Earth and Space Science

RESEARCH ARTICLE
10.1029/2019EA001005

Deriving Surface Ages on Mars Using Automated Crater Counting

G. K. Benedix1, A. Lagain1, K. Chai4, S. Meka4, S. Anderson1, C. Norman5, P. A. Bland1,2, J. Paxman5, M. C. Towner1, and T. Tan5

1Space Science and Technology Centre, School of Earth and Planetary Sciences, Curtin University, Perth, Western Australia, Australia, 2Department of Earth and Planetary Sciences, Western Australian Museum, Western Australia, Australia, 3Also at Planetary Sciences Institute, Tucson, AZ, USA, 4Curtin Institute for Computation, Curtin University, Perth, Western Australia, Australia, 5School of Civil and Mechanical Engineering, Curtin University, Perth, Western Australia, Australia

Abstract
Impact craters on solar system bodies are used to determine the relative ages of surfaces. The smaller the limiting primary crater size, the higher the spatial resolution in surface/resurfacing age dating. A manually counted database (Robbins & Hynek, 2012, https://doi.org/10.1029/2011JE003966) of >384,000 craters on Mars >1 km in diameter exists. But because crater size scales as a power law, the number of impact craters in the size range 10 m to 1 km is in the tens of millions, a number making precise analysis of local variations of age, over an entire surface, impossible to perform by manual counting. To decode this crater size population at a planetary scale, we developed an automated Crater Detection Algorithm based on the You Only Look Once v3 object detection system. The algorithm was trained by annotating images of the controlled Thermal Emission Imaging System daytime infrared data set. This training data set contains 7,048 craters that the algorithm used as a learning benchmark. The results were validated against the manually counted database as the ground truth data set. We applied our algorithm to the Thermal Emission Imaging System global mosaic between ±65° of latitude, returning a true positive detection rate of 91% and a diameter estimation error (~15%) consistent with typical manual count variation. Importantly, although a number of automated crater counting algorithms have been published, for the first time we demonstrate that automatic counting can be routinely used to derive robust surface ages.

Plain Language Summary
Crater counting is the traditional method of determining the surface ages of planets throughout the solar system. This method, up to now, has used data that have been painstakingly counted by hand. The current published database for Mars contains hundreds of thousands of craters for diameters larger than 1 km. If we can count craters smaller than this, we will be able to target specific areas of interest to date. But the rate of impacts on planetary surfaces follows a power law such that the number of small (less than 1 km) craters is exponentially higher than the number of large craters. To count these requires an automated tool. Here we show that we have developed such a tool. We have validated the results against current manual databases. Importantly and for the first time, we demonstrate that an automated crater counting tool can deliver geologically meaningful ages.

1. Introduction and Background
Planetary science encompasses the observations of the planets and moons in our solar system as well as the study of the processes by which they formed and evolved. Our place in space illustrates that planetary bodies evolve along many different geologic pathways over time. The study of the planets and other bodies in our solar system, therefore, provides a template for untangling planet formation and evolution not just here but throughout the universe; abodes for life in this neighborhood of space might well be abodes for life elsewhere in the galaxy or universe.

Astrogeologists study the planets, moons, asteroids, and comets in our solar system, using baseline data sets composed of imagery of the surface and a geologic map that includes age information (Prockter et al., 2016; Tanaka et al., 2014; Wilhelms & McCauley, 1971). We determine ages for worlds that we have never visited by counting the number of craters of a given size that we see on their surfaces. The more craters, the older the surface. The technique was invented in the early 1960s (Baldwin, 1965; Shoemaker et al., 1963; Shoemaker &
Hackman, 1962). Crater counting provides a relative age, but if we can obtain a radioisotope age from a sample of a cratered surface, that relative age can be calibrated. It becomes an absolute chronometer. This was achieved with the return of the Apollo samples. Since then, crater counting has been at the heart of virtually every major advance in planetary science, from deciphering the geological history of the Moon (Jolliff et al., 2006) to the concept of the late heavy bombardment (e.g., Strom, 1979) as well as the realization that the icy satellites of the outer gas giants might have ocean mantles (Morrison & Cruikshank, 1974). Advances in technology mean that image resolution has increased dramatically (from km/pixel to cm/pixel), but there has been essentially no change in how that data are processed to derive ages for surfaces. The default method is counting craters by hand. This effectively limits us to craters >1 km. There are ~385,000 of these on Mars. They were counted and measured by hand (Robbins & Hynek, 2012). Currently, available image resolutions can distinguish craters down to 10 m in size on several planetary bodies. Because the crater production function follows a series of power laws (Hartmann, 2005; Hartmann & Daubar, 2016), this means that there are tens of millions of these primary craters—far too many to acquire a reliable data set by manual counting. But if we could measure all of them, we could have age maps at ultimate resolution (after accounting for secondary craters), deriving ages and resurfacing ages for much smaller and younger surface features. We have created an extensive training data set and developed an advanced machine learning algorithm to do this. We validated the algorithm against the manually counted data set (Robbins & Hynek, 2012). Our algorithm is shown to successfully count craters down to 10 pixels diameter on any number of high-resolution images.

1.1. Crater Counting

Crater counting provides relative timing of the geological history of a planetary surface (Hartmann, 1964; Hartmann & Neukum, 2001). Impact craters form when a meteoroid strikes the surface of a planetary body. The age of a surface can be estimated through analysis of crater frequencies, a so-called crater size-frequency distribution (CSFD), assuming random impacts of meteoroids with known long-term average rates (Hartmann & Neukum, 2001). If an absolute external age can be accurately attributed to a surface (as happened with the Moon when Apollo and Luna samples were returned and radioisotopically dated (Stöffler & Ryder, 2001)), then the crater counting time frame can be calibrated, providing specific age information for a variety of features. The current age calibration for Mars is based on the lunar calibration modified by dynamical arguments (Hartmann, 2005; Hartmann & Neukum, 2001; Neukum et al., 2001).

