAC tiles in the training dataset; (v) average between (ii) and (iv). A manual decision was then made as to which one, if any, of these five options should be added to the training dataset. This was according to whether the combined histogram-matching, blending between LROC NAC and HiRISE pixels, and warping of the Martian feature had all collectively produced a convincing example of a Lunar pit or skylight.

This produced an additional 812 training images, with each having one feature. Fig. 3 displays some of the synthetic Moon tiles that were generated as part of this work and subsequently used for training a DL model to detect Lunar pits and skylights. Significantly more synthetic images were able to be created using Martian skylights (628) compared to Martian pits (184). This is a result of two factors. Only Martian pits or skylights that were fully contained within the extents of the source HiRISE tile were allowed to be implanted. As expected, many of the coalesced or elongated pits did not satisfy this condition. Of those that did, several pits exhibited aeolian features in their floors, which were too prominent to be convincing as a Lunar pit.

3.3. Model architecture and evaluation

The chosen model for this work is a Mask R-CNN (Region-based Convolutional Neural Network; He et al., 2017). R-CNNs are a series of Deep Convolutional Neural Networks (DCNNs) which perform object detection by proposing several regions of interest that are likely to contain target features and feeding these through a neural network for detection and, in some cases, classification (Girshick et al., 2013). Where previous R-CNN iterations used selective search to generate these region proposals (Girshick, 2015), the Faster R-CNN implements a Region Proposal Network (RPN), which is a fully convolutional network (FCN) that is trained to generate region bounds and their likelihood to contain target features (Ren et al., 2015). Mask R-CNN further develops upon Faster R-CNN by including additional FCNs that are tasked with predicting a mask of the target feature simultaneously with its bounding box (He et al., 2017). Mask R-CNN is also a well-established form of instance segmentation model, which has even seen extensive use in detecting surface features in planetary remote-sensing data (Ali-Dib et al., 2020; Alshehhi and Gebhardt, 2022; Prieur et al., 2023; Rubanenko et al., 2021; Nodjoumi et al., 2023). A diagram of the Mask R-CNN model, within the context of detecting pits and skylights within satellite imagery of the Moon and Mars, is given in Fig. 4.

As the training framework was built from scratch in Python using PyTorch, there was the ability to swap out Mask R-CNN's default backbone model (ResNet50; He et al., 2016) for a series of other CNNs. The backbone model is responsible for learning the feature maps that will be used by the rest of the model for region proposal, detection and classification. DenseNet121, 169 and 201 (Huang et al., 2016); VGG16 and 19 (Simonyan and Zisserman, 2014); as well as ResNet101 and 152, were all chosen to be trained since they were compared by Dawson et al. (2023) for scene classifications of different carbonate rocks. Dawson et al. (2023) also assessed the InceptionV3 network, however, persistent errors meant that this could not be implemented as the backbone for Mask R-CNN in this work. Smaller ResNet models (ResNet18 and ResNet34) and MobileNetV2 (Sandler et al., 2018) were also trained as backbones since they were quick to implement and train due to having fewer convolutional layers.

Before evaluating the model's performance, all detections within the validation samples undergo non-maximum suppression (NMS). NMS is a technique for iteratively removing repeat bounding box detections of the same objects made by object detection models. The algorithm starts by sorting all detections by confidence score (from highest to lowest) to create set A. An empty set, B, within which the filtered detections will be placed, is also initialised. NMS then iterates through A, and determines whether the bounding box A_i overlaps any of the lower-scoring boxes by a sufficient IoU threshold. Any of these overlapped lower-scoring boxes are removed from set A and the box A_i is added to set B.

Performance metrics have been calculated over the course of the training process and exported to TensorBoard for simple visualisation. These metrics include the training losses for the various elements of the Mask R-CNN model (classifier, box regressor, mask, object-ness, RPN box regressor); the average recall (R), precision (P) and F_1-score (F_1) of the bounding boxes and predicted masks upon the training samples; and the average R, P and F_1 of the boxes and masks upon the validation samples. Recording P, R and F_1 separately in this way allows easier assessment of whether the model has over-fit to the training data, wherein the performance on the validation samples would be significantly worse than on the training samples.

P, R and F_1 are defined in Eqs. (1)–(3), but the numbers of true positives (TP), false positives (FP) and false negatives (FN) are derived differently for the bounding boxes and predicted masks. A bounding box is assessed as a TP or FP depending on if it was detected with the correct class, a confidence score of ≥50%, and the bounding box overlaps the extent of the corresponding label with a sufficiently strict

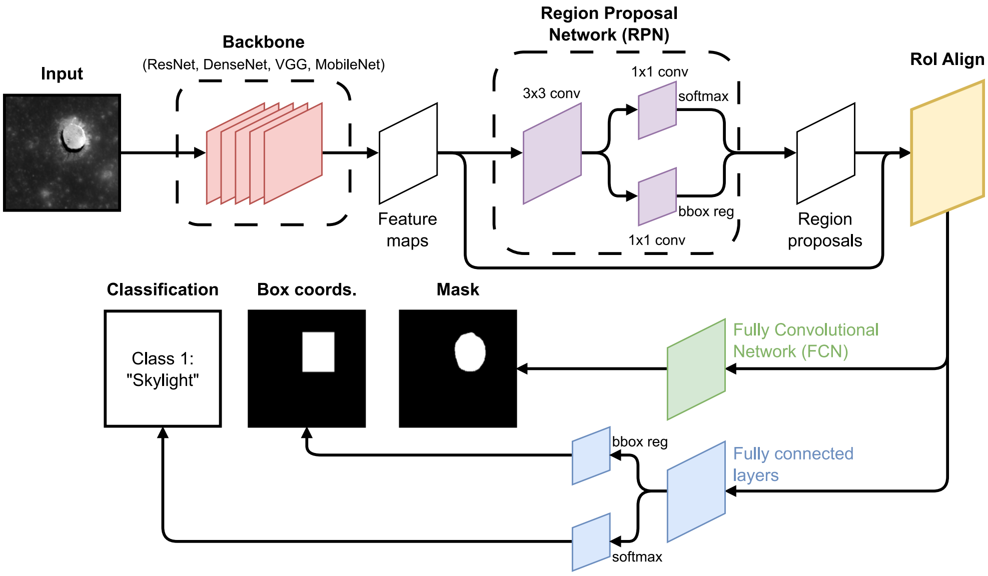

Fig. 4. Diagram of the Mask Regional Convolutional Neural Network (R-CNN) in the context of detecting pits and skylights on the Moon (He et al., 2017). The hypothetical example shows the Mare Tranquillitatis Pit (MTP) being correctly detected as a 'skylight' with a bounding box and predicted mask.

IoU value of ≥65%. IoU is given in Eq. (4), where A and B are the bounding boxes of the detection and label, respectively. Any labelled target features with no detections meeting these criteria are denoted as FNs. Meanwhile, the TP, FP and FN are calculated for the predicted masks by overlaying them with the corresponding label and finding where the pixel classifications agree or not. Of course, this can only be calculated in cases where a mask has been predicted for a feature which has a label (i.e. a TP bounding box detection).

$$\mathrm{P} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{1}$$

$$\mathrm{R} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

$$\mathrm{F}_1 = 2 \times \frac{\mathrm{P} \times \mathrm{R}}{\mathrm{P} + \mathrm{R}} \tag{3}$$

$$\mathrm{IoU} = \frac{|A \cap B|}{|A \cup B|} \tag{4}$$

There are a series of hyper-parameters (such as the optimiser, learning rate, batch size and the number of epochs) which need to be defined before training any DCNNs. Adam with decoupled weight decay (AdamW; Loshchilov and Hutter, 2017) was chosen as the optimiser, with a learning rate of 0.0001. A batch size of 16 was used as it was the maximum value possible with the large tile size and computing resources available. All models were trained for 100 epochs (the number of times all batches have been passed through the model), but this was terminated early if the average between validation F_1-scores for the bounding box and predicted mask worsened after three consecutive epochs.

4. Results and discussion

4.1. Training performance

The extent to which the contribution of Mars and synthetic Moon data improved the Mask R-CNN's ability to detect Lunar pits and skylights required assessment before the models are deployed upon the Lunar RoIs. For this, three Mask R-CNNs were trained (within the parameter space described in Section 3.3 and with the default ResNet50 backbone) upon different combinations of the various data sources. The first Mask R-CNN was trained on genuine and synthetic Moon data, the second on genuine Moon and Mars data, and the third on genuine and synthetic Moon plus Mars data. No model was trained on just genuine Moon data since there would be far too few training samples to lead to a successful model. The F_1-scores achieved by each of these models upon the validation samples (averaged over both target classes) over the course of 100 epochs are shown separately for the bounding boxes and predicted masks in Fig. 5. It is worth noting that no significant improvement was observed when training models with the two target classes merged into one for any of the combinations of training data mentioned above.

Fig. 5 shows that the highest validation F_1 was achieved by the Mask R-CNN trained upon all available datasets — 89.1 and 95.6% for the bounding boxes and predicted masks (at epoch 57), respectively. This was to be expected since there is more data to train with in this instance. Ideally, there would be sufficient examples of Lunar and Martian pits that this comparison could be made independent of dataset sizes. Therefore, it cannot be said whether the addition of Martian and synthetic data would improve the performance of these models in all cases, but it has in this scenario with limited training data.

It is important to state that the entire training/validation dataset was randomly split irrespective of whether the data was of the Moon (genuine or synthetic) or Mars. This meant that a Mask R-CNN trained upon Moon and Mars data was also validated upon data of the Moon and Mars. Therefore, it may be that synthetic Moon or Martian pits and skylights were easier to detect than their genuine Lunar analogues, rather than their inclusion leading to improvements in the overall performance. Nevertheless, the dataset deficiency of those not including Martian or synthetic Moon data is still expected to be the predominant reason for the better performance of the Mask R-CNN trained on all available data. The training and validation samples were also split by image, rather than by pit or skylight, meaning that two separate observations of the same feature could appear in both datasets. However, this is judged not to have affected the relative performance between data sources, since the multiple observations had visibly different sensing conditions and the splitting approach was the same for all three sources.

