Automated crater detection on Mars using deep learning

Christopher Lee

Toronto, Ontario, Canada

Abstract

Impact crater cataloging is an important tool in the study of the geological history of planetary bodies in the Solar System, including dating of surface features and geologic mapping of surface processes. Catalogs of impact craters have been created by a diverse set of methods over many decades, including using visible or near infra-red imagery and digital terrain models.

I present an automated system for crater detection and cataloging using a digital terrain model (DTM) of Mars — In the algorithm craters are first identified as rings or disks on samples of the DTM image using a convolutional neural network with a UNET architecture, and the location and size of the features are determined using a circle matching algorithm. I describe the crater detection algorithm (CDA) and compare its performance relative to an existing crater dataset. I further examine craters missed by the CDA as well as potential new craters found by the algorithm. I show that the CDA can find three–quarters of the resolvable craters in the Mars DTMs, with a median difference of 5-10% in crater diameter compared to an existing database.

A version of this CDA has been used to process DTM data from the Moon and Mercury (Silburt et al., 2019). The source code for the complete CDA is available at `https://github.com/silburt/DeepMoon`, and Martian crater

Preprint submitted to Journal Name *January 17, 2019*

datasets generated using this CDA are available at `https://doi.org/10.5683/SP2/MDKPC8`.

Keywords: Mars craters, Digital Terrain Model, Deep Learning, Convolutional Neural Network

1. Introduction

Information on crater populations and spatial distributions provide important constraints on the geological history of planetary surfaces. Regional differences in crater distributions and population characteristics can be used to constrain geologic processes and stratigraphy (Cintala et al., 1976; Wise and Minksowski, 1980; Barlow and Perez, 2003; Barlow, 2005), and crater populations can be used to estimate the age of surface features as well as constrain the timescale of surface processes (Arvidson, 1974; Soderblom et al., 1974; Craddock et al., 1997; Stepinski and Urbach, 2009; Tanaka et al., 2014). To enable such research, impact craters need to be identified, measured, and counted using imagery of a planet's surface (Barlow, 1988; Salamunićcar and Lončarić, 2008; Robbins and Hynek, 2012).

However, the task of creating a dataset of crater locations has traditionally been a time–consuming process of manually identifying craters in printed maps (Barlow, 1988) or digital imagery (Robbins and Hynek, 2012). Recent advances in automated image processing have lead to the development of semi-automated or fully automated Crater Detection Algorithms (CDAs, Stepinski et al., 2009; Di et al., 2014; Pedrosa et al., 2017; Silburt et al., 2019). These CDAs use different approaches and datasets, but each method attempts to identify crater–like features on the surface using digital imagery or elevation datasets and apply image processing techniques to isolate the crater features. An important benefit of these CDAs is that it becomes possible to automate large parts of the crater–finding process and reduces the effort required after the initial implementation.

Automated CDAs have tunable parameters that can be optimized for the imagery or elevation dataset being processed. In designing the algorithms, a curated list of crater locations and images are used in a "training" step to adjust these parameters. Once trained, the CDA can be applied to larger datasets from the same body or even applied to different planetary bodies through "transfer learning" (Silburt et al., 2019).

In this work I use an automated CDA based on a Convolutional Neural Network (CNN, Goodfellow et al., 2016) to identify *circular* crater–like features in a martian Digital Terrain Model (DTM). I perform three experiments with the CDA to characterize its performance on the DTM under various assumptions. In two of the experiments I attempt to find rings associated with the crater rims as in Silburt et al. (2019) using CNNs trained on lunar data (the Silburt et al. (2019) CDA) and martian data. In the third experiment, I train the CNN to find *disk* structures associated with the entire crater. The latter method is commonly used in image segmentation methods (Ronneberger et al., 2015) and a similar method was developed by Stepinski et al. (2009).

The CNN used in this work uses a standard "UNET" architecture (Ronneberger et al., 2015) that is commonly used in image segmentation and processing, and similar CNNs have been applied to identification of tumors in medical images (Çiçek et al., 2016), identification of radio frequency interference in astronomical data (Akeret et al., 2017), and crater detection on the Moon using a DTM from the Lunar Reconnaissance Orbiter (Silburt et al., 2019). The architecture of the CNN is not the primary purpose of this work, and the reader is referred to the referenced work for an in–depth discussion of the methodology.

The remainder of this paper is organized as follows: in section 2 I review prior work in developing automated CDAs; in section 3 I describe of the crater detection algorithm, the training and data processing involved, and the structure of the experiments; in section 4 I discuss the metrics calculated for each experiment and examine the new crater datasets in detail; finally, in section 5 I provide concluding remarks.

The software used to make CDA described here is based on the work of Silburt et al. (2019) with three modifications: The source data used here retains the 16–bit raw precision of the source DTM compared to 8–bit image used in Silburt et al. (2019); a disk–finding CNN is implemented with an additional processing step, discussed later; the distance and size thresholds used to determine duplicates and matches in the database were reduced to 0.25 of the crater diameter (from 2.6 and 1.8 diameter units).

The original code is available at `https://github.com/silburt/DeepMoon.git` , updates and modifications to the code can be found at `https://github.com/eelsirhc/DeepMars.git` . The datasets generated here, along with ancillary data, can be found at `https://doi.org/10.5683/SP2/MDKPC8`. Similarly to Silburt et al. (2019), I use Keras (Chollet, 2015) version 2 with Tensorflow (Abadi et al., 2016) to build, train, and test the model. In training the model I used an Nvidia 1080 Ti GPU using the CUDA and CUDNN support libraries, but the code is compatible with Intel and AMD CPUs.

2. Prior Work

One of the first large global databases for Mars was created by Barlow (1988) using printed maps from Viking orbiters and included 25,826 craters with a diameter greater than 8km. This dataset has been updated since then (Barlow, 2010) with 42,483 craters, and other datasets are available (Rodionova et al. (2000) with 19,308 craters, Salamunićcar and Lončarić (2008) with 57,633 craters). The most comprehensive dataset for Mars craters is that derived from the Thermal Emission Imaging System (THEMIS) instrument by Robbins and Hynek (2012). The Robbins and Hynek (2012) dataset includes 383,343 craters with diameters greater than 1km, including 30,473 craters above 8km diameter. These craters were identified in 100m/pixel scale THEMIS IR imagery using a customized manual image processing pipeline. The Robbins and Hynek (2012) dataset is reported to be statistically complete to 1km diameter for the majority of Mars covered by the source THEMIS dataset, reflected in the power law distribution following the expected distribution to diameters of 1km or lower (Arvidson et al., 1979). In this work I consider craters with a diameter greater than 4km based on the resolution limit of the input DTM.

In contrast to the attempts to catalog martian craters using manual methods, automated CDAs have not been extensively used to generate *global* catalogs of martian craters, but have been used to catalog small test areas containing a mixture of crater types. Two common metrics used in machine learning comparisons are *precision* and *recall*, whose mathematical definitions are given in the next section. *Precision* is the fraction of craters found by the CDA that exist in a target dataset (usually a subset of Barlow (1988) or Robbins and Hynek (2012)), and the *recall* is the fraction of craters in the target dataset that are found by the CDA. In the calculation of precision, the effect of detected craters that are real, but do not exist in the target dataset, is to decrease the precision.

Stepinski et al. (2009) developed the AutoCrat CDA using the 128pixels/degree DTM from the Mars Orbiting Laser Altimeter (MOLA) instrument. The AutoCrat CDA combines a "rule-based" module that applies gradient–based algorithms to identify local depressions as possible craters, followed by a "machine-learning" module that applies a decision tree algorithm (Witten and Frank, 2005) to determine whether the feature is a crater or not. The decision tree algorithm is used to differentiate craters from non–crater depressions using diameter, depth, and shape parameters as factors. Only a small fraction of the planet is covered in Stepinski et al. (2009), with a reported precision of 42% (1,544 known craters out of 3,666 detections) and a recall of 72% (1,544 out of 2,144 known craters were found). A global database generated using this CDA was reported in Stepinski and Urbach (2009) with 75,919 craters larger than 1.37km.

Di et al. (2014) developed a CDA that also processed DTM images. Their CDA uses a sliding window correlator to find and highlight crater edge features and a circular Hough transform to transform those highlighted crater edges into locations and sizes. Di et al. (2014) reports on the CDA performance for three sites with 11,868 craters, but they do not provide an explicit calculation of precision and recall. Di et al. (2014) reports finding 934 craters in total, with 114 false positives (a precision of 87%) with a recall rate using the same data (their table 2) of 74% for craters with a diameter greater than 6km, but a recall rate of less than 10% for all craters tested.

Pedrosa et al. (2017) developed a CDA using thermal imagery from THEMIS. The CDA processes THEMIS IR imagery by first identifying geophysical depressions using a 'watershed' transform (to find virtual floodplains) and then within each watershed identified the local minima as possible craters. A circle template matching algorithm is then used to compare the crater features to a characteristic ring representing the crater rim. Pedrosa et al. (2017) reports a precision and recall of 65% and 91%, respectively (their figure 7), compared to a target dataset of 3,600 craters. The template matching system employed by Pedrosa et al. (2017) is similar to the method used in Silburt et al. (2019) and here to find the location of each crater.