Over the last half century, and particularly in the last 25 years, the spatial resolution of planetary image data sets acquired by orbiting spacecraft has improved (Mariner 4 had an average resolution of ~3 km/pixel) and we can now see the surfaces of other planets at sub–m scales (HiRISE relays images of Mars at ~30 cm per pixel). The ability to make full use of these publicly available high-resolution image data sets, and count crater sizes to tens of meters in diameter, would allow unprecedented resolution, at fine-grained spatial scales, in the determination of timing between young and/or small geological terrains as well as the dating of geologic processes such as volcanism or erosion. To access them, we need to automate the process.

1.2. Manual Crater Counts

For Mars and the Moon, there are complete databases of manually counted craters (Robbins & Hynek, 2012; Robbins, 2019) down to a minimum size of 1 km. While these provide an invaluable resource for the community in deciphering the geological histories of these worlds, they are not definitive, in that no crater counting method—manual or automated—delivers a perfectly accurate representation of the crater population on a planetary surface. Automated crater detection approaches are tested against manual data sets, but it is notable that there is significant variability among manual data sets (Robbins et al., 2014) arising from the subjective quality of how any given human interacts with an image. Moreover, studies have shown that human attention declines rapidly after 30 min (See et al., 1995). This may explain why comparisons between manually counted data sets find an upper correspondence limit of around 80%; that is, in determining the error budget for any given isochron, we should appreciate that there will always be a 20% error arising from the lack of complete reproducibility between human operators and the subjective quality of the experience.
1.3. The Current Status of Automated Crater Detection Algorithms

A number of studies have addressed automated crater detection. DeLatte, Crites, Guttenberg, Tasker, and Yairi (2019) and DeLatte, Crites, Guttenberg, and Yairi (2019) present a very comprehensive overview of the many different studies and range of techniques applied to the problem over the last 20 years. We do not repeat this review here, as it would be redundant, but point the reader to DeLatte, Crites, Guttenberg, and Yairi (2019) for the full review. The benefit of automated approaches is that they address a number of human limitations. A computer does not get tired; there is no attention issue (See et al., 1995). Such an analysis is objective and reproducible. They can count craters to the limit of image resolution, offering the potential of reliable crater data sets into the tens of millions. But there are an equal number of discrete issues that are unique to this approach. To date, no CDAs have achieved the ultimate goal of counting and measuring craters over large areas in a reliable and timely fashion, although small area work has shown some recent progress (DeLatte, Crites, Guttenberg, Tasker, & Yairi, 2019; DeLatte, Crites, Guttenberg, & Yairi, 2019; Lee, 2019). Approaches have included edge detection, Hough transforms, and more recently machine learning and deep learning. Some of these previous efforts used elevation models in order to minimize the effects of Sun angle on the optical images (e.g., Lee, 2019). However, despite good detection rates, elevation models currently available for celestial bodies do not benefit from the same coverage and resolution that imagery can offer. Although advances have been made, it has proven to be surprisingly difficult to teach a computer to recognize the subtle variation in crater morphology and sizes, in order to count and characterize them. This is not surprising given that humans also struggle to consistently decide what is or is not a crater (Robbins & Hynek, 2014). This is especially true for Martian craters because of the diversity of circular features at the surface of the planet that are not attributable to formation by impact, such as calderas, parasitic cones, mounds, and giant polygons. For these and other reasons, no CDA has been developed to the point where it can be an easy tool to use in understanding the geology of a planetary surface, generating isochrons and geologically meaningful ages that are consistent with literature values derived from manual counts. However, machine learning has advanced significantly in recent years. Live facial recognition is now a reality in many cities. Since this is a significantly more complex computational task than identification and characterization of circular features in planetary data sets, we felt that a new approach to developing a CDA was timely.

In previous work, we described our technique of using supervised machine learning, (Benedix et al., 2018; Benedix et al., 2019; Galloway et al., 2014; Galloway et al., 2015; Norman et al., 2018) in some detail. Here we discuss the evolution of the technique. We also show that results are statistically comparable to manual count data sets for craters greater than 1 km in diameter and that the algorithm is able to recognize craters down to hundreds of meters across on Mars, allowing us to generate isochrons for surfaces on Mars (or any other cratered planetary surface using appropriate training sets), at the highest resolution, routinely.

2. Methods

2.1. Algorithm—Convolutional Neural Network

Crater counting is, at its heart, an object detection task. Initial work for the development of the CDA (Benedix et al., 2018; Norman et al., 2018) began with training a Convolutional Neural Network (CNN) using the OverFeat architecture (Sermanet et al., 2014) specifically designed for visual object detection. However, rapid advances in the last 2 years have produced newer, faster, and more accurate object detection algorithms. Since then, we have improved upon our initial CDA model by training a CNN using the You Only Look Once (YOLOv3) architecture (Redmon et al., 2016; Redmon & Farhadi, 2018).

CNNs have proven to perform well at complex classification tasks (Krizhevsky et al., 2012). “Convolution” in the context of machine learning consists of applying filters (matrices of numbers) across an image to produce intermediate outputs known as feature maps. The goal of applying machine learning is to optimize and/or learn to identify “good” features (matrices of numbers) to perform a certain task such as detecting craters in satellite imagery through a training process. The combination of computational operations such as convolutions is referred to as layers of a neural network, and the way these layers are connected to other layers is referred to as a “network architecture.”

