Automated crater shape retrieval using weakly-supervised deep learning

Mohamad Ali-Dib[a,b,e], Kristen Menou[c,d], Chenchong Zhu[c,e], Noah Hammond[b], Alan P. Jackson[b,f]

[a]*Institut de recherche sur les exoplanètes, Département de physique, Université de Montréal. 2900 boul. Édouard-Montpetit, Montréal, Quebec, H3T 1J4, Canada*
[b]*Centre for Planetary Sciences, Department of Physical & Environmental Sciences, University of Toronto Scarborough, Toronto, Ontario M1C 1A4, Canada*
[c]*Department of Astronomy & Astrophysics, University of Toronto, Toronto, Ontario M5S 3H4, Canada*
[d]*Physics & Astrophysics Group, Dept. of Physical & Environmental Sciences, University of Toronto Scarborough, 1265 Military Trail, Toronto, Ontario, M1C 1A4, Canada*
[e]*Canadian Institute for Theoretical Astrophysics, University of Toronto, 60 St. George St, Toronto, Ontario M5S 3H8, Canada*
[f]*School of Earth and Space Exploration, Arizona State University, 781 E. Terrace Mall, Tempe, AZ 85287, USA*

Corresponding authors:

- M.A-D. malidib@astro.umontreal.ca

Abstract

Crater shape determination is a complex and time consuming task that so far has evaded automation. We train a state of the art computer vision algorithm to identify craters on the moon and retrieve their sizes and shapes. The computational backbone of the model is MaskRCNN, an "instance segmentation" general framework that detects craters in an image while simultaneously producing a mask for each crater that traces its outer rim. Our post-processing pipeline then finds the closest fitting ellipse to these masks, allowing us to retrieve the crater ellipticities. Our model is able to correctly identify 87% of known craters in the holdout set, while predicting thousands of additional craters not present in our training data. Manual validation of a subset of these craters indicates that a majority of them are real, which we take as an indicator of the strength of our model in learning to identify craters, despite incomplete training data. The crater size, ellipticity, and depth distributions predicted by our model are consistent with human-generated results. The model allows us to perform a large scale search for differences in crater diameter and shape distributions between the lunar highlands and maria, and we exclude any such differences with a high statistical significance.

1. Introduction

Craters are one of the dominant morphological structures on most solar system objects. Their numbers can be used as a diagnostic tool to estimate

2

the relative surface age of airless objects, while their size distributions hold valuable information on the impactors. Finding new craters and retrieving their sizes has however generally been a manual process, and as such is rather time consuming, especially for smaller radii where the sharply increasing number of craters makes the compilation of global databases difficult. The impressive database of Martian craters published by Robbins and Hynek (2012) for example took years to compile. An automated method has been demonstrated recently by Silburt et al. (2019) who trained a convolutional neural network (CNN) in a UNET architecture ("DeepMoon") to identify craters on the surface of the moon with a recall on known craters of 92%, and a false positive rate on newly identified ones of 11±7% (see also the techniques of Harris (2018); Christoff et al. (2018)).

Characterizing the shape and ellipticity distributions of craters on the other hand is another hard task that is yet to be decisively automated. Crater ellipticity is controlled mainly by the impact angle (Gault and Wedekind, 1978; Melosh, 1989), and is hence useful in order to understand the obliquity history of bodies (Holo et al., 2018), the surface's geophysical properties (Elbeshausen et al., 2009, 2013), and improve age dating (Liu et al., 2018). Historically, crater ellipticity was determined through slow and inefficient visual inspection (Schultz and Lutz-Garihan, 1982; Barlow, 1988; Bottke et al., 2000; Herrick and Forsberg-Taylor, 2003), and it is only very recently that numerical methods have started to emerge (Liu et al., 2018).

In this work we present a new method based on MaskRCNN, a neural

network capable of identifying craters with accuracy comparable to Deep-Moon while simultaneously retrieving their shapes accurately. This allows for a large sample study of lunar crater ellipticities, and the first systematic search of any dichotomies between craters of the highlands and of the mare.