Following this investigation of the contribution of different data sources was a comparison of the various backbones described in Section 3.3. Fig. 6 plots the validation F_1 of the bounding boxes and predicted masks for each of these backbones — colour-coded according to the respective families: ResNet (pink); DenseNet (brown); VGG (grey) and MobileNetV2 (red). This shows that the ResNet family performed the best in this specific training scheme, with the ResNet50 model achieving the highest F_1-scores, in particular. Incidentally, this is also

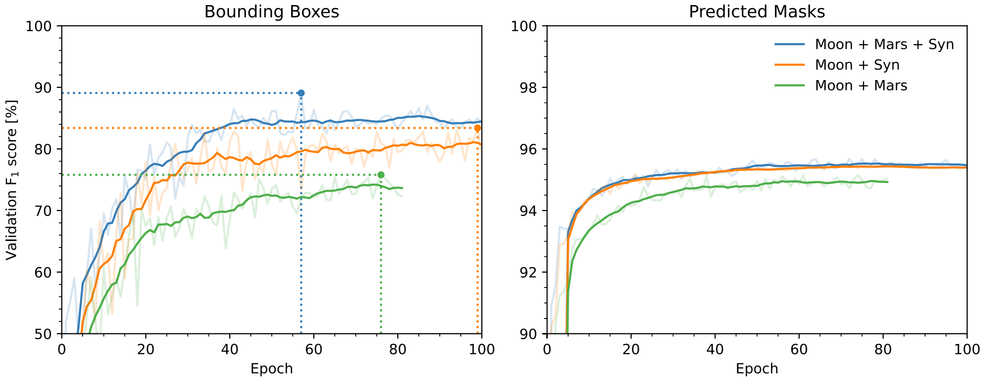

Fig. 5. The F_1-scores (F_1) achieved upon the validation dataset across 100 epochs by three separate Mask R-CNN models trained (all with the default ResNet50 backbone) on different combinations of Moon, Mars and Synthetic Moon data (see Table 1). The solid lines each represent the rolling average over 10 epochs of the fainter profile. The dotted lines annotate the maximum F_1 achieved for the bounding boxes.

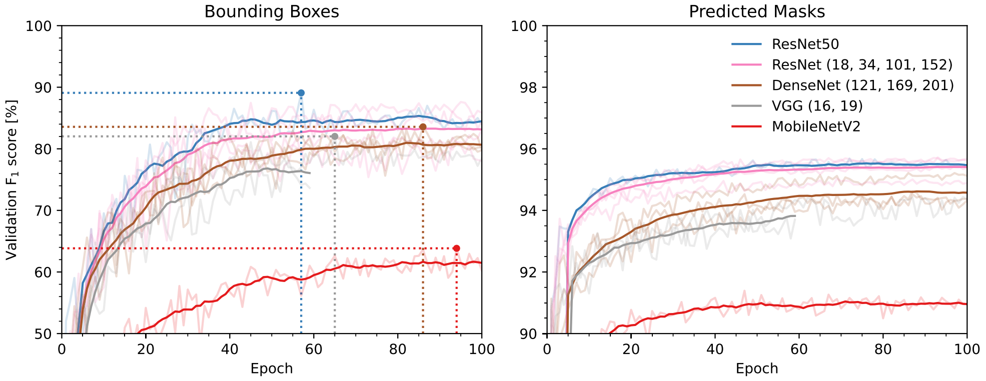

Fig. 6. The F_1-scores (F_1) achieved upon the validation dataset across 100 epochs by several Mask R-CNN models using different backbones for feature map generation. The solid lines each represent the rolling average over 10 epochs of all backbones within a given family (e.g. ResNet), with each fainter profile corresponding to a particular backbone within said family (e.g. ResNet18). The profiles have been colour-coded accordingly: ResNet (pink), DenseNet (brown), VGG (grey) and MobileNetV2 (red). The ResNet50 performance has been plotted in the same blue as in Fig. 5 to highlight that it is the same model. These models were all trained on Moon, Mars and Synthetic Moon data since this achieved the highest performance in Fig. 5.

the same model configuration which performed best in the comparison between different data sources, hence why it has been plotted in the same colour as in Fig. 5. A summary of the performances of each backbone, including a breakdown by target class, can be found in Table A.3 within Appendix A.

The higher performance of the ResNet model family may be in part due to the residual blocks indicative of their architectures. The output of these blocks is the sum of a layer's output and the layer's original input, which allows the model to learn a series of residual functions and eases the training of deeper models. Another factor may be the implementation of a Feature Pyramid Network (FPN) within the ResNet models. FPNs leverage the multi-scale nature of a CNN (wherein image resolution decreases but semantic value increases as successive convolution operations are applied) to output feature maps for all layers (Lin et al., 2016). As a result, the higher performance of the ResNet models could be attributed to how the residual blocks and the FPNs both allow for region proposals of pits and skylights at a range of scales, with greater knowledge of slight differences between other features, such as impact craters.

This result contradicts what was observed by Dawson et al. (2023), which was that the ResNet group of CNNs (ResNet50, 101, 152) performed the worst on average out of those trained upon their smallest dataset of 7000 carbonate core images (i.e. the closest in size to the datasets in this work) - in terms of the overall accuracy $\left(\frac{TP+TN}{TP+TN+FP+FN}\right)$ averaged over all target classes. The best-performing CNNs in Dawson et al. (2023) were, in fact, VGG16 and 19. However, the training of VGG16 in this work was terminated early due to a decrease in performance in three successive epochs. This highlights how the large disparity in the dataset sizes, the use of dissimilar optimisers (with Dawson et al., 2023 testing both stochastic gradient descent and Adam), and the reporting of unalike performance metrics can cause direct comparisons between CNN performances within separate works to be tenuous.

4.2. Testing performance

Since the Mask R-CNN model trained on Lunar, Martian and synthetic Moon data with a ResNet50 backbone has been shown in Section 4.1 to achieve the highest average F_1 upon the validation dataset, this model now requires testing on the data described in Section 2.2 to get a picture of its real-world accuracy before it is inferred upon the data of the eight Lunar RoIs. We refer to this particular model as ESSA (Entrances to Sub-Surface Areas), which is also the Cornish (*Kernewek*) name for the corresponding author's home-town of Saltash, Cornwall ('*Essa*').

Fig. 7 gives a visual representation of the results of applying ESSA on this data — which contains the Gruithuisen, Marius A and Marius B lava tube collapse chains proposed by Sauro et al. (2020), as well as five separate LROC NAC images of the MTP under different illumination conditions. It should be noted that some figures in this work display detection confidence scores of 100.0%. This is not an assertion that the probability of the detection being incorrect is zero, but rather that they were made with a score of 99.95% or more. The results (calculated with

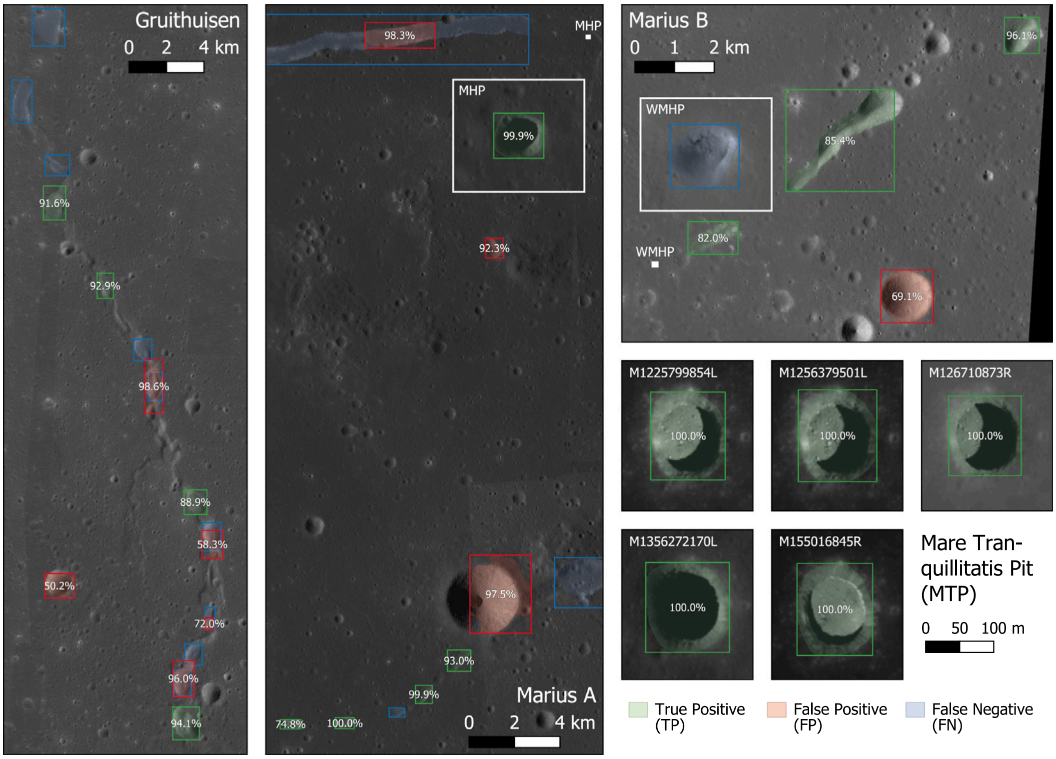

Fig. 7. Visual performance of ESSA on Lunar pits and skylights. The bounding boxes and predicted masks are overlaying self-produced mosaics of candidate lava tube collapses (Gruithuisen, Marius A and B) and five LROC NAC observations of the Mare Tranquillitatis Pit (MTP). True positives (TPs) are those with the correct classification, a confidence score ≥50%, and an IoU with the label ≥65%.

the same confidence and IoU thresholds as during training) equate to an F_1 of 82.4 and 93.7% for the bounding boxes and predicted masks, respectively. The F_1 for the bounding boxes was first calculated for each site individually and then averaged (as to negate the inequality in the number of targets), whereas for the masks, they were averaged globally. Whilst the P and R of the predicted masks remained comparably high (>91% for both), there was some difference between the P (86.3%) and R (78.8%) of the bounding boxes.