Salamunićcar et al. (2011) provides a summary of many more automated CDAs, and a discussion of the effect of combining multiple CDAs into one dataset. The various methods used in the CDAs are essentially the same as the CDA developed here. A correlation function is used to identify features that identify a crater — edges in Di et al. (2014), disks in Stepinski et al. (2009), opposing crescents in Pedrosa et al. (2017). Once the crater is identified a circle finding algorithm is used to determine the location and size — a Hough transform in Di et al. (2014), a sliding window correlator in Pedrosa et al. (2017). In the CDA developed here, discussed in detail in the next section, the CNN implements a sequence of correlation functions to identify and highlight the crater, followed by a circle matching algorithm to determine the location of the craters.

3. Methods

3.1. Input dataset

The source digital terrain model (DTM) for this work is the "Mars HRSC MOLA Blended DEM Global 200m v2" (Fergason et al., 2018) dataset available from the Astrogeology Science Center website (https://astrogeology.usgs.gov). This map is a blend of digital terrain models derived from the Mars Orbiter Laser Altimeter (MOLA, Smith et al., 2001) aboard the Mars Global Surveyor spacecraft, and the High-Resolution Stereo Camera (HRSC, Jaumann et al., 2007) aboard Mars Express. The stated scale of the dataset is 200m/pixel horizontally, chosen as a compromise between the 463m/pixel scale of MOLA and the 50m/pixel scale of HRSC. However, the HRSC data covers only 44% of the planet, so more than half of the planet is interpolated from MOLA dataset at 463m/pixel scale. Some regions have no data from either spacecraft. The stated accuracy of each point is 100m horizontally and at best 1m radially (Fergason et al., 2018). The total image size of the DTM is 106694×53347 pixels with 16–bit resolution for the elevation data, using a simple cylindrical (Plate Carrée) projection. The effective resolution of this source image is $\frac{1}{296}^{\text{th}}$ of a degree, and $\frac{1}{2}$ m vertically (better than the resolution of the input data).

The CDA takes as input a 256×256 pixel 8–bit image taken from the larger DTM and attempts to identify craters within this image. The size and bit resolution are chosen to fit in the memory limits of the hardware being used. To prepare an image for use with the CDA I use the following steps:

1. A square sample is extracted from the DTM and resampled into the required 256×256 size.

2. The bit resolution of the image is rescaled from the 16–bit source to the 8–bit resolution required for the CDA. This step occurs after resampling the image to a smaller region to mitigate the effect of the large altitude variation on Mars.
3. The image is orthographically projected using the Cartopy Python package (Met Office, 2018) to provide an image with near–constant linear scale instead of the constant angular resolution of the Plate Carrée projection.
4. Padding is added to the image as required to fill in the square image after projection. The Orthographically projected image always occupies fewer pixels than the Plate Carrée source image.

The size of the image sample (step 1) is chosen from a list of sizes from 512 to 16,384 pixels to provide a range of scales from 400m/pixel to 12.8km/pixel. The full dataset is constructed by sampling the entire planet at a range of pixel scales and with overlapping regions between adjacent images, requiring 55,000 images in total. I also tested an additional 150,000 images sampled at the original scale of the DTM (200m/pixel) but the performance of the CDA degrades substantially because of the coarser scale of the majority of the input DTM. An alternate method used by Silburt et al. (2019) was to select the location and size of the images at random, which provides similar statistical results to the systematic method above, but would not guarantee planet-wide coverage.

3.2. Experiments

In this work I performed three experiments with the CDA. The first experiment uses a CNN trained on Lunar data (Silburt et al., 2019) to find ring structures associated with the crater rim. This CNN has not been previously trained on Mars crater observations and is an example of *transfer learning*.

The second experiment modifies the first by training the CNN using a subset of the Mars crater imagery without using the previously Moon trained CNN. The target data for the training is derived from a human–generated crater database (Robbins and Hynek, 2012) using high–resolution infra–red imagery. In the second experiment the CNN is trained to identify rings associated with the rims of craters, and a summary of the training method is provided in the next subsection.

In the third experiment the CNN is modified to identify *disks* associated with craters, instead of rings. This approach follows the same training methodology as the ring finding CNN, but with a modified training dataset and crater matching algorithm. To keep the comparison as close as possible I use the same image locations for both trained CNNs, and in the *disk* finding CDA I convert crater features highlighted by the CNN to ring structures before attempting to locate the craters. This disk–ring conversion makes comparison easier between the CNNs but is not necessarily an optimized algorithm for the disk finding CDA.

3.3. Training and Validation

Training and validation follows the method outlined in Silburt et al. (2019), and an example is given in figure 1. A sample image is taken from the dataset (figure 1–left) and the locations of resolvable known craters are taken from the Robbins and Hynek (2012) and drawn as white pixels on an otherwise black 'image' (figure 1–center). The CNN is then trained by exposure to a large number of images and trained to encode the DTM into the binary ring image (figure 1–center). Figure 1 is one of the 15 images in the dataset that includes a majority of Gale crater in the image – this figure is centered on 137 degrees east longitude, 8 degrees south longitude, at 3.2km/pixel scale.

Figure 1–right super-poses the input DTM image with the known craters (Robbins and Hynek, 2012) in red and craters found by the ring finding Mars–trained CDA in blue. In this image overlapping blue and red circles identify craters correctly identified by the CNN (though some are displaced spatially), red circles with no blue counterpart are missed craters (*false negatives* in machine learning terminology), while blue circles with no red counterpart are features incorrectly identified as craters by the CNN (*false positives*). In principle, the false positives might be new craters, but I will suggest later that the majority are not new craters, although some are genuine circular features (e.g., paterae).

For the CNN trained with martian craters, I use 30,000 images distributed in location and scale in the training dataset (5,000 images were reserved for a testing stage during the training), and 25,000 images in the validation dataset. The images are distributed geographically so that both datasets contain unique images but similar spatial distributions. The image extents are also distributed between datasets, with similar numbers of each scale in each dataset. The small number of large geographically extended images (fewer than 500) means they are unevenly distributed between the two datasets.

The CNN training is assumed complete either when accuracy on the train-

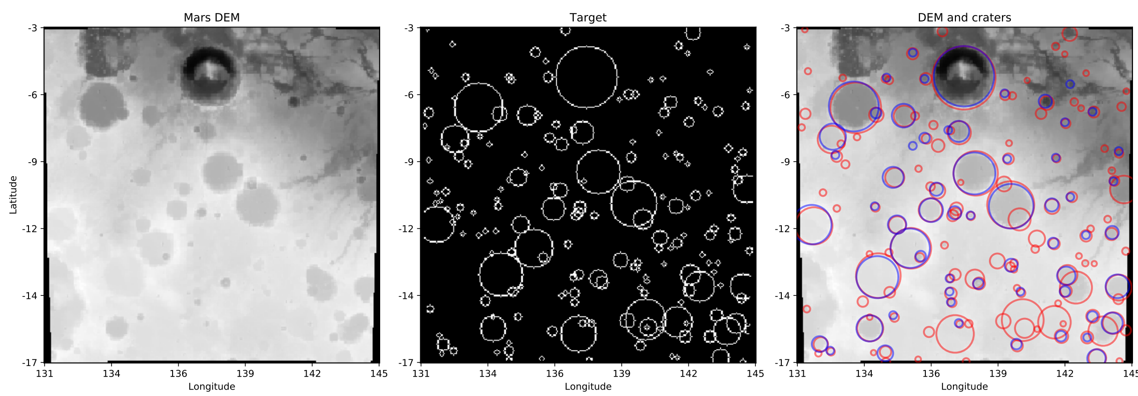

Figure 1: Example DTM image (left),target map (center), and identified craters (right). The DTM is extracted from the source HRSC+MOLA map at 137E longitude, 8S latitude with a resolution of 18 pixels per degree (approximately 3.2 km/pixel). Gale crater is located at the center–top of the image, and is found by the CNN in this example. The right plot does not include craters that are smaller than 4 pixels in diameter for clarity. Red circles show craters from the Robbins and Hynek (2012) dataset, blue circles show craters found by the CDA.

ing dataset stops improving, or when the number of iterations has reached the maximum allowed. In Silburt et al. (2019), the CNN was trained with 30,000 images of the Moon for 4 complete iterations. For the Mars trained CNN I allowed up to 30 iterations, with a typical training taking 10 iterations before the accuracy stopped improving.

The CNN does not produce a location or size for each crater in the image, but instead transforms the DTM image into a binary image that highlights topographic features that are related to craters. The CNN is best at highlighting features that are between 10 and 60 pixels in diameter in any particular image, resulting in a large number of missing craters in each image. Small craters are represented by too few pixels to be positively identified. Large craters become diffuse or incomplete circles fall below the detection threshold. In figure 1, Gale crater was among the largest identified features, even though a few larger craters are visible in the image.

