Australian Space Data Analysis Facility

A Machine Counts Craters on the Moon: Progress Towards a Complete Global Impact Crater Dataset Project Article

24th February 2022

Project collaborators: CSIRO, Pawsey Supercomputing Centre, The Australian Space Data Analysis Facility, Space Science and Technology Centre.

In this joint project with CSIRO, Pawsey Supercomputing Centre, and Space Science and Technology Centre, this project aimed to unravel the geological history of the Moon through the detection and analysis of billions of craters constellating its surface using a machine learning algorithm running on Pawsey supercomputers. Specifically, this project aims to pre-process high-resolution imagery dataset, optimise the using of the Crater Detection Algorithm co-developed by research engineers at CSIRO and scientists at Curtin, and post-process the data.

The team consists of Kosta Servis (Pawsey/ASDAF), Dr. Anthony Lagain (Curtin University), Professor Gretchen Benedix (Curtin University) and Mr. John Fairweather (PhD Candidate, Curtin University). Kosta, who sits on both the Pawsey Supercomputing and the ASDAF team, has been invaluable in gathering a high volume of data, pre-treating them and applying the Crater Detection Algorithm which he adapted for this project's purpose, and finally, post-treating the detections to create the largest lunar crater database ever built, containing craters that are ~100 m in diameter and containing ~500M of entities.

The Crater Detection Algorithm was originally developed to work on Martian images, however, it has been adapted to work on Moon images in January of 2020. With the algorithm still running in Pawsey, it will analyse the images and compile the data which will be evaluated against the ground truth dataset (craters that were manually mapped out by the team to evaluate the performance of the algorithm) – if satisfactory, the project is expected to be complete in March 2022.

Australian Space Data Analysis Facility

Challenges in the project

There were a few challenges that the team encountered at different stages of the project from the production of a georeferenced dataset to storing the files the research produced:

- Issues with the starting dataset; the team deemed the data set from NASA LRO NAC (the Lunar Reconnaissance Orbiter – Narrow Angle Camera) unsuitable to produce a georeferenced data set. There was potentially a more practical alternative with the slightly reduced resolution by JAXA Kaguya – Selene mission Thematic Camera (TC). The dataset had a far superior quality in terms of georeferencing and there is an available mosaic of those images on PDS (Planetary Data System), however the data was slightly reduced in resolution (7.4m/px)
- As the data set was divided in lunar mornings and evenings, it must be processed in pairs and the use of NMS (Non-Max Suppression) implementation was needed at the final stage to remove duplicates
- When down sampling (to reduce the images' resolution to detect large craters), it did not produce am expected size frequency distribution. The team had to finetune the downscaling process and arrived at an optimal configuration
- The default projection distorts angles and areas especially at higher latitudes, such that an additional step of geographic projection was necessary. All images were re-projected to a local stereographic projection which does preserve the geometric characteristics in the local area
- As the team's algorithm depends on the illumination to detect circular features, there was the issue of varying illumination and contrast features. This was solved by performing a histogram equalisation on the images for practicality.
- A few technical issues with running algorithm on the Pawsey Cray Cluster that relate to Nextflow (the workflow framework). The first issue was that Nextflow does not support multicluster jobs which is required by the workflow as there are different steps that require different resources from different clusters. To overcome this the Nextflow framework was modified to support multicluster Slurm job (the name of the process scheduler used at Pawsey)

- Lastly, the number of intermediate files and links produced by the Nextflow framework were more numerous than the default allowed by the Pawsey work storage system – this was addressed by Pawsey allowing more files temporarily and tweaking a number of parameters

Outcome and Next Steps

The project will be the subject of a few publications in rank A journals, presenting the method, the output (the crater database), and the scientific applications. Besides the possible applications of this project, the crater database will allow future researchers to investigate: (1) the spatial and temporal evolution of the cratering rate on the Moon, (2) date lunar surfaces in a semi-automatic way, (3) investigate the material transport from impact cratering to the Chang'e, Apollo, and Luna missions landing sites, and thus understand some contaminations in the samples that have been collected on the Moon, and (4) the spatial variation of the erosion rate on the Moon.

Through this project, it allowed the team to build the largest crater database compiling the size and location of ~500M impact craters on the Moon. The team is now investigating the possibility to adapt the algorithm to classify the morphology (shape and degradation state) of impact craters. This information will be particularly useful in determining the approach conditions of asteroids (angle, velocity, density) that have formed those craters, as well as the age of the craters.

Impact of Project

The Crater Detection Algorithm applied on the largest planetary image dataset, demonstrated its usefulness in planetary science through a few publications in a top-ranked journal based on its application on Mars (with the same to be expected with its application on the Moon). Through the retraining of this algorithm to work on terrestrial dataset can help in nature preservation and agriculture such as the automatic detection of bushfires and classifying land use.