These discrepancies between the training and testing performance metrics were expected as the testing data contained several features which were likely to be more confusing for the model. For example, the WMHP exhibits no visible shadow within the Marius B mosaic. This partly explains why it was missed (i.e. not correctly detected as a skylight, nor incorrectly as a pit), as one of the main characteristics of a pit or skylight is a shadow. Several of the other missed pits in the Gruithuisen and Marius A chains also lacked a shadow. This implies that there may be a lower bound for the α (≈30°) that an image can have for it to be passed to ESSA. Moreover, impact craters on the same scales as the ≈3 km-diameter example detected as an FP within Marius A were rarely seen during training. Five out of the nine FPs Fig. 7 also overlap a FN, but by an insufficient IoU.

Nonetheless, the MTP was detected in all five LROC NAC images with a score >99.9%, with no FPs found elsewhere in any of the images. Meanwhile, the MHP was confidently detected within the Marius A mosaic despite the fact that it was completely covered by shadow. Sufficient pits have also been detected in the proposed lava tube collapses such that a sinuous alignment can be seen, which is necessary to be able to search for any aligned chains in the RoIs. There were also no examples of any FPs being caused by misclassification (i.e. a pit being detected as a skylight, and vice-versa). Furthermore, Fig. B.12 in Appendix B shows how ESSA correctly recalled all 12 LPA mare pits and skylights when it was applied to new LROC NAC observations with different sensing conditions compared to those seen during training. These results were not included in the analysis of this section since these features were seen in some form during training.

This testing has also allowed for reassessment of the confidence score threshold. A higher score threshold would reduce the number of FPs and, hence, preserve the time taken to assess any such detections manually. Although, this should not be at the expense of missing too many true pits or skylights. Fig. 8 plots the change in the P, R and F_1 of both target classes detected and shown in Fig. 7 against an increasing score threshold. This shows that the P and R both decrease with an increasing score threshold. The decrease in R is most obvious since it means that TPs need to be detected with higher and higher scores to remain as such. However, the decrease in P is a result of this reduction in the TPs, despite some high-scoring FPs still meeting the threshold. A score threshold of 80% appears to be the loosest approximate threshold before which R begins to decrease noticeably, leading to true pits or skylights to be missed. With a stricter threshold of 80%, the 17 TPs shown in Fig. 7 were reduced to 16, whilst the FPs also decreased from 9 to 5. Therefore, detections with a confidence score below 80% were removed when applying ESSA to the eight RoIs.

4.3. *Inference upon eight regions of interest on the moon*

Across the eight Lunar RoIs, the model detected a total of 420 features — 116 as class 1 (skylights) and 304 as class 2 (pits). These detections required manual inspection to confirm whether they were genuine Lunar pits or skylights. Although, it should be stated that many of these detections were of the same features across multiple overlapping tiles or were separate sections of the same elongated pit or rille.

The same morphological descriptions used during labelling were also used during this manual inspection. Other factors, such as the Sun's position relative to the surface, were also considered. The Sun's position was not knowable by ESSA when making detections, which makes sense of some of the FPs observed. Examples of some of the common false positives were small and/or debris-ridden impact craters,

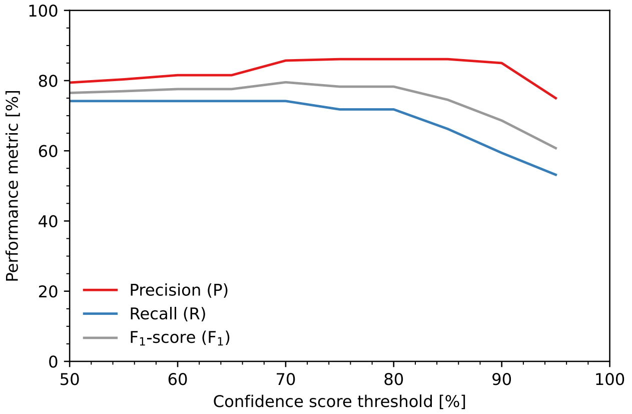

Fig. 8. A plot of the change in the precision (P), recall (R) and F_1-score (F_1) of the testing results displayed in Fig. 7 due to an increasing score threshold. The P, R and F_1 were calculated individually (with an IoU threshold of 65%) for each image, then averaged across all images and both target classes.

wrinkle ridges, and boulders, which were not frequently seen during training.

Fig. 9 presents some of the elongated pits and Lunar rilles, which were correctly detected by the model across the RoIs, including a mapping of such features across the entire Marius Hills region. Regarding Marius Hills, applying ESSA to all 4905 image tiles covering the region took only 4.35 mins (using one of 13 GPU nodes which share an NVIDIA A100 40 GB PCIE GPU and 256 GB of RAM), proving that it can analyse data at far quicker time-scales than could be reached through manual methods. Unfortunately, no new clear instances of linearly or sinuously aligned pits, suggesting the presence of a partially collapsed lava tube, were detected by ESSA within the RoIs. However, given the relatively small sample size considered in this work when compared to the total Lunar maria, this is not to say that no more sinuous pit chains exist to be found. Despite this, Fig. 9 shows how ESSA was able to detect the MHP and WMHP within the vicinity of the Marius A and B collapse chains, despite the WMHP being missed in Section 4.2.

Fig. 9 also presents the location (at 13.0188°N and 56.9322°W) of a previously uncatalogued skylight detection made by ESSA within the Marius Hills region. This is intriguing since there was LROC NAC imagery of the feature available (albeit with lesser coverage) at the time when PitScan was used in the creation of the LPA (Wagner and Robinson, 2021). Due to its proximity (within ≈4.3 km) to the Marius A proposed lava tube collapse, it is also surprising that its existence was not mentioned by Sauro et al. (2020). Fig. 10 displays this skylight, which we call the South Marius Hills Pit (SMHP) in convention with the LPA, within the three LROC NAC images that it was detected within by ESSA — all with scores above 96.1%. SMHP is an elliptical skylight, with semi-major and minor axis of roughly 80 and 60 m, respectively.

A possible explanation for why the SMHP has been missed by previous investigations in this area is exhibited by the images in Fig. 10 - in that there is very little contrast between its floor and the surrounding surface. Therefore, the sensitivity of the PitScan tool may have been too weak to detect the shadow as being caused by a depression. One instance in which this is not true is M193332423R, shown in Fig. 10, which was not likely to have been passed to PitScan due to its significant α (75.6°). In this near-nadir image (ϵ = 1.2°), the westward direction of the Sun's illumination reveals that its western wall is not visible, which may mean that it is overhanging or providing access to a sub-surface cavity.

The PITS tool (see Le Corre et al., 2023 for more details) has been applied to two images, including M193332423R, containing the SMHP. The resulting h profiles, which are given in Fig. C.13 in Appendix C, both suggest a bowl-shaped floor with a maximum h between $14.0^{+0.1}_{-0.7}$ and $15.4^{+0.1}_{-0.8}$ m. With the Sun's illumination coming from opposite directions and the shadow being cast to a similar location within the floor, it can be said that the rims to the east and west of SMHP are approximately level. The northeastern rim of the SMHP may also provide the safest access to any robotic or crewed explorers with an average slope of ≈31°down towards the floor.

Aside from Marius Hills, another skylight was detected by ESSA within Bel'kovich A crater. At first glance, this appeared as an impact crater itself, due to the wide (≈170 m in diameter) funnel around the detection and the fact that shallow depressions at such high latitudes can exhibit prominent shadows. However, upon closer inspection of the highest resolution LROC NAC image available (M128727218L), shown in Fig. 11, it appears as though the western flank of the shadow is being cast against a steep or overhanging wall due to the high reflectance immediately opposite the shadow edge. Therefore, we name this feature the Bel'kovich A Pit (BAP), which is the closest to the Lunar North Pole of any skylight found within a mare unit (at 58.6979°N, 87.5990°E).

As is also true for the SMHP, further observations at high resolutions, with different sensing conditions, are required to determine whether the BAP potentially harbours a cave entrance. However, this is unlikely to be possible for the BAP with current visual imagery datasets due to the periapsis of the LRO's elliptical orbit now being at the Lunar South Pole and the BAP being at a latitude where the sub-solar point at noon will always be to its south. Therefore, other data sources, such as infrared or radar imagery, may need to be employed.

Nonetheless, two LROC NAC observations of the BAP have been passed through the PITS tool. One of the images (M1233723761R) has been cropped to the extent of the BAP's funnel since the low Sun angle means that the shadow's edge falls within it. From these h profiles, which are also shown in Fig. C.14 in Appendix C, we can deduce that the combined h of the BAP and its surrounding funnel is ≈26 m. With consideration for the fact that the Sun is casting the shadow from different points along its length, the maximum h of the BAP (minus the funnel) was $14.0^{+0.1}_{-0.7}$ m. There is also a steep slope (≈46°) from the eastern rim towards the possible overhang in the west.