3.4. CNN Processing

The location and size of the craters in the CNN images are determined using the `match_template` algorithm from scikit-image (van der Walt et al., 2014). The `match_template` algorithm finds the location and size of each circular feature by maximizing its correlation with a *template circle* of known size. To allow comparison with the Silburt et al. (2019) study I keep the same threshold parameters for the circle matching algorithm, though small improvements may be possible with more extensive re–training of the algorithm. For each crater map generated by the CNN the location and size of craters are found with the following steps:

1. A candidate circle size is chosen and used to generate a template image.

2. The candidate circle is compared against all locations in the binary crater image generated by the CNN. The resulting map becomes a "heat map" of correlation between the candidate circle and the template.
3. Where the correlation between the crater map and the candidate circle exceeds a confidence threshold that location is identified as the crater location and the size of the candidate circle is used as a crater size.

This template matching process is conducted for circles with integer radii from 5 pixels to 40 pixels. For architectural reasons the CNN rarely predicts craters smaller than 10 pixels in diameter or larger than 60 pixels, with typical minimum and maximum diameters of 10 pixels and 30 pixels, respectively. The circle matching algorithm performs poorly for circles with a diameter smaller than 10 pixels where it considers diffuse segments of larger craters as potential small craters, resulting in 'rings' of small craters around larger craters. Duplicate craters are removed in each image by identifying craters that are within 0.25 diameter units in size and within 0.25 diameter units in location of another crater. These values are smaller than those used by Silburt et al. (2019).

The result of this post-processing is a list of unique craters found in each input image, before any comparison with the Robbins and Hynek (2012) database. In the right panel of figure 1 this post-processing produced the blue circles, while the Robbins and Hynek (2012) craters with a diameter of at least 4 pixels are shown as red circles.

The disk–finding CNN is trained to highlight craters by replacing the crater in the DTM with a solid disk in the binary image, instead of a ring surrounding the crater. After the disk CNN has processed the DTM scene a

Sobel and Feldman (1968) transform is used to convert this disk into a ring to emulate the output of the ring–finding CNN. After this additional step, the analysis follows the ring finding method above.

A downside of the disk finding method is that overlapping craters are not easily separated. An overlapping crater system is filled with the same binary value so that overlapping craters are identified as single non-circular features and are rejected by the circle matching algorithm. The CDA developed by Stepinski et al. (2009) is also a type of disk finding algorithm but uses a pre-processing step of Gaussian blurring to provide images with 3 different length scales to overcome the problems associated with small craters within large craters. The blurring does not separate similarly sized overlapping craters.

3.5. Post Processing

The image dataset contains 55,000 images with resolutions ranging from 150 pixels/degree to 4 pixels/degree covering the planet. As a result, a single location would appear at up to 7 different resolutions in 15 ± 6 images (the variation is due to the use of overlapping images). For example, Gale crater appears at 5 resolutions in 9 images.

Each of the 55,000 images is processed by the CNN and template matching algorithm to find craters independently of the other images. The location and size of each crater is found in pixel space during the template matching stage, and then converted into geographic coordinates using the known limits of each image and the orthographic projection parameters.

As a result of the overlapping images and multiple resolutions, single craters may be identified in multiple images and appear multiple times in the generated global crater list. Duplicate craters are removed by comparing the

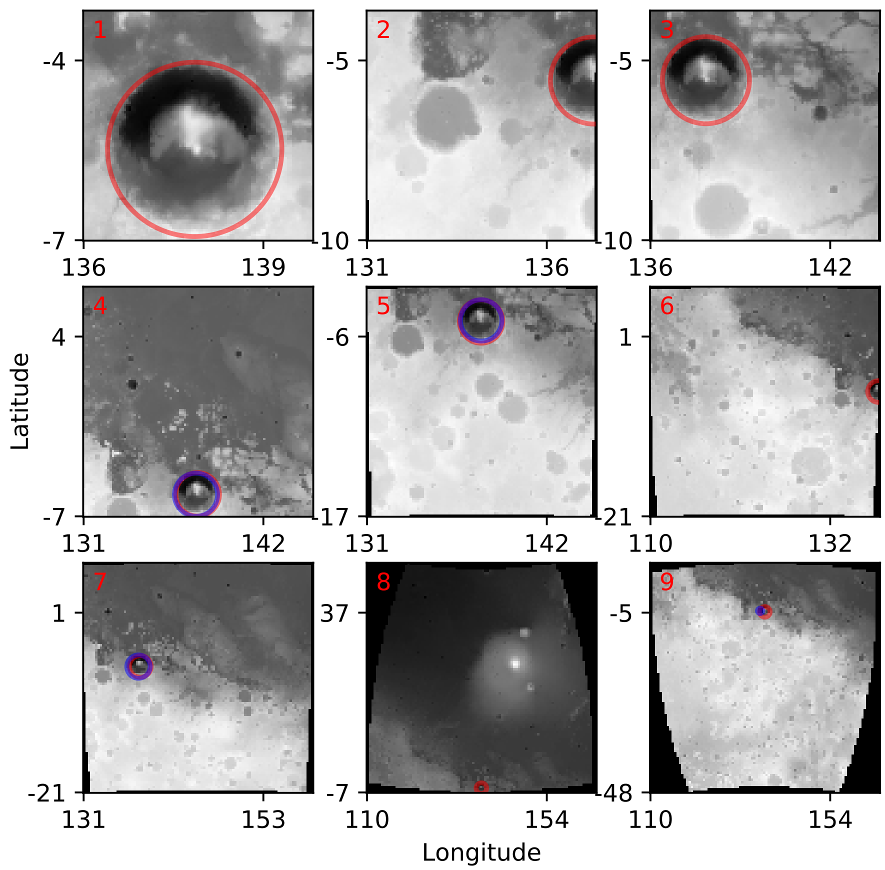

Figure 2: DTM images where more than 50% of Gale crater is contained in the image. Gale crater is highlighted with a red circle (using the Robbins and Hynek (2012) location) and a blue circle where it was identified by the CDA in each image. The CDA identified the crater in 5 images in this example data (4,5,7,8,9). The crater in images 1,2, and 3 is probably too big for the current algorithm. Image 2 and 6 only include partial circles and lie below the detection threshold.

diameter and location of the crater with other craters with similar location and size, using the same parameters as in the last section.

Figure 2 shows the 9 images that include more than 50% of Gale crater. In this example, Gale crater was found in 5 images. The 5 candidate craters are combined in the post-processing stage to provide 1 location and size for Gale crater, preferentially using the values found in the highest resolution image.

Once all duplicates are removed the final result is a database with approximately 60,000 craters found by the CDA. This list includes only craters larger than 4km in diameter, the lower limit allowed in the algorithm. Above this 4km limit, the ring CDAs found 75% of all craters listed in the Robbins and Hynek (2012) database. Above 10km diameter, the ring CDAs found more than 80% of all craters in the Robbins and Hynek (2012) database. The algorithm itself has no lower limit in geophysical space but does have a lower limit in pixel space. The circle–matching algorithm works well for circles 10 pixels in diameter or larger, and continues to work down to 6–pixel diameter circles but with a higher false positive rate. Ten pixels in diameter represents a physical limit of 4km in diameter using the 400m/pixel scale images. As a comparison, Robbins and Hynek (2012) used 100m–230m/pixel THEMIS infra-red imagery that covers more of the planet than the DTM used here and the crater size limit reported in Robbins and Hynek (2012) is 1km, or 10 pixels at the highest THEMIS image resolution. Robbins and Hynek (2013) suggested a lower diameter limit of 10km for MOLA derived DTM data, noting that imagery derived DTMS (e.g., from HRSC) are better at resolving smaller craters.

3.6. Accuracy Metrics

In each experiment, the performance of the CDA is measured against the Robbins and Hynek (2012) database using a number of standard metrics. The crater locations in the CDA database and the Robbins and Hynek (2012) database are compared using the same methodology used to find duplicate craters. If the CDA finds a crater within 0.25 diameter units in location and size of a crater from the Robbins and Hynek (2012) database then it is considered a match. A *True Positive* is a match between the CNN and Robbins and Hynek (2012) database, a *False Positive* is a crater in the CDA database without a matching crater in the Robbins and Hynek (2012) database, and a *False Negative* is a crater in the Robbins and Hynek (2012) database without a matching crater in the CDA database. *True Negatives* are not used.

Using these definitions, the precision P is defined as the ratio of true positive to all identifications, and the recall R as the ratio of true positives to all craters in the Robbins and Hynek (2012) database.

$$P = \frac{T_p}{T_p + F_p}, \tag{1}$$

$$R = \frac{T_p}{T_p + F_n} \tag{2}$$

Where T_p, F_p, and F_n are the numbers of true positives, false positives, and false negatives, respectively. A high precision suggests the CDA has a high fractional true positive rate, while a high recall suggests the CDA finds a high fraction of the existing craters.