The YOLO architecture is based on the Darknet-53 network for feature extraction (Redmon et al., 2016). Darknet-53 uses 53 convolutional layers and performs on par with other state-of-the-art classifiers on
standardized tests and with greater speed (Redmon & Farhadi, 2018). YOLOv3 adds three separate output
structures to enable predictions of objects at three different scales. Consequently, the algorithm performs
well at detecting small, medium, and large objects simultaneously within the same image. This is highly
suitable for the crater detection problem, where craters from tens of meters to tens of kilometers in size
are of interest.

The object detection methods used in our initial CDA model relied upon applying a CNN classifier on each
satellite image tile at multiple locations and scales. The YOLO architecture differs from these prior
approaches in that a single neural network is applied to the image tile once, which directly predicts bound-
ing boxes for objects at different scales. This means that YOLO predictions are informed by the global context
of the image, which can result in better model accuracy, and that the predictions are much faster than
prior methods.

2.2. Training

A supervised machine learning algorithm requires a ground truth database of examples labeled by experts.
We used the most complete catalog database of craters with diameters greater than 1 km across the entire
Martian surface, currently numbering some 384,343 entries (Robbins & Hynek, 2012), as the ground truth
data set. We assumed this Mars Crater Database (MCD) to be consistent enough to create a substantial
library for the training of the algorithm with the objective of generalizing to craters smaller than 1 km
in diameter.

2.2.1. Input Imagery

The training data set was established using Thermal Emission Imaging System (THEMIS) Day IR images
(Christensen et al., 2004) in the latitudinal band between 35°N and 35°S of the equator. This band was
chosen because no image projection is required. This band represents ~60% of the Martian surface. For this
purpose, the latest available version of the global mosaic was used (Edwards et al., 2011), with a typical
spatial resolution of 100 m per pixel. We used the traditional “Mars Chart” regions: 30, approximately equal
area, quadrangles commonly used by the mapping community (Batson, 1973; Gaither et al., 2015; Skinner
et al., 2006). Data were selected from the 16 quadrangles in the equatorial band (±30° latitude) for training
and testing of the algorithm.

2.2.2. Ground Truth Data Set

An initial training data set was established by partitioning THEMIS mosaics into 6,380 tiles, each 1,280 × 960
pixels, and overlaying the MCD to establish ground truth. It was clear, after a cursory visual check, that
intermittent projection errors and missing annotations meant that this initial data set produced some
inconsistent results. A decision was made to reduce the size of the training set by reducing the tile size
and selecting tiles with the most complete and correct annotations. This involved meticulous checking of
each tile to ensure that there was no ambiguity. This process took the most time in the development of
the algorithm, amounting to around 1 month of person-time. The reduced data set comprised 889 tiles
(1,762 craters) of 416 × 416 pixels, which were augmented (rotated and reflected, vertically and horizontally)
to produce a final data set of 3,556 tiles (7,048 craters). Figure S1 in the supporting information shows
examples of annotated tiles that were used to build the training and test data sets.

This data set was then divided into a training set of 2,490 tiles (70%), from which the algorithm learned, and a
validation set of 1,066 tiles (30%), allowing us to assess performance. A CNN model based on the YOLOv3
neural network architecture was trained with a learning rate of 0.001 for 40,000 iterations and achieved a
YOLO loss (error) of 0.196. The YOLO loss function combines the mean square errors in classification,
localization, and confidence (Redmon et al., 2016).

2.3. Validation

2.3.1. Error Analysis and Evaluation

The performance of the CDA on the validation set was evaluated through the measurement of intersection
over union (IoU), precision, and recall. IoU (Figure S2a) is defined to be the ratio between the intersection
and the union of the areas of a detected crater and a corresponding ground truth crater (or more precisely,
the bounding box around those craters). IoU can be used both to define (through a choice of threshold) when
a crater is determined to be correctly detected or not and to measure the accuracy of such detections.
Precision is the ratio of true positive detections (determined by an IoU > 0.5) to total detections. Recall is
the ratio of true positive detections to total ground truth craters. Precision measures the proportion of
detected crater which are true craters (according to the chosen IoU threshold), while recall measures the proportion of true craters that are detected. In general, a good result is when both precision and recall are simultaneously high.

When applying the CDA to the validation set, each identified object (crater) was assigned a confidence score from 0% to 100%, equating approximately to the probability of a correct detection multiplied by the expected IoU. A confidence threshold is required in order to determine which detected craters are counted, with 50% being a fairly typical choice. An initial confidence threshold of 50% was chosen; however, the performance on the validation set was quite robust against different choices of confidence threshold (see Figure S2b for a comparison between three choices of threshold). A confidence threshold of 20% was eventually chosen, due to a similar precision and slightly higher recall on the validation set.

The CDA applied to the validation set of 1,066 tiles resulted in 90% precision, 94% recall, and an average IoU of 70%. This is a very strong result, suggesting that the CDA can be applied broadly to the THEMIS Daytime IR data set with great confidence (note that the CDA was not trained on larger (complex) craters (there are 24,539 craters >10 km in diameter in the MCD), although it does succeed in detecting a significant number of larger craters). Validation is described in terms of true and false positives and negatives (as reported for other crater detection algorithms); this is illustrated visually in Figure S3.