5. Conclusions

5.1. Summary

In this work, we have trained a series of Mask R-CNN (Region-based Convolutional Neural Network; He et al., 2017) instance segmentation Deep Learning (DL) models to detect Lunar pits and skylights. This was with the aim of finding such features that may act as potential entrances to sub-surface cavities, such as intact lava tubes. The various Mask R-CNN models were trained upon different combinations of LROC NAC and MRO HiRISE imagery of the Moon and Mars, and a range of backbone models from the ResNet, DenseNet, VGG and MobileNet families. Images were labelled using various catalogues of pits and skylights on the Moon and Mars, such as the Lunar Pit Atlas (LPA; Wagner and Robinson, 2021), Mars Global Cave Candidate Catalog, (MGC^3; Cushing, 2015), lava-tube-related rilles mapped by Hurwitz et al. (2013), and potential cave entrances found by Watson and Baldini (2024). Synthetic Moon training data was also generated by implanting warped HiRISE pixels covering Martian pits and skylights into empty LROC NAC images.

During validation, it was found that the highest average F_1-scores (F_1) of the bounding boxes and predicted masks (89.1 and 95.6%, respectively) were achieved by the Mask R-CNN when it was trained on Lunar, Martian and synthetic Lunar data with a ResNet50 backbone. This particular model, named ESSA (Entrances to Sub-Surface Areas), then underwent testing upon self-produced mosaics of LROC NAC observations of the Gruithuisen, Marius A and Marius B lava tube candidates proposed by Sauro et al. (2020), as well as five LROC NAC images containing the Mare Tranquillitatis Pit (MTP). This testing led to an average F_1 of 82.1 and 93.7% for the bounding boxes and predicted masks, respectively, with the MTP being detected in every image with

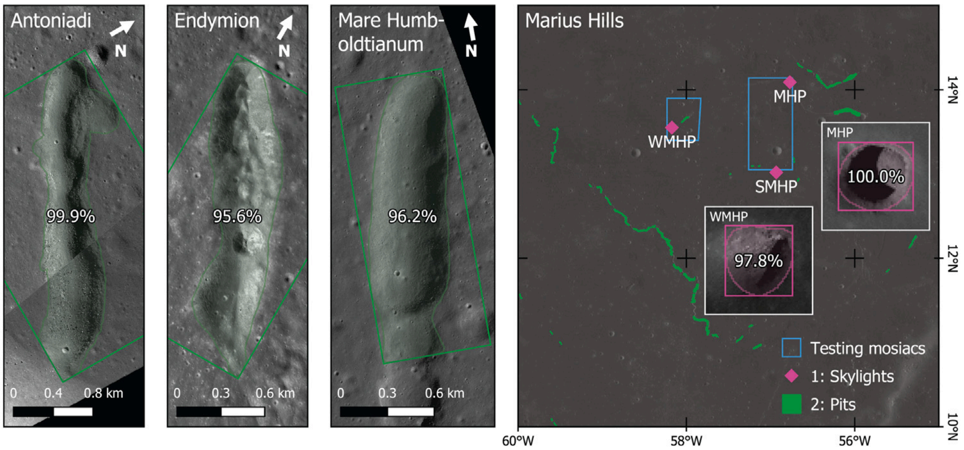

Fig. 9. Examples of some of the elongated pits and rilles detected in four out of the eight RoIs. The map of Marius Hills plots the detected pits (in green) and skylights (in pink) upon the LROC WAC Global Morphology Mosaic (Speyerer et al., 2011). The inset maps zoom in on the detections of the Marius Hills and West Marius Hills skylights (MHP and WMHP). The footprints of the Marius A and B mosaics used in testing (see Fig. 7) are also given in blue.

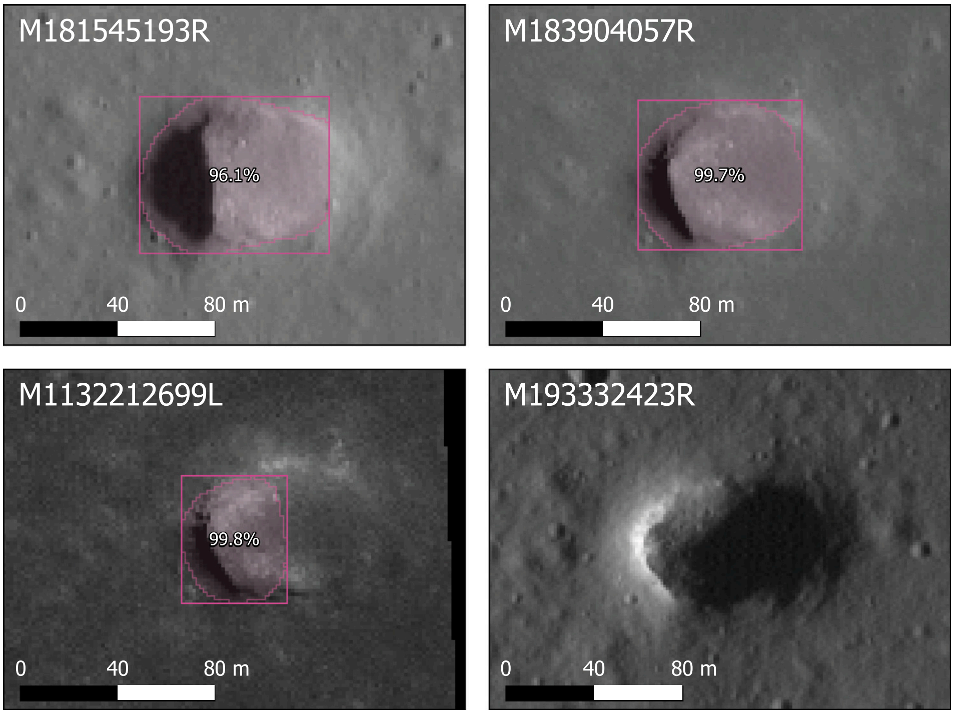

Fig. 10. Four LROC NAC images of the skylight detected by ESSA within the Marius Hills region of the Moon, nicknamed the 'South Marius Hills Pit' (SMHP). (Top row, bottom left) The LROC NAC images that the SMHP was detected within. (Bottom right) An observation of the SMHP with the Sun's illumination coming from the opposite direction and revealing a possible over-hanging rim.

high confidence. It also highlighted that some features that were not seen during training, such as kilometre-scale impact craters, commonly led to false positive detections.

Following on from this, ESSA was applied to LROC NAC imagery covering eight regions of the Moon's mare deposits (six being impact craters or basins and two being Lunar maria themselves). Considering the rarity of skylights on the Moon and the fact that it was applied to just 1.92% of the total maria, two previously uncatalogued skylights have been detected by ESSA, which could prove to be potential cave entrances upon further investigation. Moreover, ESSA has achieved these findings within time scales that would not have been reachable through manual means.

The first, named the South Marius Hills Pit (SMHP), was detected in a region which had already been surveyed by the LPA and is within ≈4.3 km of the Marius A proposed lava tube collapse analysed by Sauro et al. (2020). Therefore, this particular area of Marius Hills represents a potential site of interest for space missions, such as Moon Diver, which are tasked with investigating cave entrances.

The second, situated within the satellite crater Bel'kovich A, appears as though the prominent shadow cast within it is adjacent to an overhanging rim. However, due to the high latitude of the feature, the Sun's illumination will only cast northwards, meaning that observations with other data types may be required to confirm this. This shows that ESSA has extended the latitude range within which the search for cave

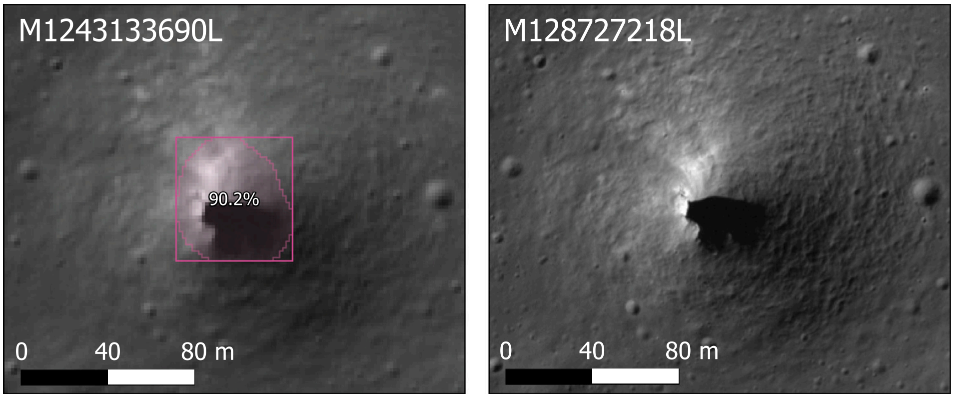

Fig. 11. Two LROC NAC images showing the skylight detected by ESSA within Bel'kovich A crater, nicknamed the Bel'kovich A Pit (BAP). (Left) The LROC NAC image M1243133690L (downscaled to 1.5 m/px) that the feature was detected within. (Right) The highest resolution LROC NAC image available for this feature, highlighting the possible overhanging rim.

entrances can be carried out, but that there is also scope for repeat surveys of areas already thought to have been completed.

Instructions for inferring ESSA upon a user's own data can be found on GitHub (Le Corre, 2025a), with the learnt model weights being available as a PyTorch model checkpoint on Zenodo (Le Corre et al., 2025). The same Zenodo repository also contains a list of the LROC NAC and MRO HiRISE products used in the training of ESSA, as well as geo-referenced shapefiles of the locations of the detections made by ESSA within the eight Lunar RoIs. The above image quantities differ slightly from Table 2 as tiles could not be generated in some cases due to proximity to no-data regions. Instructions on how to set up PITS are also hosted on GitHub (Le Corre, 2025b). Bash scripts for processing batches of LROC NAC EDR products and correcting the map projection of HiRISE RDRV11 images are publicly available on GitHub (Le Corre, 2025c).