As an extreme example, the CDA could identify craters everywhere on

Mars whether they exist or not, resulting in a perfect recall (all craters are found) but almost no precision (many false positives are found). Alternatively, the CDA could identify a single crater correctly, resulting in perfect precision (no false positives) but almost no recall (many missing craters, or having many false negatives). A common metric used to balance the precision and recall is the harmonic average of the two metrics, commonly called the F_1 score,

$$F_1 = \frac{2PR}{P + R} \tag{3}$$

where the same F_1 value can be found using different combinations of precision and recall. None of these metrics reward identification of new craters that do not exist in the Robbins and Hynek (2012) database. All new identifications are assumed false positives and reduce the precision and F_1 score. The possible new crater fraction N is calculated as the ratio of false positives to the sum of false positive and Robbins and Hynek (2012) craters,

$$N = \frac{F_P}{F_P + T_R} \tag{4}$$

Where T_R is the number of true craters in the Robbins and Hynek (2012) database. This is an upper limit on the number of new craters found, and because the Robbins and Hynek (2012) database is statistically complete below the 4km limit used here it is likely that many false positive are genuine false positives and not new craters. In Silburt et al. (2019) a sample of the false positive craters was studied and they estimated that 90% of the false positive craters are new craters.

For further comparison with the Robbins and Hynek (2012) database I also calculated the difference in longitude, latitude, and diameter between the CNN craters and Robbins and Hynek (2012) craters, both in pixel units and geophysical units. Each of these metrics is calculated as the ratio relative to the smallest crater diameter in the comparison and given as the mean value and interquartile ranges for the dataset.

3.7. Error Sources

A number of sources of error are present in the experiments, from observational constraints, pixelization of the source data, and the algorithm design.

The source DTM combines HRSC and MOLA data at a stated scale of 200m/pixel. However, this is obtained by upsampling the MOLA data from 463m/pixel for 56% of the surface, and downsampling HRSC images for the remaining 44%. The smallest image scale used in this experiment was 400m/pixel, close to the MOLA laser footprint of 300m, giving a accuracy limit of .4km in crater location and diameter (i.e., 1 pixel).

The crater position is extracted by matching circular templates on images. The discrete nature of the images limits the matches to 1 pixel in any image. At the highest image resolution used this corresponds to the .4km accuracy limit above, but in images with a larger pixel scale the accuracy decreases at a corresponding rate. At the 12.8km/pixel scale for the largest images, the accuracy drops to about 6km at the equator. In practical terms, a crater is likely to be found when it is between 10 pixels and 30 pixels in diameter, making the 1 pixel uncertainty equivalent to a 3–10% error in position or diameter. When the global CDA database is generated, the highest resolution image that included the crater detection was used to obtain the best overall position and size data for each crater in the final dataset. For example, in the Gale crater example in figure 2, image 4 or 5 would be used when calculating the location and size of the crater.

Projecting the DTM from Plate Carée into orthographic and back introduces some errors depending on the extent of the image, as distortion increases away from the center point of the projection. Silburt et al. (2019) estimated an error of 2% in the crater size for typical images, which becomes larger than the pixelization errors for craters larger than about 50 pixels in diameter. Few craters were found larger than 30 pixels in diameter so the contribution of this error is negligible.

Finally, algorithmic implementation also introduces some uncertainty. In the CNN step, the image bit resolution is limited to 8–bits of data, which for large images with vertically extended topography would obfuscate shallower craters – 1km of vertical extent requires a vertical resolution of 4m at best, while an image that includes Olympus Mons and the surrounding terrain might be limited to 100m vertical resolution. In the template matching step, spurious matches can occur when comparing small candidate craters against large craters as the template matches along the crater wall, or comparing candidate circles against the space *between* two or more nearby craters if the 'void' between the craters can be identified as a crater.

4. Results

Table 1 gives the metrics for the three different experiments. All of the metrics are calculated using the same source images but the training of each network is different. The "Moon" trained CDA uses the network generated

	Moon (image)	Moon (global)	Mars (image)	Mars (global)	Disk (image)	Disk (global)
Crater count	9.9 ± 10.0	$57,564$	9.9 ± 10.0	$57,564$	9.9 ± 10.0	$57,564$
Craters detected	4.8 ± 5.1	$54,739$	4.9 ± 5.2	$57,767$	5.1 ± 4.9	$75,733$
Craters matched	4.3 ± 4.7	$42,445$	4.4 ± 4.8	$42,891$	4.3 ± 4.7	$39,149$
Latitude Error	4^{+2}_{-1}	2^{+1}_{-1}	4^{+2}_{-2}	2^{+1}_{-1}	4^{+2}_{-2}	2^{+1}_{-1}
Longitude Error	5^{+2}_{-2}	2^{+2}_{-1}	5^{+2}_{-2}	3^{+2}_{-1}	5^{+2}_{-2}	2^{+2}_{-1}
Diameter Error	6^{+3}_{-3}	5^{+4}_{-3}	7^{+3}_{-3}	6^{+4}_{-3}	9^{+4}_{-4}	6^{+5}_{-3}
Percentage new craters	5 ± 8	18	5 ± 8	21	7 ± 11	39
Maximum diameter (pix)	34.1 ± 20.0	–	33.5 ± 19.8	–	32.7 ± 18.9	–
Precision	90 ± 18	78	89 ± 19	74	84 ± 23	52
Recall	42 ± 21	74	43 ± 21	75	44 ± 23	68
F1	58 ± 17	76	59 ± 17	74	58 ± 18	59

Table 1: Metrics calculated for three neural network based CDAs. "Moon trained" and "Mars trained" refer to the data used to train the initial model. "Disk" trained using Mars crater imagery in training, but identified disks associated with craters instead of crater "rings". All metrics were calculated using the same martian crater images from MOLA/HRSC. Values are given as mean ± 1 standard deviation (for single values after the ±) or median ± inter–quartile range (for two values after the ±) as in Silburt et al. (2019). Precision, recall, and F_1 scores are given as percentages. Each model appears twice, with the "image" column given the per image metrics aggregated over the ensemble of 55,000 images (after removing duplicates per image), and the "global" column gives the post-processed metrics (after removing duplicates globally).

by Silburt et al. (2019) with no further training. The "Mars" trained CDA uses the network trained on a subset of the martian crater database, where the network is trained to find the crater rim. The "Disk" trained network is also trained on martian crater images, but is trained to find the whole disk of the crater, instead of just the rim. Table 1 includes data from 55,000 images derived from the global MOLA/HRSC DTM. A smaller dataset using only images that were withheld during the training phase for the "Mars trained" CNN is shown in table 2. None of the CDAs were shown the images summarized in table 2 during training.

	Moon (image)	**Moon (global)**	**Mars (image)**	**Mars (global)**	**Disk (image)**	**Disk (global)**
Crater count	8.1 ± 5.3	$32,979$	8.1 ± 5.3	$32,979$	8.1 ± 5.3	$32,979$
Craters detected	3.9 ± 3.1	$26,808$	4.0 ± 3.2	$28,198$	4.2 ± 3.2	$34,419$
Craters matched	3.4 ± 2.9	$21,732$	3.5 ± 2.9	$21,985$	3.5 ± 2.9	$19,949$
Latitude Error	4^{+2}_{-1}	2^{+1}_{-1}	4^{+2}_{-2}	1^{+1}_{-1}	4^{+2}_{-2}	1^{+1}_{-1}
Longitude Error	5^{+2}_{-2}	2^{+2}_{-1}	5^{+2}_{-2}	2^{+2}_{-1}	5^{+3}_{-2}	2^{+2}_{-1}
Diameter Error	6^{+3}_{-3}	5^{+4}_{-3}	7^{+3}_{-3}	6^{+4}_{-3}	8^{+5}_{-4}	6^{+5}_{-3}
Percentage new craters	4 ± 8	13	5 ± 8	16	8 ± 11	30
Maximum diameter (pix)	34.1 ± 20.3	–	33.6 ± 20.2	–	32.6 ± 19.1	–
Precision	90 ± 18	81	89 ± 19	78	83 ± 24	58
Recall	42 ± 22	66	43 ± 22	67	44 ± 23	60
F1	58 ± 18	73	59 ± 18	72	58 ± 19	59

Table 2: Metrics calculated using the validation dataset as in table 1, but for a subset of the images not used in training the "Mars trained" or "Disk trained" networks.

When training machine–learning algorithms there is a risk of 'overfitting' where the algorithm becomes significantly better (by some metric) on the dataset it is trained with, at the expense of performing poorly on data it has not been shown. This overfitting can be seen when the precision (or recall, or F1 score) of the algorithm is much higher for a 'training' dataset than an unseen 'validation' dataset. Comparing the results in table 1 and 2, the metrics calculated for the validation dataset and the complete (validation and training) dataset suggests the networks are not overfitting the training data. This is reinforced by the performance of the "Moon" trained CDA that has never been trained using the Mars dataset. Differences between the global and validation metrics for this CDA reflect statistical differences in the performance of the CDA on the two datasets.