A baseline validation of the accuracy of the CDA is shown in Figure S4, where the diameter is compared for a number of craters (Table S1 contains details of the locations of each crater). The comparison is excellent down to very small sizes of crater, where the limitation is due to the resolution of the THEMIS images, rather than an error of the CDA.

2.3.2. Data Processing Pipeline
To handle the large volumes of data for Mars (and other planetary bodies), we have incorporated the CDA into a data pipeline, as described here and shown in Figure S5. First, the image is converted to 8-bit JPG and divided into tiles (960 × 960 pixels) using the open-source software package, ImageMagick. Tiles are generated with overlapping patterns, to prevent issues of craters being ignored due to being split between tiles. Processed tiles are sent to the CDA tool, and crater coordinate results are collated from all tiles. Data that are extracted include crater diameters, total area searched, and metrics such as confidence ratings for each candidate crater.

The current iteration of the CDA tool was trained to recognize craters in a particular size range in terms of the number of pixels (circular features that are roughly 10 to 100 pixels across) so is optimally suited to these pixel dimensions. The size range in absolute diameter is then dependent on image resolution. To make the best use of the algorithm for a larger range of crater sizes, we processed each source image both in its native resolution of 100 m/pixel and also at artificially reduced resolutions of 500 and 1,500 m/pixel (before tile division; Figure S5). This ensures that the larger craters (up to 10 km) and craters that might be bigger than a single tile at raw resolutions are detected. Craters that are counted more than once at different resolutions are merged together by comparing the overlapping area of detections against each other.

Given that the CDA was not trained on large craters—the Robbins database is complete for these craters, so automated detection was not considered a priority—we did not expect that the algorithm would perform optimally for craters larger than 10 km on Mars, that is, beyond the simple-to-complex crater transition. The principal focus of our research is developing a method to automate counting of small, mostly simple craters. The algorithm was not trained on complex craters, but we include that data here for completeness.

The results of the CDA are included as a .csv table (Data Set S1) in the supporting information.

2.3.3. Application to Test Data Set—Global THEMIS Mosaic
After training and validation, we applied the CDA to all THEMIS DayIR mosaic quadrangles, aside from the polar regions (MC-01 and MC-30). The two polar quadrangles were excluded because polar processes result in modifications of the original morphology of the craters in these regions, and therefore, reliable crater identification is difficult.

Each quadrangle between ±30° latitude is a 26,674 × 17,783 pixel image (at 100 m per pixel; this is equivalent to 4.7 ×10⁶ km²). With minimal preprocessing of the publicly available images, the image was evaluated using the pipeline described above. The tiling sequence initially took 3 min. Each tile was run through the detection algorithm with all tiles being fully analyzed in 2 min using a desktop GPU (Nvidia GTX 1060).
The most time-intensive part of the pipeline, to begin with, was the visualization of the detections, which involved annotating each tile and then merging them back together into a single image. This process took 10 min. The total time to analyze a single THEMIS image was 15 min, which takes significantly less time than it would take compared to the manual approach. However, these timings were vastly improved when the input image format was compressed from PNG to JPG formats. With this change, each of the above steps took 15, 50, and 50 s, respectively. This effectively means that the algorithm can interrogate an area of \(4.7 \times 10^6\) km\(^2\) in 2 min without compromising accuracy.

### 3. Results and Discussion

In this section, we report the results of the CDA as applied to the Global THEMIS mosaic and then compare to the MCD of Robbins and Hynek (2012) which was built using the same data set. We discuss the statistical significance of the differences between the two approaches and show that over the diameter range that is optimally counted by the algorithm, the CDA is quite robust. Table S2, in the supporting information, contains the diameter, latitude, and longitude values for all automatically counted craters.

Table 1, below, has the specific comparison metrics of the two databases. The results are evaluated over two latitude ranges. The area between \(\pm 65^\circ\) represents a significant portion of the Martian surface, importantly where crater morphology is generally less circular. We also excerpt the data between \(\pm 45^\circ\) to show the areas where the crater morphology is more similar to the labeled training data set. The results are further broken down to illustrate the capability of the CDA over the entire range of diameters (i.e., all craters >1 km) and the subset of that data covering the span of diameters between 1.5 and 10 km.

#### 3.1. Comparison to Manual Counts: Global Analysis

In terms of absolute numbers, the algorithm results are within 10% of the manual counting across all latitudes. When only detections between \(\pm 45^\circ\) are taken into account for all diameter values, the return rate is 97% (Table 1). When comparing the diameter subset of 1.5 to 10 km (for which the algorithm is optimized), it is interesting to note that the algorithm tends to count more than the manual counting. We discuss this below.

This initial comparison is encouraging, but a more complete evaluation of the algorithm is a comparison of the true positive population of craters. For this, we use the MCD as the “truth” with the caveat that there are some limitations in the MCD—as with any manual count—as discussed above. We identify the craters that appear in both data sets. The ACD has 302,188 true positive craters located between \(\pm 65^\circ\) latitude. Comparatively, the MCD contains 364,247 craters for the same latitude range. This result means that the ACD is roughly 85% of the MCD count rate. Robbins et al. (2014) note that this is approximately the percentage of variation between different crater counting experts. The variation is largest for craters larger than 10 km, but, as discussed above, the CDA was not trained to identify complex craters that are larger.