5.2. Limitations

The main limitation that requires clarification is that no pits, skylights or intact lava tubes have been visited by robotic or crewed missions on the surface of the Moon or Mars to date. Therefore, any categorisation of these features has been made on the basis of remote observations. For this reason, terms such as 'candidate', 'proposed', or 'possible' are frequently used to try to make this distinction.

Secondly, due to the downscaling of LROC NAC and HiRISE data to a common resolution of 1.5 m/px, catalogued features from the MGC^3 or the LPA that were less than ≈10 m in diameter were barely resolvable, meaning that labelling them could lead to small clusters of pixels being falsely detected. Therefore, due to the non-uniformity of resolutions found within planetary remote-sensing datasets, there is often a trade-off between spatial detail and the ability to produce more training data.

Despite ESSA being able to survey LROC NAC data at speeds unachievable through manual means, a bottleneck in the workflow was caused by the processing of planetary remote-sensing data. LROC NAC image products are not publicly available in a GIS-ready format, and developing ESSA for use with LROC NAC CDR (calibrated, but not map-projected) was avoided due to potential clashes with the addition of MRO HiRISE RDRV11 data with its embedded map projection. It was also a requirement for all input data fed to the PITS tool, including imagery, to be geo-referenced. This meant that, whilst it only took 4.35 mins for ESSA to survey Marius Hills, it took over 10 h to process and map-project the 446 LROC NAC EDR images covering it. This is not, by any means, a reason to avoid using automated tools, such as ESSA, since this would still be the case if the same task were performed manually. However, for ESSA to be applied to all Lunar mare deposits globally, it may be necessary for it to be inferred (and possibly retrained) upon LROC NAC CDR data. Any detections made by it could then have their locations converted from lines/samples to latitude/longitude coordinates.

5.3. Future work

Since ESSA has surveyed just ≈0.23% of the Moon's surface so far, there are still vast amounts of data to which it can be applied. In the context of searching for pits and skylights which relate to potential cave entrances, mare regions should still be prioritised for being fed to ESSA. Up to now, ESSA has largely been applied to smaller mare deposits, which have well-defined boundaries, such as those within impact craters. However, with sufficient time allocated for the processing of imagery, ESSA could search for pits and skylights within some of the larger Lunar maria (such as Mare Frigoris) by iterating through latitude–longitude intervals.

With the detections made by ESSA in this work, such as the SMHP/BAP skylights, these could all be used as training data to further improve its performance. Since these features have been imaged several times by LROC, the different illumination conditions could help to bolster the dataset. The same is also true of the detected elongated pits or rilles, which could be another source of additional training data. In fact, the detections made in this work could be used to train Mask R-CNNs or similar models to specifically map Lunar rilles at pixel-level scale. Or rather, the learnt weights within ESSA could be transfer-learnt to the task of solely mapping Lunar rilles at the scale of LROC NAC imagery.

As highlighted by the reviewer, a potential way of further improving ESSA's performance would be to train it upon imagery where the Sun's illumination direction is constant (i.e. aligned vertically or horizontally within the image). A new dataset of synthetic data could also be generated where the Martian features being implanted are also azimuth-matched to have the Sun's illumination direction remain uniform. This could potentially reduce FP detections of positive elevation features, such as boulders, where the brightness gradient across them would differ from a depression. PitScan utilised a similar concept in its automated workflow, whereas this would tailor it for use within the training of a DL model.

Since it was partly trained on MRO HiRISE imagery, the model used in this work for detecting Lunar pits and skylights could also be transferred to the surface of Mars. However, there is less global coverage of Mars, with imaging sensors having comparable resolutions to that of LROC NAC images (McEwen et al., 2024). Similar to how the model has been used to map areas which had not been scanned by PitScan, the model could be periodically applied to new releases of HiRISE imagery to search for Martian cave entrances as the surface coverage increases.

CRediT authorship contribution statement

Daniel Le Corre: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Nigel Mason:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **Jeronimo Bernard-Salas:** Writing – review & editing, Supervision, Project administration, Funding acquisition. **David Mary:** Writing – review & editing, Supervision. **Nick Cox:** Writing – review & editing, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This project is part of the Europlanet 2024 RI which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 871149. Model training, testing and inference, as well as batch image processing, were all enabled thanks to the CPU and GPU nodes within the Specialist and High Performance Computing systems provided by Information Services at the University of Kent.

Appendix A. Training performances of various mask R-CNN backbones

Table A.3 presents the average precision (P), recall (R) and F_1-scores (F_1) of the bounding box detections upon the validation dataset for all Mask R-CNN backbones tested in this work. As explained in Section 4.1, these models were all trained on genuine Moon (LROC NAC) and Mars (MRO HiRISE) imagery, as well as synthetic Moon tiles. Table A.3 only displays the bounding box performance, and not that of the predicted masks, since it best describes the ability of these models to find pits and skylights, and the performance of the latter was high (F_1 > 90%) for all backbones. These performance metrics have been averaged for each of the two target classes individually, as well as globally over the entire validation dataset (as was the case in Figs. 5 and 6). The particular epoch at which these metrics were achieved was when the average between the bounding box and predicted mask F_1-scores was at its peak over the 100 epochs of training. The same thresholds for the confidence score (50%) and IoU overlap (65%) as described in Section 3.3 were also used to derive these metrics.

Appendix B. Applying ESSA to LPA mare features

Due to the scarcity of known Lunar pits and skylights found within mare deposits, the majority of the LPA mare features were used as training data, leaving only the MTP, WMHP and MHP for use as independent testing data. Despite this, LROC NAC images with different sensing conditions (in terms of the α and/or the solar azimuth angle) compared to those used for training were acquired for each of these 12 LPA mare features. This allowed for a gauging of ESSA's ability to recall known pits and skylights within the regions that it is to be applied to — the Lunar maria. Fig. B.12 shows how ESSA detected all 12 LPA mare features with the correct classes and often with scores over 99.95%. The only feature which was detected with a score below the stricter threshold of 80% (see Section 4.2) was the Runge skylight, which itself is less than 30 m in diameter. These results were not included within the testing performance analysis of Section 4.2 since these features were seen in some form during training. Nonetheless, these additional results add credence to ESSA's ability to recall known pits and skylights within the Lunar maria.

Appendix C. PITS apparent depth profiles for SMHP and BAP

Fig. C.13 plots the apparent depth (h) profiles derived by PITS (Le Corre et al., 2023) upon two LROC NAC images of the South Marius Hills Pit (SMHP). The first image (M181545193R), as shown in Fig. 10, is one of the observations within which the SMHP was detected by ESSA, with the Sun illuminating it in an approximately westward direction. M181545193R was able to be cropped to the extent of the SMHP using the geo-referenced bounding box automatically detected by ESSA. The second image (M193332423R), also shown in Fig. 10, has the Sun's illumination coming from the east, but the shadow edges in both images are cast to roughly the same spot within its floor.

Fig. C.14 plots the h profiles calculated by PITS upon two LROC NAC images of the Bel'kovich A Pit (BAP). The first image (M1233723761R) has been cropped in such a way as to include the wide funnel surrounding the BAP. The extreme α (83.8°) in M1233723761R means that a shadow is being cast by the funnel's edge, as opposed to that of BAP's steep wall. This allows for PITS to calculate a h profile for the funnel and, as a result, to also get an estimate of the total depth of the BAP relative to the immediate terrain. The second image (M128727218L), shown in Fig. 11, is the highest resolution observation available of the BAP. The shadow exhibited in M128727218L is now clearly being cast by the BAP's steep wall.

Data availability

Both the ESSA model and the necessary code required to infer it on new data have been linked to within this article.

Table A.3
Average performance metrics of the bounding boxes achieved upon the validation dataset when changing the backbone of the Mask R-CNN model. The precision (P), recall (R) and F_1-score (F_1) have been averaged individually for each target class (1: skylight and 2: pit), as well as globally over all samples. The maximum of each column has been highlighted in bold.

Backbone	Epoch	Class 1: Skylight			Class 2: Pit			Global Average		
		P [%]	R [%]	F_1 [%]	P [%]	R [%]	F_1 [%]	P [%]	R [%]	F_1 [%]
ResNet18	76	88.8	86.3	87.5	71.9	84.6	77.7	80.3	85.4	82.8
ResNet34	93	81.8	81.3	81.6	82.6	81.6	82.1	82.2	81.5	81.8
ResNet50	57	**90.9**	91.4	**91.2**	87.9	86.2	87.0	**89.4**	88.8	**89.1**
ResNet101	73	90.4	87.4	88.9	71.6	**96.2**	82.1	81.0	91.8	86.1
ResNet152	95	84.2	**93.4**	88.5	82.1	91.8	86.7	83.2	**92.6**	87.6
DenseNet121	97	69.0	90.3	78.2	82.2	88.0	85.0	75.6	89.2	81.8
DenseNet169	86	73.7	80.7	77.0	**92.1**	87.8	**89.9**	82.9	84.3	83.6
DenseNet201	82	73.2	84.1	78.3	82.5	94.4	88.0	77.8	89.2	83.1
VGG16	54	63.1	83.0	73.1	71.8	88.8	83.0	71.6	85.9	78.1
VGG19	65	81.4	85.5	83.4	73.4	89.0	80.4	77.4	87.2	82.0
MobileNetV2	94	50.4	59.7	54.7	68.6	78.0	73.0	59.5	68.9	63.8