In the following subsections I examine the performance of the CDA in more detail. The results are separated by the type of detection: section 4.1 examines all CDA crater detections in comparison to the Robbins and Hynek (2012) dataset; section 4.2 examines at the matched (*true positives*) in the CDA datasets; section 4.3 examines the craters missed by the CDA (*false negatives*), and I use the extended data provided in the Robbins and Hynek (2012) dataset to identify the characteristics of the those missing craters; finally, section 4.4 examines the craters found by the CDAs that do not exist in the Robbins and Hynek (2012) dataset – the false positives.

4.1. All Craters

First, I compare the complete dataset generated by CDAs to the Robbins and Hynek (2012) dataset. Figure 3 shows the crater distribution binned by diameter following the power law distribution used in Robbins and Hynek (2012), and shows good agreement between the CDAs and the expected power law distribution. I used 16 bins per octave (Robbins and Hynek, 2012) of crater size instead of the 2 bins used in (Stepinski et al., 2009) and Arvidson et al. (1979). The discretization present in the diameter measurements from the CDAs has been removed from the data by applying a Gaussian noise multiplier (with magnitude of 5%, smaller than the global mean diameter error in table 1 of 7%) to each data point. With only 2 bins/octave (Arvidson et al., 1979) the distributions would agree with each other without the need for the de–aliasing jitter in the CDA data. The peaks at 8km and 16km are residuals of this jittering process and represent the smallest crater sizes found in the most common image resolutions used in the experiments. The CDA finds 80% of the craters larger than 10 km diameter listed in the Robbins and Hynek (2012) database, and 75% of craters larger than 4km in diameter. Craters below 4km are omitted from this dataset because of the lack of DTM data that resolve these craters.

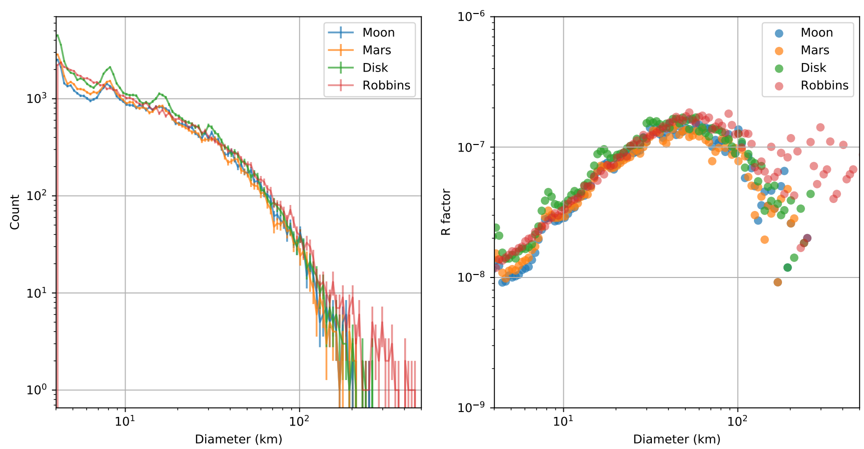

Figure 3: (left) Crater population as a function of crater diameter (km) for the datasets generated by the CDAs. (Right) R factor (Arvidson et al., 1979) for the same dataset. The raw dataset from the CDA contains aliasing due to the small number of image resolutions used in the algorithm. This discretization has be removed from the data by applying a random jitter to the crater sizes equal to 5%, smaller than the mean diameter error over all CDA datasets in table 1.

Table 3 gives the crater numbers in each of the geologic unit types listed in Tanaka et al. (2014) for the 3 CDAs and the Robbins and Hynek (2012) database. The numbers are similar in the two ring–finding CDAs and the Robbins and Hynek (2012) database, although there are craters listed in the CDA datasets that are not present in the Robbins and Hynek (2012) database (the TPR percentage shown in the table reflects this). The disk–

		Apron	Basin	Highland	Impact	Lowland	Polar	Transition	Volcanic
Robbins	Count	116	467	41,749	3,016	2,858	660	3,587	5,112
Mars	Count	128	585	40,181	3,129	3,322	911	3,645	5,866
	TPR (%)	49	53	79	72	68	44	68	69
Moon	Count	108	521	38,376	2,990	3,089	785	3,397	5,473
	TPR (%)	56	58	81	76	73	52	72	74
Disk	Count	218	1,008	48,328	4,020	5,314	1,389	5,716	9,740
	TPR (%)	20	26	59	52	42	26	40	39

Table 3: Distribution of craters by geologic unit type given in Tanaka et al. (2014). The 'Robbins' row gives the distribution of craters derived from craters in the Robbins and Hynek (2012) database. The True Positive Rate (TPR) gives the percentage of craters found by the CDA that exist in the Robbins and Hynek (2012) database.

finding CDA tends to find many more craters in all geologic units and has more false positives (lower TPR) as a result. Figure 4 shows the same results but binned by longitude and latitude instead of geology. The two ring–finding CDAs tend to under–predict craters in regions with many craters, and over–predict in regions with few craters. The disk–finding CDA over–predicts the number of craters almost everywhere.

4.2. Matched Craters

The Robbins and Hynek (2012) dataset used here contains 57,564 craters greater than 4km in diameter. The ring CDAs tested find 75% of the craters in the Robbins and Hynek (2012) dataset with a median difference in location of 2% and diameter of 5% measured in geophysical units relative to the crater diameter. This difference is typical of variability between human analysts in crater studies (Robbins et al., 2014). In raw pixel terms, the differences in position and size between the CDA and Robbins datasets are typically 1 or 2 pixels. The distribution for each metric in pixel space is shown in figure 5,

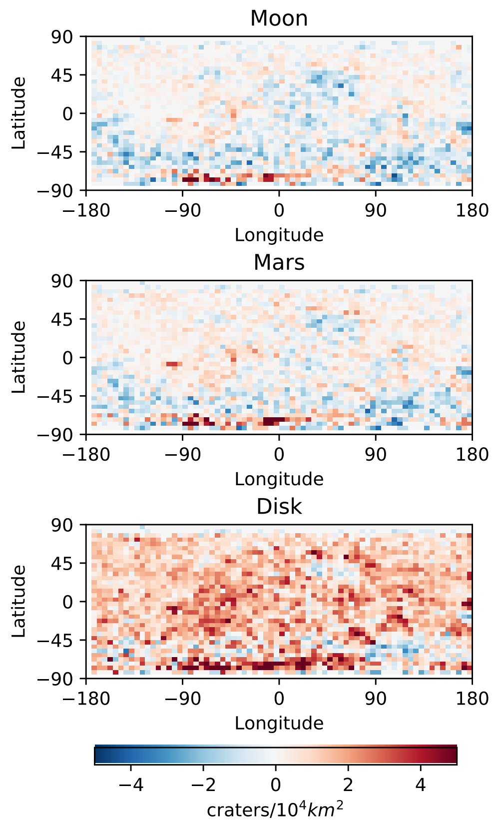

Figure 4: Plate Carée maps of the crater number predictions from the CDA relative to the Robbins and Hynek (2012) dataset, binned into 5 degree square bins and scaled to represent the number of craters per 10,000 square kilometer predicted by each CDA in excess of the Robbins and Hynek (2012) database. Positive numbers (reds) represent an over–prediction by the CDA and negative numbers (blues) represent an under–prediction.

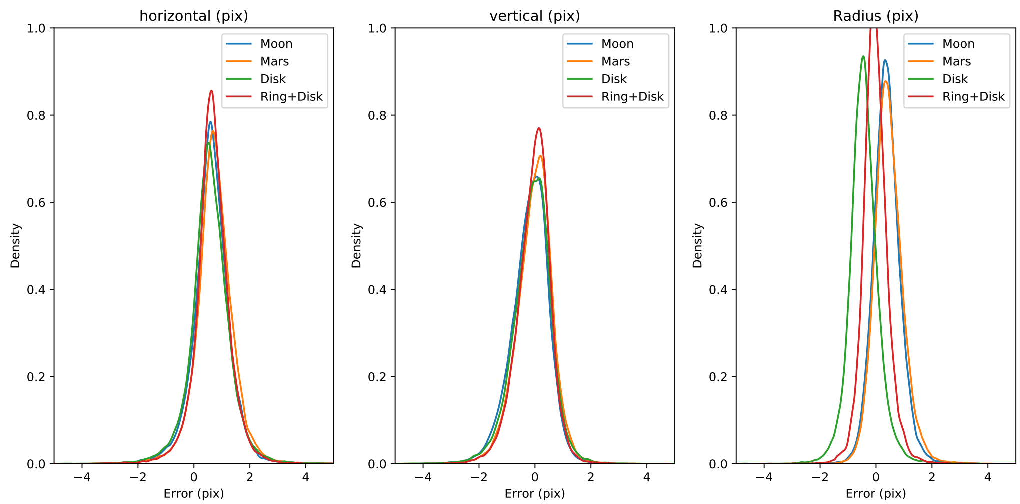

Figure 5: Distribution of pixel level differences between the CDA crater detection and the Robbins and Hynek (2012) dataset. Two additional merged CDA datasets are also included. “Ring+Disk” includes craters found in both the Mars CDA and the Disk CDA. Density is given in units of ”per pixel” and is normalized.

with the median and inter-quartile ranges given in table 4.