---

**Table 1**

Comparison of Absolute Crater Counts for the Crater Detection Algorithm (ACD) and the Mars Crater Database (MCD)

|                  | 65°S < Latitude < 65°N | 45°S < Latitude < 45°N |
|------------------|------------------------|------------------------|
|                  | All diam | 1.5 to 10 km diam | All diam | 1.5 to 10 km diam |
| MCD              | 364,247 | 163,410 | 305,239 | 132,638       |
| ACD              | 328,833 | 168,360 | 297,923 | 144,521       |
| Overall detection comparison (ACD/MCD) | 0.90 | 1.03 | 0.98 | 1.09 |
| True positive    | 302,188 (83% of MCD) (92% of ACD) | 125,372 (77% of MCD) (75% of ACD) | 262,602 (86% of MCD) (88% of ACD) | 104,041 (79% of MCD) (72% of ACD) |
| False Positive   | 26,645 (8%) | 42,988 (25%) | 55,321 (12%) | 40,480 (28%) |
| Uncorrelated Craters | 62,509 (17%) | 38,038 (23%) | 42,637 (14%) | 28,597 (21%) |

*Note.* True positive and false positive detections are visually defined in Figure S3. The percentages in italics denote the comparison to each database. True positive detections are actual craters; false positive detections indicate where the CDA does not match the manual count; uncorrelated craters indicate where the algorithm found a crater that is not present in the MCD. The use of italics is to deemphasize relative to the absolute number but give some statistical relevance to it.
Figure 1 presents the test results as crater density using a base grid cell of 100 km$^2$ (Figures 1a and 1b), regardless of diameter. The percentage difference between these two is shown in Figure 1c. This image allows differences to be rapidly identified for further inspection. Areas where the CDA is within 10% of Robbins et al., 2012, results are shown in Figure 1c. Red areas correspond to areas where the CDA counts more, whereas blue areas indicate where the CDA counts fewer than the MCD. Black circles are the locations of large primary impact craters that have produced kilometer-sized secondary craters on Mars (Robbins & Hynek, 2014).
the MCD are white, areas where the CDA underestimates are blue, and where it overestimates are in red. The main differences between the two databases are primarily related to the underlying terrain as discussed below.

The CDA consistently detects fewer craters (Figure 1c) than the MCD in the vicinities of large primary impact craters (black circles on Figure 1c). These are thought to be the sources of secondary craters larger than 1 km in diameter (Robbins & Hynek, 2014). Impact craters recognized as secondaries in these areas generally exhibit irregular, noncircular morphology, forming some crater chains which the current CDA has not been trained to detect. The MCD includes and labels secondary craters. As it is necessary to remove secondary craters when calculating planetary surface ages using crater counting, this difference between the techniques does not necessarily impact the value of the CDA in determining surface ages.

The CDA finds fewer craters than the MCD at high latitudes (>45°; Figure 1c), which is likely due to the high degradation state of large craters in the Southern Hemisphere and to the large number of secondaries in the form of crater chains contained in the MCD, especially found around Lomonosov (64.9°N, 9.2°W) and Lyot (50.8°N, 330.7°W) craters (Robbins & Hynek, 2014).

The CDA tends to detect (and therefore count) more craters (Figure 1c), than the MCD, in younger terrains typified by volcanic and tectonic regions such as Elysium Planitia, Hadriacus Mons (Northeast of Hellas basin), and the Tharsis bulge. This also occurs in the areas of the northern lowlands such as Chryse, Utopia, and Arcadia Planitia and more generally at latitudes below than 45°N and 45°S. These regions exhibit circular features that are not craters (i.e., parasitic cones, collapse pits of lava tubes, circular grabens, and calderas). Similar nonimpact formed circles found in the northern lowlands are mounds, giant polygons, and mud volcanoes. Discriminating these features from impact craters is the focus of ongoing work and can be accounted for with further training of the CDA.

To determine the effect of these differences on the calculation of any ages derived from the automatically counted population, we compared the global CSFD of each database. Considering the number and range of crater sizes here, we follow the procedure outlined by Robbins and Hynek (2012) where we display and analyze the results incrementally using 20 diameter bins per log decade to build the CSFD of each database. This is in contrast to binning diameter values in the usual root-2 bins (Crater Analysis Group, 1979). This alternate method highlights small variations between one diameter bin and another. Figure 2 shows the comparison of the absolute results and the percentage difference between the MCD and the ACD for each diameter bin. The two databases show good agreement, especially for diameters between 1.5 and 3 km, where the average difference is less than 10%. There is also good agreement in the range 7 to 10 km. For diameters between 3 and 6 km, the CDA has detected up to 25% more than the MCD, a difference that will be discussed below. As expected, given that the CDA was not trained on complex craters, it consistently underestimates the numbers of larger craters. Significantly, since the slope is the same for both data sets over this size range, and the number of craters is very similar within each diameter bin, we can expect that isochrons developed from these data sets should be very similar.

The major reason for the difference in total numbers is due to the training data set that the CDA learned from, which contained labeled, nondegraded, mostly circular objects. Given that our focus is on the small crater population, the CDA has performed extremely well. For the rest of the paper, we will focus on craters detected with diameters between 1.5 and 10 km, for which the CDA is optimally trained.

### 3.2. The True Positive Population

From the ACD we performed a spatial distribution analysis to isolate entries automatically detected by the CDA which also occur in the MCD, that is, the “true positive” (Figure S3) population. This analysis consists of tying each entry of the ACD to its closest match in the MCD using latitude, longitude, and diameter. From this spatial join, we then calculated the overestimation or underestimation of the detected crater. We used a threshold of 2% for the latitude and longitude variation between each linked entries. For diameter, we applied a threshold of 50% variation. All entries in the ACD that did not satisfy these three conditions were removed from the list of the true positive detections and are therefore considered false negatives.