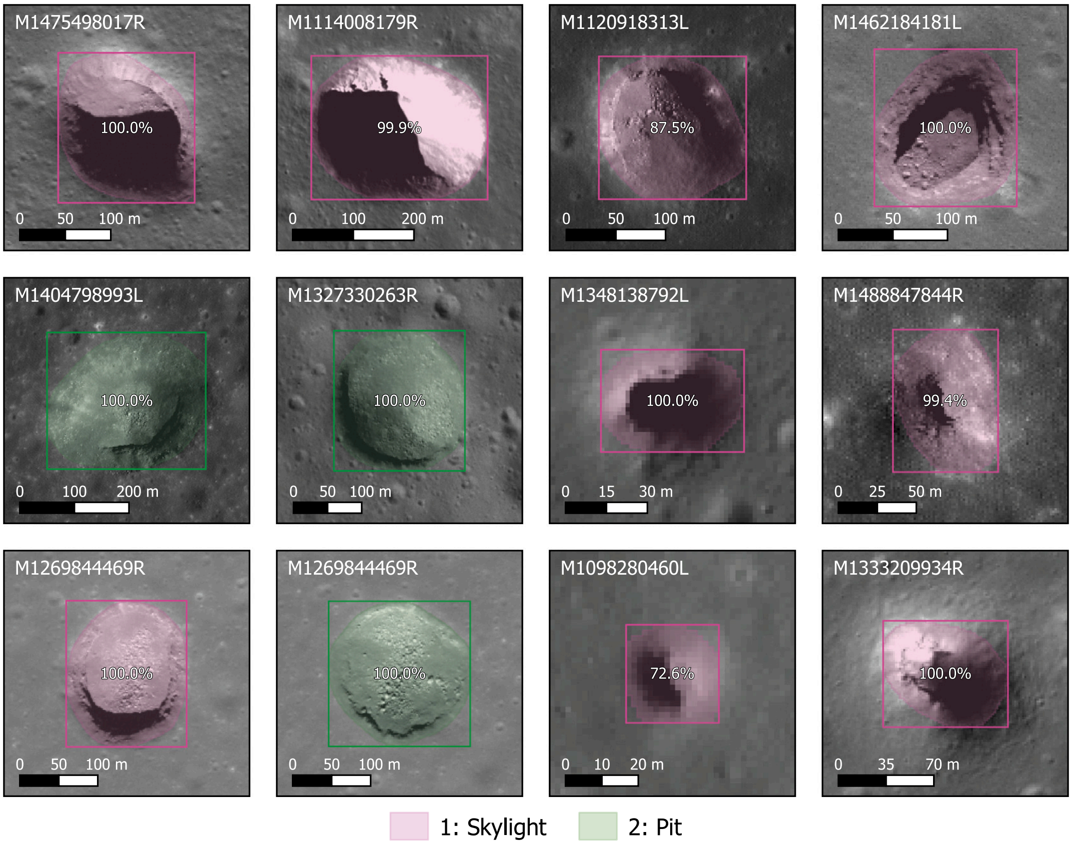

Fig. B.12. 12 LPA mare pits (green) and skylights (pink) successfully detected by ESSA. The LPA names for these features are as follows (from left to right, row by row): Compton, Lacus Mortis, Central Fecunditatis, Mare Ingenii, Mare Insularum, Mare Moscoviense, Mare Serenitatis, Southwest Mare Tranquillitatis, North Procellarum 1 and 2, Runge, and Sinus Iridium.

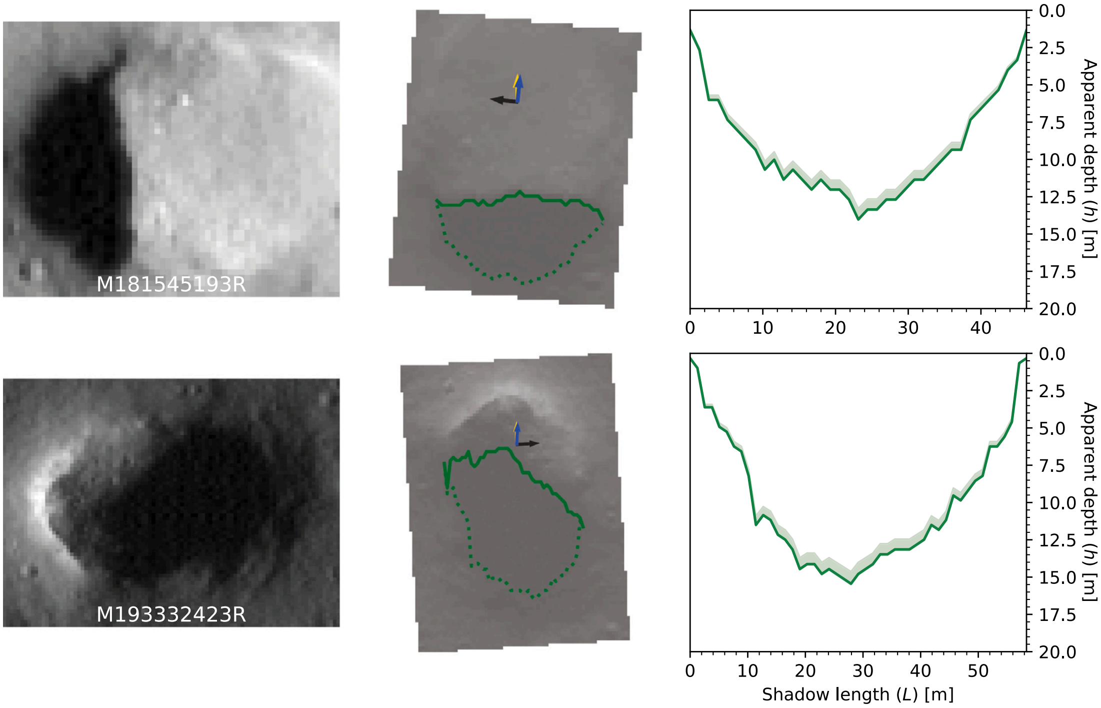

Fig. C.13. PITS apparent depth (h) profiles calculated upon two LROC NAC observations of the South Marius Hills Pit (SMHP). (Left column) The cropped input images fed to the PITS tool. (Middle column) The input images aligned to the Sun's line of sight such that it passes from bottom to top, overlaid with the automatically detected shadow edge (solid) and pit rim (dashed). The arrows in the middle column represent the directions of the Sun's illumination (yellow), LRO's sensing (blue), and north (black). (Right column) The h profiles taken from left to right along the solid green line in the middle column. The x and y axes of the h profiles are not to scale.

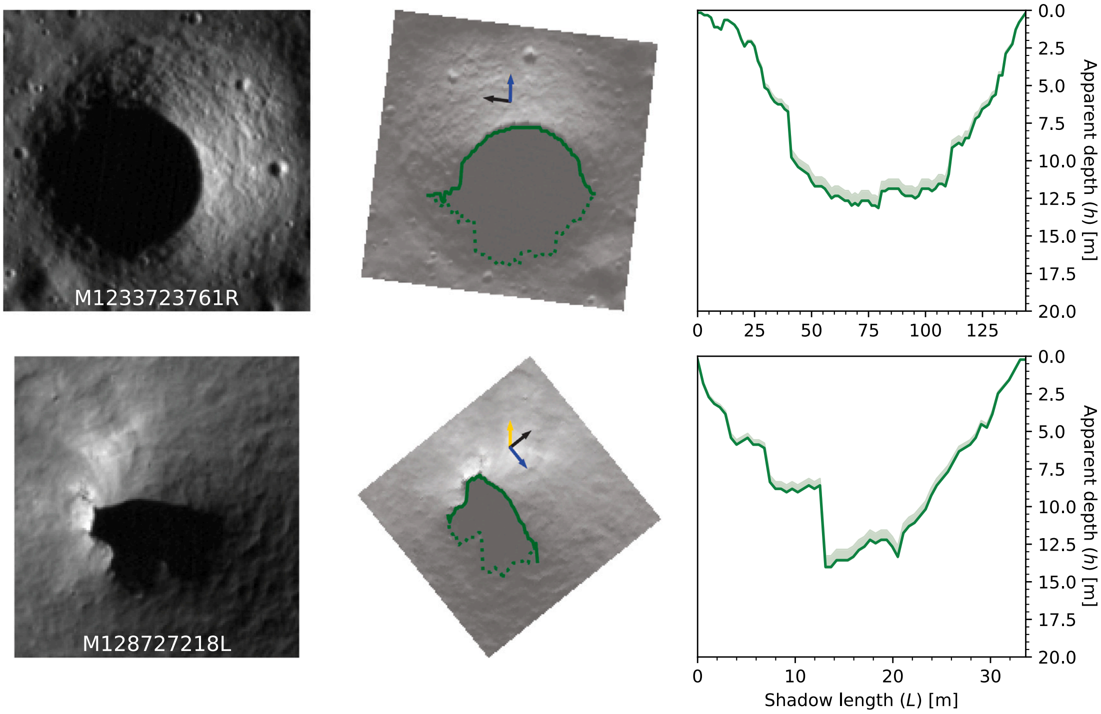

Fig. C.14. PITS apparent depth (*h*) profiles derived from two LROC NAC images of the Bel'kovich A Pit (BAP). The symbology is the same as in Fig. C.13.

References

Ali-Dib, M., Menou, K., Jackson, A.P., Zhu, C., Hammond, N., 2020. Automated crater shape retrieval using weakly-supervised deep learning. Icarus 345, 113749. http://dx.doi.org/10.1016/j.icarus.2020.113749.

Alshehhi, R., Gebhardt, C., 2022. Detection of martian dust storms using mask regional convolutional neural networks. Prog. Earth Planet. Sci. 9 (1), 4. http://dx.doi.org/10.1186/s40645-021-00464-1.

Atri, D., et al., 2022. Crewed missions to Mars: Modeling the impact of astrophysical charged particles on astronauts and their health. http://dx.doi.org/10.48550/arXiv.2208.00892, ArXiv E-Prints.

Barker, M.K., Mazarico, E., Neumann, G.A., Zuber, M.T., Haruyama, J., Smith, D.E., 2016. A new lunar digital elevation model from the lunar orbiter laser altimeter and SELENE terrain camera. Icarus 273, 346–355. http://dx.doi.org/10.1016/j.icarus.2015.07.039.