	Moon	**Mars**	**Disk**	**Ring+Disk**
Horizontal (longitudinal)	$0.6^{+0.4}_{-0.4}$	$0.7^{+0.4}_{-0.4}$	$0.6^{+0.4}_{-0.4}$	$0.7^{+0.3}_{-0.3}$
Vertical (latitudinal)	$-0.1^{+0.4}_{-0.4}$	$0.03^{+0.4}_{-0.4}$	$-0.06^{+0.5}_{-0.5}$	$0.009^{+0.3}_{-0.4}$
Diameter	$0.4^{+0.3}_{-0.3}$	$0.4^{+0.3}_{-0.3}$	$-0.4^{+0.3}_{-0.3}$	$-0.05^{+0.3}_{-0.2}$

Table 4: Median and inter-quartile ranges for the image level differences between the Robbins and Hynek (2012) crater database and the CDA predictions. All values are given as median and interquartile values of the pixel level differences between the CDA and Robbins and Hynek (2012) data.

The ring trained CDAs and the disk trained CDA have similar accuracy on the location but the opposite sign in the crater diameter differences. This apparent bias may be a result of the method used to generate each prediction, as the disk–finding CNN uses a Sobel and Feldman (1968) transform to convert the predicted disks into rings, and places the ring within the perimeter of the disk, instead of on the outer edge.

This bias can be reduced by combining the results from the ring trained CDA and disk trained CDA are combined such that only craters found by both CDAs are considered detections. This is shown as the "Ring+Disk" result in table 4 and figure 5. The absolute mean difference in diameter between the CDA and Robbins and Hynek (2012) dataset decreases from 0.5 pixels to 0.05 pixels. The trade–off for this improved accuracy is that only craters found by both CDAs can be improved, and the recall of the worst CDA limits the number of craters that can be improved. In this dataset, 63% of the existing craters are found by both CDAs and can be improved with this method.

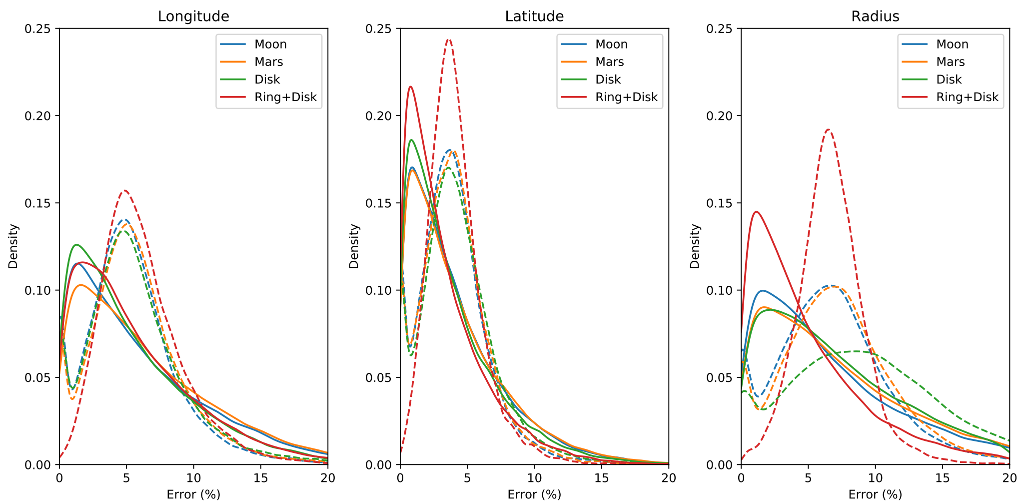

Figure 6: Error density plots for craters found by each CDA with a matching crater in the Robbins and Hynek (2012) dataset, given as the absolute fractional error relative to the crater diameter. (Left) Longitude errors, (center) latitude errors, (right) absolute diameter errors. Per–image statistics are shown with dashed lines, globally aggregated data is shown with solid lines The summary median and inter–quartile ranges for this data is given in table 1.

In terms of geophysical location and size, the distributions of error in the longitude, latitude, and diameter of the matched craters are shown in figure 6, with median and inter–quartile values given in table 1 and 2. After aggregating the per–image metrics to produce the global dataset, the CDA errors decrease as duplicate craters are filtered for higher precision crater location determined using the highest resolution image.

In the global dataset size errors decrease from 6% to 4% (medians) in the combined "Ring+Disk" CDA, but the improvement comes at the expense of recall. In the globally aggregated data, the recall of the combined dataset is worse (at 60%) than the recall of the worst individual CDA (the Disk CDA), while the precision is better (at 80%) than the best CDA (the Mars CDA). The resulting F_1 score drops to 69%, worse than the Mars Ring CDA and better than the Mars Disk CDA.

As a comparison with the errors shown here, Robbins and Hynek (2008) performed a similar study using human–derived datasets from MOLA DTMs and THEMIS imagery and noted that the DTM derived crater sizes are typically 1km larger than the imagery resolved counterparts. In this work the DTM derived crater sizes are 0.05km to 0.92km larger than the Robbins and Hynek (2012) data (25% to 75% percentiles) with the median crater being 0.44km larger. Twenty three percent of the DTM derived craters are smaller than their Robbins and Hynek (2012) counterpart.

4.3. Missed Craters

None of the CDAs found every crater in the Robbins and Hynek (2012) list even if they found more than 57,564 craters in total. The missing craters don't need to share any characteristics but the Robbins and Hynek (2012) dataset includes a large number of parameters that might illustrate why some craters were missed. In particular, the Robbins and Hynek (2012) dataset contains the depths for each crater, including the depth relative to the crater edge (`DEPTH_RIMFLOOR`), relative to the surrounding terrain (`DEPTH_SURFFLOOR`), and the degradation/ preservation state (`DEGRADATION_STATE`) that rates the condition of the crater from highly–degraded (1) to not–degraded (4). A 'random decision forest' algorithm (Tin Kam Ho, 1998) was used to identify these three parameters as most correlated with missing craters in this CDA relative to the Robbins and Hynek (2012).

Comparing the Mars ring CDA with the Robbins and Hynek (2012) dataset (the other CDAs perform similarly), shallow craters are more likely to be missed than deep craters, and highly–degraded craters are more likely to be missed than non–degraded craters. For example craters with a rim-floor depth of 105m or less account for 15% of the dataset, but accounted for 36% of the missed craters. Highly degraded craters made up 45% of all craters but 75% of the missed craters (all other degradation states have a false negative rate of less than 5%). When combined, crater depth is a stronger determinant than the degradation state. In all degradation states, shallower craters were more likely to be missed than deep craters. In the worst case of highly degraded craters, shallower craters are missed at a rate 10 times higher than the deeper craters.

Examples of missed and detected craters are shown in figure 7. Some of the less degraded craters can be found more easily in the THEMIS IR dataset used by Robbins and Hynek (2012) because of the contrasting effect of sunlight on the exposed edges of the crater.

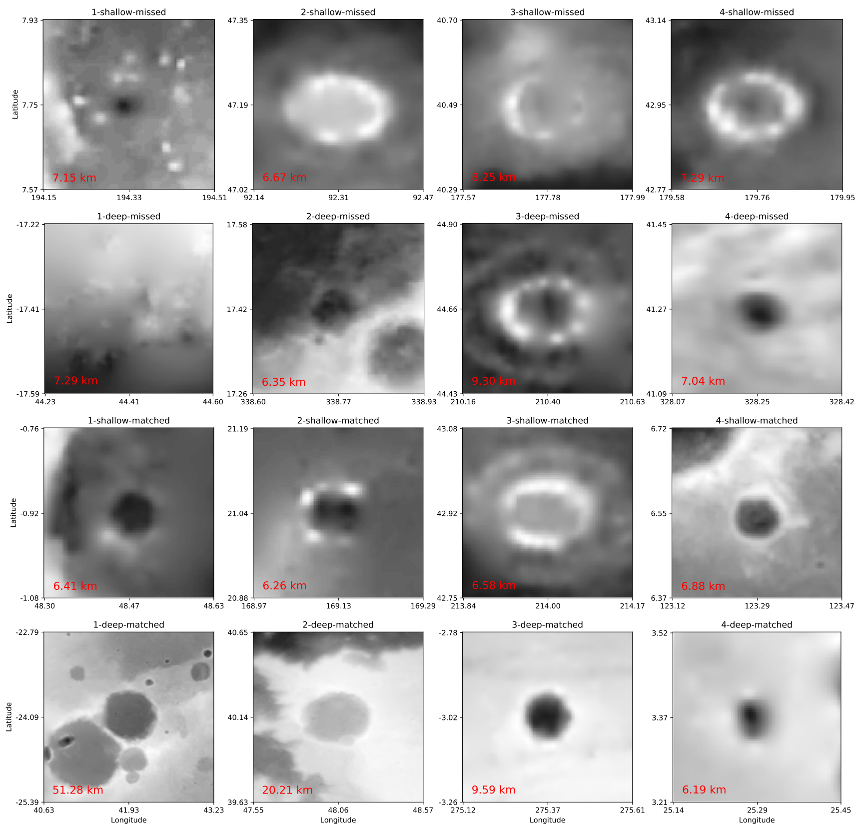

Figure 7: Randomly selected examples of craters from each degradation state (columns) and depth (alternating rows) that were missed (top two rows) or matched (bottom two rows). Each image includes the crater at the center of the image, and a border of 1 crater width on each side. The shades in each image are indicative of local topography in the image, but not necessarily the images presented to the CDA.