Across the entire crater diameter population detected by the CDA (328,833), 302,188 entries were recognized as true positives, giving a true positive detection rate of 92%, well above the counting variation of manual inspections. Within the 1.5–10 km diameter range, the true positive detection rate reached 75%, that is,
Figure 2. Incremental representation of the crater size-frequency distribution of the manual (in red) and the automatic crater database (in blue) with 20 diameter bins per log decade. The bottom plot represents the percentage difference between the two CSFDs over the diameter range 1.5 to 10 km, the range the algorithm targets for this data set. Data on the x axis were plotted as the mean of each bin boundary.

Figure 3. Location and diameter variation of true positive detections of impact craters between 1.5 and 10 km (~125,372 impact craters). The majority of true positive detections have central latitude and longitude (left and middle panels) errors less than 1%, indicating that the CDA finding is the same craters as Robbins and Hynek (2012). Diameter (right panel) variations are slightly higher, with positive error indicating a larger diameter prediction than the associated crater in the MCD, but the majority of diameter estimations are within 20% of the manual count.
125,372 detection relative to 168,360. Locations and diameters of this population were then compared in detail to the MCD. Figure 3 shows the variation of the CDA detections in the narrow diameter range for the latitude, longitude, and size of each entry relative to the Robbins and Hynek (2012) database.

Figure 4. Comparison of ages derived from the results of the CDA and entries from the MCD. Three geological units from the geological map of Mars (Tanaka, 2014) were chosen: Sinai Planum, a late Hesperian unit located south of Valles Marineris; Isidis Planitia, another late Hesperian unit but located at the dichotomy of the Northern and Southern Hemispheres; and Arabia Terra, a heavily cratered, late Noachian unit. Crater size-frequency distributions were derived from the manual count (Robbins & Hynek, 2012), compared with the results of our algorithm. The isochrons have been fitted over the same diameter range (shown as a gray bar on the x axis for each panel) for both databases to ensure that the ages are comparable and defined as the typical range of crater size from which an age can be derived for these geological units (Tanaka et al., 2014). The results shown in each panel (green for MCD and red for CDA) are identical to one another, within error, for all three geologic units. The figure also contains the number of craters for each identical area for comparison.
Negative values in latitude indicate a shift to the south from the corresponding entry in the MCD, while positive values correspond to a shift to the north (Figure 3a). Positive values of longitude variation correspond to an easterly shift relative to reported MCD values and vice versa for negative longitude variation (Figure 3b). Finally, diameter variations indicate larger (positive) or smaller (negative) (Figure 3c) sizes compared to matched MCD data. The analysis of the CDA diameter distribution shows that 15% of impact craters in the automatic database between 1.5 and 10 km in diameter are 10–15% smaller than their MCD counterpart in most cases (Figure 3c). The diameter variation does not exceed 15% for approximately two thirds of all (diameters) true positive detections. About 73% are within 20%. Over the 1.5 to 10 km diameter subrange, 67% have diameters that are within 20% of the MCD population (Figure 3c). This close correspondence in measured diameters between the MCD and ACD is encouraging. The CDA is not only identifying the same population as the MCD, but it is also accurately measuring their diameters. In this context, we use “accurate” in the sense that the observed differences in diameter can be expected to have minimal impact on isochrons and age derivation thereof; that is, CDA-derived isochrons and manual count isochrons should be similar. Since the final scientific use of these data sets is as tools to determine the age of a planetary surface, that is the ultimate test. This is defined and quantified below.

3.3. Uncorrelated Craters

There are two types of detections which do not correlate with the ground truth data set—craters that appear in the MCD but are not found by the algorithm and craters which the algorithm locates, but the MCD does not list. The former describes the false positive detections defined in the supporting information (Table 1 and Figure S3). The latter are the craters we discuss here. Approximately 17% (62,509) of craters detected by the CDA were not listed in the MCD. The abundance of uncorrelated detections varies with latitude and diameter. Most of the uncorrelated craters (>80%) are located below 30°S and above 40°N latitude, where the algorithm struggles a bit more due to limitations of the training data set.

3.4. Age Calculations Derived From Detection

A definitive test to evaluate the accuracy of any automated algorithm is to compare surface ages derived from manual counts with the results of the CDA counts for specific areas. We picked three different geological units with well-defined ages from the Martian geological map (Tanaka et al., 2014): a late Hesperian volcanic unit on Sinus Planum, a late Hesperian lowland unit on Isidis Planitia, and finally a late Noachian highland unit on Arabia Terra. We then produced CSFDs for these units (presented as before) using the detections from the ACD (shown in red in Figure 4) and the entries in the MCD (shown in green in Figure 4).

We fit isochrons for both CSFDs using the Hartmann and Neukum et al. (2001) chronology function and the production function of Ivanov (2001) over the same diameter range (denoted in Figure 4) within each unit. The ages calculated using both databases are within error of each other. Despite there being a slight difference in the cratering density in each diameter bin between the ACD and the MCD, the CSFD as a whole (or at least the part of the CSFD from which the isochron has been derived) produces similar ages for each unit.