Blair, D.M., et al., 2017. The structural stability of lunar lava tubes. Icarus 282, 47–55. http://dx.doi.org/10.1016/j.icarus.2016.10.008.

Boston, P.J., et al., 2004. Extraterrestrial subsurface technology test bed: Human use and scientific value of martian caves. In: El-Genk, M.S. (Ed.), Space Technology and Applications. In: American Institute of Physics Conference Series, vol. 699, AIP, pp. 1007–1018. http://dx.doi.org/10.1063/1.1649667.

Calvari, S., Pinkerton, H., 1999. Lava tube morphology on etna and evidence for lava flow emplacement mechanisms. J. Volcanol. Geotherm. Res. 90 (3), 263–280. http://dx.doi.org/10.1016/S0377-0273(99)00024-4.

Carrer, L., Pozzobon, R., Sauro, F., Castelletti, D., Patterson, G.W., Bruzzone, L., 2024. Radar evidence of an accessible cave conduit on the moon below the Mare Tranquillitatis pit. Nat. Astron. 8, 1119–1126. http://dx.doi.org/10.1038/s41550-024-02302-y.

Chappaz, L., et al., 2017. Evidence of large empty lava tubes on the moon using GRAIL gravity. Geophys. Res. Lett. 44 (1), 105–112. http://dx.doi.org/10.1002/2016GL071588.

Christensen, P.R., et al., 2009. JMARS - a planetary GIS. In: AGU Fall Meeting Abstracts. pp. IN22A–06, URL: https://jmars.asu.edu/.

Coombs, C.R., Hawke, B.R., 1992. A search for intact lava tubes on the moon: Possible lunar base habitats. In: Mendell, W.W., Alred, J.W., Bell, L.S., Cintala, M.J., Crabb, T.M., Durrett, R.H., Finney, B.R., Franklin, H.A., French, J.R., Greenberg, J.S. (Eds.), Lunar Bases and Space Activities of the 21st Century. p. 219.

Crown, D.A., Scheidt, S.P., Berman, D.C., 2022. Distribution and morphology of lava tube systems on the Western Flank of Alba Mons, Mars. J. Geophys. Res. (Planets) 127 (6), e07263. http://dx.doi.org/10.1029/2022JE007263.

Cushing, G.E., 2015. Mars global cave candidate catalog PDS4 archive bundle. http://dx.doi.org/10.17189/1519222, PDS Cartography and Imaging Sciences Node (IMG). URL: https://astrogeology.usgs.gov/search/map/mars_global_cave_candidate_catalog_v1_cushing.

Cushing, G.E., Okubo, C.H., Titus, T.N., 2015. Atypical pit craters on Mars: New insights from THEMIS, CTX, and HiRISE observations. J. Geophys. Res. (Planets) 120 (6), 1023–1043. http://dx.doi.org/10.1002/2014JE004735.

Cushing, G.E., Titus, T.N., Wynne, J.J., Christensen, P.R., 2007. THEMIS observes possible cave skylights on Mars. Geophys. Res. Lett. 34, 17201. http://dx.doi.org/10.1029/2007GL030709.

Davey, S.C., Ernst, R.E., Samson, C., Grosfils, E.B., Willamson, M.-C., 2013. Hierarchical clustering of pit crater chains on venus. Can. J. Earth Sci. 50 (1), 109–126. http://dx.doi.org/10.1139/cjes-2012-0054.

Dawson, H.L., Dubrule, O., John, C.M., 2023. Impact of dataset size and convolutional neural network architecture on transfer learning for carbonate rock classification. Comput. Geosci. 171, 105284. http://dx.doi.org/10.1016/j.cageo.2022.105284.

Ferrill, D.A., 2015. Pit. In: Hargitai, H., Kereszturi, A. (Eds.), Encyclopedia of Planetary Landforms. Springer New York, New York, NY, pp. 1566–1570. http://dx.doi.org/10.1007/978-1-4614-3134-3_267.

Fortezzo, C.M., Spudis, P.D., Harrel, S.L., 2020. Release of the digital unified global geologic map of the moon at 1:5,000,000-scale. In: 51st Annual Lunar and Planetary Science Conference. In: Lunar and Planetary Science Conference, p. 2760.

Gadányi, P., van der Bogert, C.H., 2015. Lava tube. In: Hargitai, H., Kereszturi, A. (Eds.), Encyclopedia of Planetary Landforms. Springer New York, New York, NY, pp. 1181–1186. http://dx.doi.org/10.1007/978-1-4614-3134-3_474.

GDAL/OGR contributors, 2025. GDAL/OGR geospatial data abstraction software library. http://dx.doi.org/10.5281/zenodo.5884351, URL: https://gdal.org.

Gillis-Davis, J.J., et al., 2009. Pit-floor craters on Mercury: Evidence of near-surface igneous activity. Earth Planet. Sci. Lett. 285 (3–4), 243–250. http://dx.doi.org/10.1016/j.epsl.2009.05.023.

Girshick, R., 2015. Fast R-CNN. http://dx.doi.org/10.48550/arXiv.1504.08083, ArXiv E-Prints.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2013. Rich feature hierarchies for accurate object detection and semantic segmentation. http://dx.doi.org/10.48550/arXiv.1311.2524, ArXiv E-Prints.

Greeley, R., 1971. Lava tubes and channels in the Lunar Marius Hills. Moon 3 (3), 289–314. http://dx.doi.org/10.1007/BF00561842.

Greeley, R., Gault, D.E., 1979. Endogenic craters on basaltic lava flows: size frequency distributions. Lunar Planet. Sci. Conf. Proc. 3, 2919–2933.

Gregg, T.K.P., Greeley, R., 1992. Formational constraints on venusian "canali". In: Lunar and Planetary Science Conference. In: Lunar and Planetary Science Conference, vol. 23, p. 449.

Haase, I., Wählisch, M., Gläser, P., Oberst, J., Robinson, M.S., 2019. Coordinates and maps of the apollo 17 landing site. Earth Space Sci. 6 (1), 59–95. http://dx.doi.org/10.1029/2018EA000408.

Hagen, T.H., 2015. Pit crater chain, pit chain. In: Hargitai, H., Kereszturi, A. (Eds.), Encyclopedia of Planetary Landforms. Springer New York, New York, NY, pp. 1575–1579. http://dx.doi.org/10.1007/978-1-4614-3134-3_261.

Haruyama, J., et al., 2009. Possible lunar lava tube skylight observed by SELENE cameras. Geophys. Res. Lett. 36 (21), L21206. http://dx.doi.org/10.1029/2009GL040635.

Haruyama, J., et al., 2010. New discoveries of lunar holes in mare tranquillitatis and mare ingenii. In: 41st Annual Lunar and Planetary Science Conference. In: Lunar and Planetary Science Conference, p. 1285.

He, K., Gkioxari, G., Dollár, P., Girshick, R., 2017. Mask R-CNN. http://dx.doi.org/10.48550/arXiv.1703.06870, ArXiv E-Prints.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, p. 1. http://dx.doi.org/10.1109/CVPR.2016.90.

Horstman, K.C., Melosh, H.J., 1989. Drainage pits in cohesionless materials: implications for the surface of Phobos. J. Geophys. Res. 94, 12433–12441. http://dx.doi.org/10.1029/JB094iB09p12433.

Horvath, T., Hayne, P.O., Paige, D.A., 2022. Thermal and illumination environments of lunar pits and caves: Models and observations from the diviner lunar radiometer experiment. Geophys. Res. Lett. 49 (14), e99710. http://dx.doi.org/10.1029/2022GL099710.

Huang, G., Liu, Z., van der Maaten, L., b Weinberger, K.Q., 2016. Densely connected convolutional networks. http://dx.doi.org/10.48550/arXiv.1608.06993, ArXiv E-Prints.

Hurwitz, D.M., Head, J.W., Hiesinger, H., 2013. Lunar sinuous rilles: Distribution, characteristics, and implications for their origin. Planet. Space Sci. 79, 1–38. http://dx.doi.org/10.1016/j.pss.2012.10.019.

Kaku, T., et al., 2017. Detection of intact lava tubes at marius hills on the moon by SELENE (kaguya) lunar radar sounder. Geophys. Res. Lett. 44 (20), 10,155–10,161. http://dx.doi.org/10.1002/2017GL074998.

Kempe, S., 2012. Volcanic rock caves. In: White, W.B., Culver, D.C. (Eds.), Encyclopedia of Caves, second ed. Academic Press, Amsterdam, pp. 865–873. http://dx.doi.org/10.1016/B978-0-12-383832-2.00125-0.

Kim, J.-R., Lin, S.-Y., Oh, J.-W., 2022. The survey of lava tube distribution in Jeju Island by multi-source data fusion. Remote. Sens. 14 (3), 443. http://dx.doi.org/10.3390/rs14030443.

Klem, S.M., Henriksen, M.R., Stopar, J., Boyd, A., Robinson, M.S., LROC Science Team, 2014. Controlled LROC narrow angle camera high resolution mosaics. In: 45th Annual Lunar and Planetary Science Conference. In: Lunar and Planetary Science Conference, p. 2885.

Laura, J., et al., 2023. Integrated software for imagers and spectrometers. http://dx.doi.org/10.5281/zenodo.7644616, URL: https://isis.astrogeology.usgs.gov/.

Le Corre, D., 2025a. Entrances to sub-surface areas (ESSA). GitHub Repository. URL: https://github.com/dlecorre387/Entrances-to-Sub-Surface-Areas.

Le Corre, D., 2025b. Pit topography from shadows (PITS). GitHub Repository. URL: https://github.com/dlecorre387/Pit-Topography-from-Shadows.