Although the impact of the degradation state and crater depth were not known during the training step of this experiment, the different crater types were well represented in the crater populations used in training and validation datasets. If this were not the case, it might have been possible to improve the performance of the CNN on the shallow degraded craters by ensuring a representative sample of these craters in the training dataset.

4.4. False Positives

The CDAs each detected craters that do not exist in the Robbins and Hynek (2012) dataset that are considered false positives. A large fraction of these detections were likely correctly identified as false positives (i.e., the craters do not exist), with a much smaller fraction being real craters missing from the Robbins and Hynek (2012) dataset.

Table 3 gives the number of craters in each CDA and the Robbins and Hynek (2012) dataset, grouped by geologic type (Tanaka et al., 2014). The table also gives the *true positive rate* or the fraction of craters in each CDA that correspond to a known crater. The remaining craters are the *false positives*. The relatively poor performance of the CDAs in the *Apron*, *Basin*, and *Polar* terrain only has a small impact on the overall results — These terrains account for less than 1,500 craters in total.

Examples of false positives in the Mars ring dataset are shown in figure 8, grouped by the crater diameter. Some of the false positive detections have the appearance of craters while others are not obviously circular features (with 10,000+ false positives the small sample shown is random and not necessarily representative). For the larger detected features, many are paterae that are, correctly, not listed in the Robbins and Hynek (2012) *crater* databse. Fifteen of the 20 largest diameter 'false positives' correspond to mountains or paterae, and another 20 of the next largest 80 'false positive' detections are named features on Mars *that are not craters*. The CDA is correct in identifying these circular features in the DTM, but incorrect in labelling them as craters.

For smaller sized features the results are less promising. A review of a random sample of 300 features below 5km in diameter did not identify any definitively new craters — Approximately 30% were depressions related to valleys or topography, but are not craters; 5% were detections of craters with a diameter of 4km in the CDA but below this threshold in the Robbins and Hynek (2012) dataset (and are therefore removed from the dataset) ; 5% of the craters are circular features in the DTM data, but disappear in higher resolution imagery. Most of the remaining 60% are appropriately labelled as false positives and were not crater like even in the available DTM data. Only a small number of samples are possibly new craters, resulting in fewer than 100 new crater detections in the CDA datasets.

Silburt et al. (2019) attempted to answer a similar question by providing a sample of the false positives to researchers to categorize as crater or not. In that case 90% were identified as craters, in stark contrast to the numbers here. However, according to Robbins and Hynek (2012) their database is statistically complete below the lower limit of 4km considered here. For Lunar data, the crater database was less complete and 15% of the new detections by the (Silburt et al., 2019) CDA were below the lower limit of their "ground truth" database. Additionally, the test posed in Silburt et al. (2019) is framed differently, asking whether a human researcher would identify the feature as a crater, rather than asking whether the feature is actually a crater given all the available information.

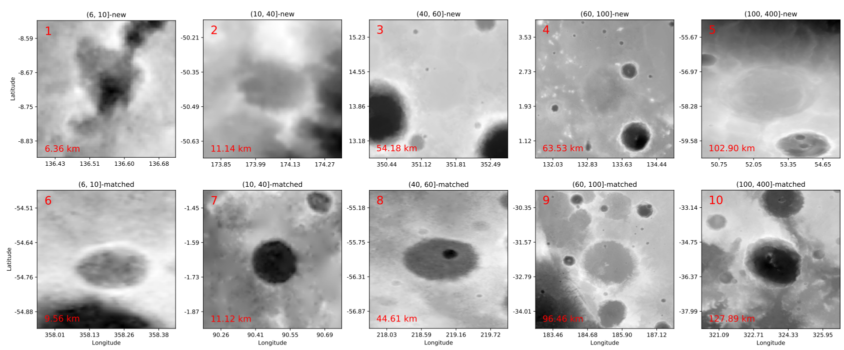

Figure 8: (top row) false positives in the Mars trained dataset, (bottom) true positives in the Mars trained dataset. The feature size increases from left to right (with the diameter range given in the title in kilometres) but the feature is randomly chosen from the CDA dataset. As in figure 7 the identified crater is centered in the image with a 1 diameter border around it. (image 5 is Peneus Patera).

5. Conclusions

In this paper, I have applied a new Crater Detection Algorithm (CDA) to find craters in Mars digital terrain model. The CDA combines a multi–layer neural network to highlight circular features and a template correlation algorithm to determine their location and size. The best CDA used here finds 75% of the craters listed in a comprehensive existing dataset (Robbins and Hynek, 2012), in line with typical human performance on similar datasets (Wetzler et al., 2007). I also showed that a CDA trained on lunar data (Silburt et al., 2019) performed well on the martian DTMs without further training.

The performance of each CDA was measured against the Robbins and Hynek (2012) crater list, and the predicted locations and sizes of craters compare well with that dataset. The CDAs find craters over the entire martian surface with no significant bias in location, size, or geology, and with differences of around 5% of the crater size and location relative to the Robbins and Hynek (2012) dataset, in line with estimated errors from human–generated crater datasets (Robbins et al., 2014).

The best CDA developed here misses many existing craters, and misidentifies other features as craters. The ring trained CDA missed 54% of those craters in the most degraded state, and 80% of those craters shallower than 105m from rim to floor. Given the large number of shallow craters missed, it might be possible to improve the performance of the CNN stage by increasing the 'contrast' of the DTM images by limiting the vertical extent in each image, similar to the pre–processing technique used in Stepinski et al. (2009) to limit the horizontal scale of craters in each image.

A key feature of any automated CDA is the ability to make predictions rapidly and without human intervention. The CDA developed here can work with any standard DTM dataset from planet orbiting spacecraft and can generate 100–1,000 crater predictions per second on consumer hardware. DTMs generated from high resolution imagery can be used to generate catalogs of craters not available in current databases (Lee, 2018), and could be incorporated into existing data processing pipelines. Additionally, this work and others (Silburt et al., 2019; Lee, 2018) have shown that the CDA can be applied across different planets providing consistent datasets are available, allowing meaningful comparison between different planetary bodies using a consistent processing algorithm.

6. Acknowledgments

Computations were performed on clusters and GPU enabled workstations at the University of Toronto. I am grateful to Ari Silburt and co-authors for making the source code for the DeepMoon CDA available. I thank the anonymous reviewer and Christian Riedel for providing numerous comments and questions that have improved and clarified this work.

7. Bibliography

M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng. TensorFlow: A system for large-scale machine learning. In *OSDI'16 Proceedings of the 12th USENIX conference on Operating Systems Design and Implementation*, 2016. ISBN 978-1-931971-33-1. doi: 10.1038/nn.3331. URL `http://arxiv.org/abs/1605.08695`.

J. Akeret, C. Chang, A. Lucchi, and A. Refregier. Radio frequency interference mitigation using deep convolutional neural networks. *Astronomy and Computing*, 18:35–39, 2017. ISSN 22131337. doi: 10.1016/j.ascom.2017.01.002. URL `http://dx.doi.org/10.1016/j.ascom.2017.01.002`.

R. Arvidson, J. Boyce, C. Chapman, M. Cintala, M. Fulchignoni, H. Moore, P. Schultz, L. Soderblom, R. Strom, A. Woronow, and R. Young. Standard techniques for presentation and analysis of crater size-frequency data.

Icarus, 37(2):467–474, 1979. ISSN 10902643. doi: 10.1016/0019-1035(79)90009-5.

R. E. Arvidson. Morphologic classification of Martian craters and some implications. *Icarus*, 22(3):264–271, 1974. ISSN 10902643. doi: 10.1016/0019-1035(74)90176-6.

N. G. Barlow. Crater size-frequency distributions and a revised Martian relative chronology. *Icarus*, 75(2):285–305, 1988. ISSN 10902643. doi: 10.1016/0019-1035(88)90006-1.

N. G. Barlow. A review of Martian impact crater ejecta structures and their implications for target properties. In T. Kenkmann, F. Hörz, and A. Deutsch, editors, *Large Meteorite Impacts III*. Geological Society of America, mar 2005. ISBN 9780813723846. URL `https://doi.org/10.1130/0-8137-2384-1.433`.