3.5. Applying the Algorithm to Other Data Sets

Since a complete data set of craters >1 km already exists for Mars, determined by manual counting, the principal target of our CDA is to detect impact craters over the size range where manual counting becomes intractable (i.e., <1 km diameters). A CDA should also be flexible enough to be applied to multiple image resolutions and planetary bodies without having to retrain for all potential morphological features each time. To evaluate how effective our current algorithm is at doing this, we applied it to a ConText Camera (CTX) data set, which has a resolution of 5 m per pixel (Malin et al., 2007). We used the beta 01 version of the Murray Lab CTX mosaic (NASA/JPL/MSSS/The Murray Lab) covering an area of 8° (equals four image tiles) centered on the Gusev impact crater (14°S, 175°E).
Analysis of the four tiles of the Murray Lab mosaic over this area (~14,000 km²) took roughly 20 min. In this region, the CDA detected ~60,000 craters larger than 40 m. To investigate the accuracy of the detection, because there is no manual count to compare to, we use the results to determine the ages of specific areas. We focused on Gusev crater where independent and detailed crater counting ages are available for comparison (Parker et al., 2010). Most of these counts were performed with the aim of dating specific areas in and around this impact crater to retrace its geological history. For example, Parker et al. (2010) counted impact craters on several geological units of Gusev Crater using the Mars Express High-Resolution Stereo Camera.

Figure 6. Crater size-frequency distributions and associated ages (in gray) based on the results of our algorithm applied to the available CTX mosaic for the areas outlined in Figure 5 (a: Area 3, b: Area 5, c: Area 7, d: Area 10). For comparison, the age determined by Parker et al. (2010) is shown in red on each panel.
(Neukum & Jaumann, 2004) data (25 m per pixel). We reproduced 4 of the 14 counting areas described in Parker et al. (2010), specifically focusing on areas 3, 5, 7, and 10 (Figure 4a and Table A in Parker et al., 2010, and Figure 5 below). These regions correspond, respectively, to the Gusev crater floor, the Thira East formation, the Ma’adim Vallis, and the Highland terrains around Gusev crater. These areas were chosen according to two criteria: (1) relevant areas for such a comparison needed to exhibit a statistically large number of impact craters and (2) we required that manually measured model ages needed to show a large variability within the selected region. The model ages of the selected areas fall between 2.89 and 4.04 Gyr, thus offering a sufficiently significant range of ages from Early Amazonian to Early Noachian.

From these areas, we extracted the CSFD of the CDA detection and fitted them with an isochron by using the CraterStatsII software (Michael & Neukum, 2010), with the chronology function of Hartmann and Neukum et al. (2001) and the production function of Ivanov (2001). This is the same chronology system used by Parker et al. (2010) and what we used above. We utilized the same diameter range for the isochron fitting as Parker et al. (2010) (see supporting information of their paper). Results are shown in cumulative representation in Figure 6. For three out of the four areas, the CDA-derived ages (in black in Figure 6) are within the stated errors of Parker et al. (2010) (in red in Figure 6).

4. Conclusion and Perspectives

The manual identification and size estimation of an impact crater on any terrestrial body is subject to several sources of errors. Robbins et al. (2014) shows, for the Moon, that the estimation spread of the size and location of impact craters between several experts could reach 20%. The subjective quality of the process is revealed in variability in age derivation over the same area obtained from different experts and even between two counts performed by the same person at different times (Robbins et al., 2014). A CDA can avoid human limitations (a computer does not get tired, lose attention, the analysis is objective and reproducible, and they can count craters to the limit of available image resolution), but previous efforts have revealed their own set of issues. We constructed a CDA and validated it against manual count data sets obtaining a variability that is of a similar order to that observed between human analysts. We used the results to generate isochrons and, thus, ages indistinguishable from literature values. By definition, it is, therefore, a tool that can be used to extract geologically meaningful information from planetary data sets.

We trained the CDA based on the YOLOv3 algorithm on a selection of Controlled THEMIS DayIR, 100 m per pixel, images using the Robbins and Hynek (2012) database as a ground truth data set. We evaluated the detection accuracy of our algorithm through comparison of all THEMIS DayIR mosaic data covering the latitudinal band 65°N and 65°S of the equator to the Robbins database:

1. For the detected population above 1 km in diameter, 92% are true positive detections—meaning they were found in both the manual and automatic counting databases. Only 17% of the MCD was not accounted for by the algorithm. These nondetections mainly comprise craters located at high latitude (>45°), secondary craters, and large complex impact structures.
2. Among the true positive detections, although 70% have larger diameters relative to the same crater in the MCD, the difference is less than 20% in the majority of these cases.
3. The CDA accurately finds the latitude and longitude of 95% of the craters detected (position errors of less than 0.5%).
4. The training data set is optimized to detect simple craters between 10 and 100 pixels in diameter. This translates to a diameter range of 1.5 to 10 km on THEMIS resolution (100 m per pixel) data sets.

As a trial, we applied the CDA to a higher-resolution imagery data set without retraining, that is, CTX images (5 m per pixel) over areas of Gusev Crater previously studied by Parker et al. (2010). We demonstrated that it generates derived model ages that are indistinguishable (within error) from the literature.