Le Corre, D., 2025c. Planetary image processing bash scripts. GitHub Repository. URL: https://github.com/dlecorre387/Planetary-Image-Processing.

Le Corre, D., Mary, D., Mason, N., Bernard-Salas, J., Cox, N., 2023. Automatically calculating the apparent depths of pits using the pit topography from shadows (PITS) tool. RAS Tech. Instrum. 2 (1), 492–509. http://dx.doi.org/10.1093/rasti/rzad037.

Le Corre, D., Mason, N., Bernard-Salas, J., Mary, D., Cox, N., 2025. ESSA PyTorch model checkpoint and shapefiles containing pit and skylight detections on the moon. http://dx.doi.org/10.5281/zenodo.15438463.

Léveillé, R.J., Datta, S., 2010. Lava tubes and basaltic caves as astrobiological targets on earth and Mars: A review. Planet. Space Sci. 58 (4), 592–598. http://dx.doi.org/10.1016/j.pss.2009.06.004.

Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., Belongie, S., 2016. Feature pyramid networks for object detection. http://dx.doi.org/10.48550/arXiv.1612.03144, ArXiv E-Prints.

Loshchilov, I., Hutter, F., 2017. Decoupled weight decay regularization. http://dx.doi.org/10.48550/arXiv.1711.05101, ArXiv E-Prints.

Malin, M.C., et al., 2007. Context camera investigation on board the Mars reconnaissance orbiter. J. Geophys. Res. (Planets) 112 (E5), E05S04. http://dx.doi.org/10.1029/2006JE002808.

Martin, R.P., Benaroya, H., 2023. Pressurized lunar lava tubes for habitation. Acta Astronaut. 204, 157–174. http://dx.doi.org/10.1016/j.actaastro.2022.12.013.

McEwen, A., 2007. Mars reconnaissance orbiter high resolution imaging science experiment, reduced data record, MRO-M-HIRISE-3-RDR-V1.1. http://dx.doi.org/10.17189/1520303, NASA Planetary Data System. URL: https://hirise-pds.lpl.arizona.edu/PDS/RDR/.

McEwen, A.S., et al., 2007. Mars reconnaissance orbiter's high resolution imaging science experiment (HiRISE). J. Geophys. Res. (Planets) 112 (E5), E05S02. http://dx.doi.org/10.1029/2005JE002605.

McEwen, A.S., et al., 2024. The high-resolution imaging science experiment (HiRISE) in the MRO extended science phases (2009–2023). Icarus 419, 115795. http://dx.doi.org/10.1016/j.icarus.2023.115795.

Nesnas, I.A.D., et al., 2023. Moon diver: Exploring a pit's exposed strata to understand lunar volcanism. Acta Astronaut. 211, 163–176. http://dx.doi.org/10.1016/j.actaastro.2023.05.042.

Nodjoumi, G., Pozzobon, R., Sauro, F., Rossi, A.P., 2023. DeepLandforms: A deep learning computer vision toolset applied to a prime use case for mapping planetary skylights. Earth Space Sci. 10 (1), e2022EA002278. http://dx.doi.org/10.1029/2022EA002278.

Okubo, C.H., Martel, S.J., 1998. Pit crater formation on Kilauea volcano, Hawaii. J. Volcanol. Geotherm. Res. 86 (1), 1–18. http://dx.doi.org/10.1016/S0377-0273(98)00070-5.

Pizer, S.M., et al., 1987. Adaptive histogram equalization and its variations. Comput. Vis. Graph. Image Process. 39 (3), 355–368. http://dx.doi.org/10.1016/S0734-189X(87)80186-X.

Prieur, N.C., et al., 2023. Automatic characterization of boulders on planetary surfaces from high-resolution satellite images. J. Geophys. Res. (Planets) 128 (11), e2023JE008013. http://dx.doi.org/10.1029/2023JE008013.

QGIS Development Team, 2025. QGIS geographic information system. QGIS Association. URL: https://www.qgis.org.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster R-CNN: Towards real-time object detection with region proposal networks. http://dx.doi.org/10.48550/arXiv.1506.01497, ArXiv E-Prints.

RGCPS, 2025. Lava tube database. Ronald Greeley Center for Planetary Studies (RGCPS), Arizona State University. URL: https://rgcps.asu.edu/ltdb/.

Robbins, S.J., Kirchoff, M.R., Hoover, R.H., 2023. Fully controlled 6 meters per pixel equatorial Mosaic of Mars from Mars reconnaissance orbiter context camera images, version 1. Earth Space Sci. 10 (3), e2022EA002443. http://dx.doi.org/10.1029/2022EA002443.

Robinson, M., 2009. Lunar reconnaissance orbiter narrow angle camera, experiment data record, LRO-L-LROC-2-EDR-v1.0. http://dx.doi.org/10.17189/1520250, NASA Planetary Data System (PDS). URL: https://pds.lroc.asu.edu/data/LRO-L-LROC-2-EDR-V1.0/.

Robinson, M.S., et al., 2010. Lunar reconnaissance orbiter camera (LROC) instrument overview. Space Sci. Rev. 150 (1–4), 81–124. http://dx.doi.org/10.1007/s11214-010-9634-2.

Rubanenko, L., Perez-Lopez, S., Schull, J., Lapotre, M.G.A., 2021. Automatic detection and segmentation of Barchan Dunes on Mars and earth using a convolutional neural network. IEEE J. Sel. Top. Appl. Earth Obs. Remote. Sens. 14, 9364–9371. http://dx.doi.org/10.1109/JSTARS.2021.3109900.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.-C., 2018. MobileNetV2: Inverted residuals and linear bottlenecks. http://dx.doi.org/10.48550/arXiv.1801.04381, ArXiv E-Prints.

Sauro, F., Pozzobon, R., Massironi, M., De Berardinis, P., Santagata, T., De Waele, J., 2020. Lava tubes on Earth, Moon and Mars: A review on their size and morphology revealed by comparative planetology. Earth Sci. Rev. 209, 103288. http://dx.doi.org/10.1016/j.earscirev.2020.103288.

Sauro, F., et al., 2019. Volcanic caves of lanzarote: A natural laboratory for understanding volcano-speleogenetic processes and planetary caves. In: Mateo, E., Martínez-Frías, J., Vegas, J. (Eds.), Lanzarote and Chinijo Islands Geopark: From Earth To Space. Springer International Publishing, Cham, pp. 125–142. http://dx.doi.org/10.1007/978-3-030-13130-2_9.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. http://dx.doi.org/10.48550/arXiv.1409.1556, ArXiv E-Prints.

Spencer, J.R., Fanale, F.P., 1990. New models for the origin of Valles Marineris closed depressions. J. Geophys. Res. 95, 14301–14313. http://dx.doi.org/10.1029/JB095iB09p14301.

Speyerer, E.J., Robinson, M.S., Denevi, B.W., LROC Science Team, 2011. Lunar reconnaissance orbiter camera global morphological map of the moon. In: 42nd Annual Lunar and Planetary Science Conference. In: Lunar and Planetary Science Conference, p. 2387.

Theinat, A.K., Modiriasari, A., Bobet, A., Melosh, H.J., Dyke, S.J., Ramirez, J., Maghareh, A., Gomez, D., 2020. Lunar lava tubes: Morphology to structural stability. Icarus 338, 113442. http://dx.doi.org/10.1016/j.icarus.2019.113442.

van der Bogert, C.H., Ashley, J.W., 2015. Skylight. In: Hargitai, H., Kereszturi, A. (Eds.), Encyclopedia of Planetary Landforms. Springer New York, New York, NY, pp. 1964–1969. http://dx.doi.org/10.1007/978-1-4614-3134-3_342.

Wagner, R.V., Robinson, M.S., 2014. Distribution, formation mechanisms, and significance of lunar pits. Icarus 237, 52–60. http://dx.doi.org/10.1016/j.icarus.2014.04.002.

Wagner, R.V., Robinson, M.S., 2021. Occurrence and origin of lunar pits: Observations from a new catalog. In: 52nd Lunar and Planetary Science Conference. In: Lunar and Planetary Science Conference, p. 2530, URL: https://www.lroc.asu.edu/atlases/pits.

Wagner, R.V., Robinson, M.S., 2022. Lunar pit morphology: Implications for exploration. J. Geophys. Res. (Planets) 127 (8), e07328. http://dx.doi.org/10.1029/2022JE007328.

Wagner, R.V., et al., 2024. Where is that crater? Best practices for obtaining accurate coordinates from LROC NAC data. Planet. Sci. J. 5 (7), 157. http://dx.doi.org/10.3847/PSJ/ad54c6.

Watson, T.H., Baldini, J.U.L., 2024. Martian cave detection via machine learning coupled with visible light imagery. Icarus 411, 115952. http://dx.doi.org/10.1016/j.icarus.2024.115952.

Williams, K.E., McKay, C.P., Toon, O.B., Head, J.W., 2010. Do ice caves exist on Mars? Icarus 209 (2), 358–368. http://dx.doi.org/10.1016/j.icarus.2010.03.039.

Wyrick, D., Ferrill, D.A., Morris, A.P., Colton, S.L., Sims, D.W., 2004. Distribution, morphology, and origins of Martian pit crater chains. J. Geophys. Res. (Planets) 109 (E6), E06005. http://dx.doi.org/10.1029/2004JE002240.

Zhang, S., et al., 2024. Detecting lunar linear structures based on multimodal semantic segmentation: The case of Sinuous Rilles. Remote. Sens. 16 (9), 1602. http://dx.doi.org/10.3390/rs16091602.

Zhou, M., et al., 2024. High-resolution morphology of lunar lava tube pits using photogrammetric modeling of multiple stereo images. Earth Space Sci. 11 (11), 2024EA003532. http://dx.doi.org/10.1029/2024EA003532.