N. G. Barlow. What we know about Mars from its impact craters. *Bulletin of the Geological Society of America*, 122(5-6):644–657, 2010. ISSN 00167606. doi: 10.1130/B30182.1.

N. G. Barlow and C. B. Perez. Martian impact crater ejecta morphologies as indicators of the distribution of subsurface volatiles. *Journal of Geophysical Research*, 108(E8):5085, 2003. ISSN 0148-0227. doi: 10.1029/2002JE002036. URL `http://doi.wiley.com/10.1029/2002JE002036`.

F. Chollet. Keras, 2015.

Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger. 3D U-net: Learning dense volumetric segmentation from sparse annotation.

Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9901 LNCS: 424–432, 2016. ISSN 16113349. doi: 10.1007/978-3-319-46723-8_49.

M. J. Cintala, J. W. Head, and T. A. Mutch. Martian crater depth/diameter relationships : Comparison with the Moon and Mercury. In *Proc. Lunar. Sci. Conf. 7th*, pages 3575–3587, 1976.

R. A. Craddock, T. A. Maxwell, and A. D. Howard. Crater morphometry and modification in the Sinus Sabaeus and Margaritifer Sinus regions of Mars. *Journal of Geophysical Research E: Planets*, 102(E6):13321–13340, 1997. ISSN 01480227. doi: 10.1029/97JE01084.

K. Di, W. Li, Z. Yue, Y. Sun, and Y. Liu. A machine learning approach to crater detection from topographic data. *Advances in Space Research*, 54(11):2419–2429, 2014. ISSN 18791948. doi: 10.1016/j.asr.2014.08.018. URL `http://dx.doi.org/10.1016/j.asr.2014.08.018`.

R. Fergason, T. Hare, and J. Laura. HRSC and MOLA Blended Digital Elevation Model at 200m v2, 2018.

I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016. ISBN 0262035618.

R. Jaumann, G. Neukum, T. Behnke, T. C. Duxbury, K. Eichentopf, J. Flohrer, S. v. Gasselt, B. Giese, K. Gwinner, E. Hauber, H. Hoffmann, A. Hoffmeister, U. Köhler, K. D. Matz, T. B. McCord, V. Mertens, J. Oberst, R. Pischel, D. Reiss, E. Ress, T. Roatsch, P. Saiger, F. Scholten, G. Schwarz, K. Stephan, and M. Wählisch. The high-resolution stereo camera (HRSC) experiment on Mars Express: Instrument aspects and experiment conduct from interplanetary cruise through the nominal mission. *Planetary and Space Science*, 55(7-8):928–952, 2007. ISSN 00320633. doi: 10.1016/j.pss.2006.12.003.

C. Lee. Martian Crater Identification Using Deep Learning. In *American Geophysical Union*, 2018.

Met Office. *Cartopy: a cartographic python library with a matplotlib interface.* Exeter, Devon, 2018. URL `http://scitools.org.uk/cartopy`.

M. M. Pedrosa, S. C. de Azevedo, E. A. da Silva, and M. A. Dias. Improved automatic impact crater detection on Mars based on morphological image processing and template matching. *Geomatics, Natural Hazards and Risk*, 8(2):1306–1319, 2017. ISSN 19475713. doi: 10.1080/19475705.2017.1327463. URL `https://doi.org/10.1080/19475705.2017.1327463`.

S. J. Robbins and B. M. Hynek. MOLA data may introduce significant artifacts in crater diameters. In *Mars Crater Consortium 11*, pages 4–5, 2008.

S. J. Robbins and B. M. Hynek. A new global database of Mars impact craters larger than 1km: 1. Database creation, properties, and parameters. *Journal of Geophysical Research E: Planets*, 117(5):1–18, 2012. ISSN 01480227. doi: 10.1029/2011JE003966.

S. J. Robbins and B. M. Hynek. Utility of laser altimeter and stereoscopic terrain models: Application to Martian craters. *Planetary and Space Sci-*

ence, 86:57–65, 2013. ISSN 00320633. doi: 10.1016/j.pss.2013.06.019. URL `http://dx.doi.org/10.1016/j.pss.2013.06.019`.

S. J. Robbins, I. Antonenko, M. R. Kirchoff, C. R. Chapman, C. I. Fassett, R. R. Herrick, K. Singer, M. Zanetti, C. Lehan, D. Huang, and P. L. Gay. The variability of crater identification among expert and community crater analysts. *Icarus*, 234:109–131, 2014. ISSN 00191035. doi: 10.1016/j.icarus.2014.02.022. URL `http://dx.doi.org/10.1016/j.icarus.2014.02.022`.

F. J. Rodionova, K. I. Dekchtyareva, A. Khramchikhin, G. G. Michael, S. V. Ajukov, S. G. Pugacheva, and V. V. Shevchenko. Morphological catalogue of the craters of Mars. Technical report, European Space Agency, 2000.

O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241. Springer International Publishing, Cham, 2015. ISBN 978-3-319-24574-4. doi: 10.1007/978-3-319-24574-4_28. URL `http://arxiv.org/abs/1505.04597http://link.springer.com/10.1007/978-3-319-24574-4{%}5C{_}28`.

G. Salamunićcar and S. Lončarić. GT-57633 catalogue of Martian impact craters developed for evaluation of crater detection algorithms. *Planetary and Space Science*, 56(15):1992–2008, 2008. ISSN 00320633. doi: 10.1016/j.pss.2008.09.010.

G. Salamunićcar, S. Lonarić, P. Pina, L. Bandeira, and J. Saraiva.

MA130301GT catalogue of Martian impact craters and advanced evaluation of crater detection algorithms using diverse topography and image datasets. *Planetary and Space Science*, 59(1):111–131, 2011. ISSN 00320633. doi: 10.1016/j.pss.2010.11.003.

A. Silburt, M. Ali-Dib, C. Zhu, A. Jackson, D. Valencia, Y. Kissin, D. Tamayo, and K. Menou. Lunar crater identification via deep learning. *Icarus*, 317:27–38, 2019. ISSN 10902643. doi: 10.1016/j.icarus.2018.06.022.

D. Smith, M. T. Zuber, and G. A. Neumann. Seasonal variations of snow depth on Mars. *Science*, 294:2141–2146, 2001. URL `http://www.sciencemag.org/content/294/5549/2141.short`.

I. Sobel and G. Feldman. A 3x3 Isotropic Gradient Operator for image Processing. In *Stanford Artificial Intelligence Project*, 1968.

L. A. Soderblom, C. D. Condit, R. A. West, B. M. Herman, and T. J. Kreidler. Martian planetwide crater distributions: Implications for geologic history and surface processes. *Icarus*, 22(3):239–263, 1974. ISSN 10902643. doi: 10.1016/0019-1035(74)90175-4.

T. F. Stepinski and E. R. Urbach. The First Automatic Survey of Impact Craters on Mars: Global Maps of Depth/ Diameter Ratio. In *Lunar and Planetary Science Conference*, 2009. ISBN 1875-5550 (Electronic) 1389-2037 (Linking). doi: 10.2174/138920312803582960.

T. F. Stepinski, M. P. Mendenhall, and B. D. Bue. Machine cataloging of impact craters on Mars. *Icarus*, 203(1):77–87, 2009. ISSN 00191035.

doi: 10.1016/j.icarus.2009.04.026. URL http://dx.doi.org/10.1016/j.icarus.2009.04.026.

K. L. Tanaka, J. A. Skinner, J. M. Dohm, R. P. Irwin, E. J. Kolb, C. M. Fortezzo, T. Platz, G. G. Michael, and T. M. Hare. Geologic map of Mars: USGS Sci. Invest. Map 3292, scale 1:20,000,000. *U.S. Geological Survey, Reston, Va.*, 2014. doi: 10.3133/sim3292.

Tin Kam Ho. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8): 832–844, 1998. ISSN 01628828. doi: 10.1109/34.709601. URL http://ieeexplore.ieee.org/document/709601/.

S. van der Walt, J. L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J. D. Warner, N. Yager, E. Gouillart, and T. Yu. scikit-image: image processing in Python. *PeerJ*, 2:e453, 2014. ISSN 2167-8359. doi: 10.7717/peerj.453. URL https://peerj.com/articles/453.

P. G. Wetzler, R. Honda, B. Enke, W. J. Merline, C. R. Chapman, and M. C. Burl. Learning to detect small impact craters. *Proceedings - Seventh IEEE Workshop on Applications of Computer Vision, WACV 2005*, pages 178–184, 2007. doi: 10.1109/ACVMOT.2005.68.

D. U. Wise and G. Minksowski. Dating methodology of small, homogeneous crater populations applied to the Tempe-Utopia Trough region on Mars. Technical report, NASA, 1980.

I. H. Witten and E. Frank. *Data Mining: Practical machine learning tools and techniques.* Morgan Kaufmann, 2005.