Our algorithm is capable of identifying impact craters on two different imagery data sets (THEMIS Day IR and CTX) reproducing consistent scientific results. These results are encouraging. With further training of more complex data sets, we anticipate that it will be able to characterize and identify a larger variety of crater types, constraining their degradation, and distinguish other features of interest. It is interesting to note that, based on our experience with the CTX comparison, little extra training may be needed before applying the CDA to higher-resolution image data sets.
Acknowledgments
This work was supported by the resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government, Curtin University, and the Government of Western Australia. G. K. B. acknowledges the Australian Research Council (DP170102972) for funding of this project. The authors acknowledge the following links which include data sources for access to NASA mission data [THEMIS DayIR maps (https://tinyurl.com/y567542), CTX (http://murray-lab.caltech.edu/CTX/indexhtml), Mars Crater Database (https://tinyurl.com/y5c2qy6)] as well as open source software [YOLOv3 (https://github.com/qqiudl/darknet and https://github.com/AlexeyAB/darknet)] as well as as open source software [YOLOv3 (https://github.com/qqiudl/darknet and https://github.com/AlexeyAB/darknet), ImageMagick (https://imagemagick.org/index.php)]. Data usage: The results of the CDA as applied to the THEMIS global data set are available in The results of the CDA as applied to the THEMIS global data set are available in The results of the CDA as applied to the THEMIS global data set are available in B. L. Jolliff, M. A. Wieczorek, C. K. Shearer, C. R. Neal (2006). New views of the Moon. Hartmann, W. K., & Neukum, G. (2001). Cratering chronology and the evolution of Mars. Lee, C. (2019). Automated crater detection on Mars using deep learning. Michael, G. G., & Neukum, G. (2010). Planetary surface dating from crater size frequency distribution measurements: Partial resurfacing events and statistical age uncertainty. Earth and Planetary Science Letters, 294, 223–229. https://doi.org/10.1016/j.epsl.2009.12.041 Morrison, D., & Cruikshank, D. P. (1974). Physical properties of the natural satellites. Science, 185(4), 641–739. Neukum, G., Ivanov, B. A., & Hartmann, W. K. (2001). Cratering records in the inner solar system in relation to the lunar reference system. In Chronology and Evolution of Mars Space Sciences Series of ISSI (pp. 55–86). Dordrecht: Springer. https://doi.org/10.1007/978-94-017-1053-0_3 Neukum, G., & Jaumon, G. (2004). HRSC: The high resolution stereo camera of Mars Express, Mars Express: The Scientific Payload 17-36. https://www.esa.int/Enspub/Sp1240/Sp1240web.pdf Norman, C. J., Paxman, J., Benedix, G. K., Tan, T., Bland, P. A., & Towner, M. C. (2018). Automated detection of craters in Martian satellite imagery using convolutional neural networks. Planetary Science Informatics and Data Analytics Conference. Abstr. #6004. Parker, M. van K., Tanja Zegers, T., Kneissl, T., Ivanov, B., Foing, B., & Neukum, G. (2010). 3D structure of the Gusev Crater region. Earth and Planetary Science Letters, 294, 411–423. https://doi.org/10.1016/j.epsl.2010.01.013 Proctor, L. M., Kiczynski, J. M., Byrne, P. K., Denevi, B. W., Head, J. W., Fassett, C. I., et al. (2016). The first global geological map of Mars. 47th Lunar and Planetary Science Conference, Abstr. #1245. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 779-788. Redmon, J., & Farhadi, A. (2018). YOLOv3: An incremental improvement. https://arxiv.org/abs/1804.02767v1 Robbins, S. J. (2019). A new global database of lunar impact craters ≥1–2 km: 1. Crater locations and sizes, comparisons with published databases, and global analysis. Journal of Geophysical Research: Planets, 124, 871–892. https://doi.org/10.1002/2018JE005592 Robbins, S. J., Antonenko, I., Kirchhoff, M. R., Chapman, C. R., Fassett, C. I., Herrick, R. R., et al. (2014). The variability of crater identification among expert and community crater analysts. Icarus, 234, 109–131. https://doi.org/10.1016/j.icarus.2014.02.022 Robbins, S. J., & Hynek, B. M. (2012). A new global database of Mars impact craters≥1 km: 1. Database creation, properties, and parameters. Journal of Geophysical Research, 117, E04004. https://doi.org/10.1029/2011JE003966 Robbins, S. J., & Hynek, B. M. (2014). The secondary crater population of Mars. Earth and Planetary Science Letters, 400, 66–76. https://doi.org/10.1016/j.epsl.2014.05.015 See, J. E., Howe, S. R., Warm, J. S., & Dember, W. N. (1995). Meta-analysis of the sensitivity decrement in vigilance. Psychological Bulletin, 117, 230–249.
Sermanet, P., Eigen D., Zhang, X., Mathieu, M., Fergus, R., & Lecun, Y. (2014). OverFeat: Integrated recognition, localization and detection using convolutional networks. (https://arxiv.org/abs/1312.6229v4)

Shoemaker, E. M., & Hackman, R. J. (1962). Stratigraphic basis for a lunar time scale. The Moon, 14, 289–300.
Shoemaker, E. M., Hackman, R. J., & Eggleton, R. E. (1963). Interplanetary correlation of geologic time. ed. H. Jacobs. Advances in Astronomical Sciences 8, 70–89.

Skinner, J. A., Jr., Hare, T. M., & Tanaka, K. I. (2006). Digital renovation of the atlas of Mars 1:15,000,000-scale global geologic series maps, 37th Lunar and Planetary Science Conference, Abstr. #2331.

Stöffler, D., & Ryder, G. (2001). Stratigraphy and isotope ages of lunar geologic units: Chronological standard for the inner solar system (pp. 9–54). Dordrecht: In Chronology and Evolution of Mars Space Sciences Series of ISSI. Springer. https://doi.org/10.1007/978-94-017-1035-0_2

Strom, R. G. (1979). Mercury—A post-Mariner 10 assessment. Space Science Reviews, 24, 3–70.

Tanaka, K. L., Skinner, J. A. Jr., Dohn, J. M., Irwin, R. P. III, Kolb, E. J., Fortezzo, C. M., et al. (2014). Geologic map of Mars. Retrieved from http://pubs.usgs.gov/sim/3292/

Wilhelms D. E., & McCauley J. F. (1971). Geologic map of the near side of the Moon, DOI:10.3133/i703.