We discuss our data sources, training set generation algorithm, methods and neural network architecture in section 2, present our results on crater size distribution, ellipticity, and depth in section 3, and finally conclude in section 4.

2. Methods

2.1. Data Preparation

As with any machine learning model, MaskRCNN needs to be trained on a significant amount of data in order to tune the network's parameters, including its internal parameters such as neuron weights and biases, and external hyperparameters such as number of training epochs and regularization options. The training data format however needs to be consistent with the neural network's architecture. MaskRCNN works by searching for instances of specific objects (craters) in an image, and returns the bounding boxes containing the object in addition to the associated object masks. This implies that, for each training example, the input needs to be an image of the surface of the moon containing multiple craters, while the training target is a series of independent masks individually covering the crater ar-

4

eas. This data format allows the network to learn to map a lunar surface image into crater masks containing all of the relevant physical information. In practice however, using actual optical images introduces the complex problem of shadows, and so we follow Silburt et al. (2019) in using digital elevation maps (DEMs) instead. This training data format is analogous to the one used by DeepMoon, and hence we use a slightly modified version of the data generation pipeline of Silburt et al. (2019), which we refer readers to for further details. This method generates training data by randomly cropping "poststamp" images representing local areas from a global digital elevation map of the Lunar Reconnaisance Orbiter (LRO) and Kaguya merged digital elevation model (LOLA Team and Kaguya Team, 2015). This has a resolution of 512 pixels/degree, or 59 meters/pixel.

The data used to construct the targets was obtained by merging the global crater dataset assembled by Povilaitis et al. (2018) using the LRO Wide Angle Camera (WAC) Global Lunar DEM, and, for diameters larger than 20 km, the global crater dataset assembled by Head et al. (2010) using the LOLA DEM. Our training and validation sets span the region of the moon between -60 and +180 degrees longitude, over the entire latitude range we are considering (-60 to +60 degrees), while the test set covers the remaining area (where large lunar mare are concentrated). We note that during the final phases of writing this manuscript a new global dataset of craters was published by Robbins (2019) that includes craters with diameters down to 1-2 km. We leave exploring this dataset for future works.

While we largely utilise the same data processing pipeline as Silburt et al. (2019) our usage here has the following simple but crucial differences:

- We downsample the cropped images to $512 \times 512\,\text{pixels}$ instead of $256 \times 256\,\text{pixels}$, as this allows for higher resolution instance masks, leading to a better shape retrieval at the expense of increased memory usage.

- Since our ultimate goal is to retrieve the craters shapes, we represent the target craters as binary full masks (disks) instead of binary rings. This is to preserve more informations on the shapes and facilitate ellipse fitting of the predicted masks in the postprocessing phase, and allows us to be consistent with MaskRCNN full object masks.

- Instead of the target being one image containing multiple crater binary rings, each crater now is contained within its own separate target image containing a binary mask. This allows for the detection of overlapping craters that would have partially (or completely) disappeared if their binary masks were present in the same image.

- We eliminated the padding along the edges of images, as we found these to significantly affect ellipticity retrievals. This is done by transforming a region about twice the width and height of the desired $512 \times 512\,\text{pixels}$ to an orthographic projection, then cropping the central $512 \times 512\,\text{pixels}$ to obtain a padding-free image. This means that

each input image contains, to varying degrees, a larger portion of lunar surface than those used by Silburt et al. (2019).

- In the test set, we moreover separate highlands and mare craters.

2.2. MaskRCNN

The central algorithm we use to identify craters and retrieve their shapes is MaskRCNN [1] (Mask Region Convolutional Neural Network, He et al. (2017)). It is a general framework for object instance segmentation that detects multiple objects in an image while simultaneously generating a segmentation mask for each instance. MaskRCNN extends the concept of Region-CNNs (Girshick et al., 2013) by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition.

In practice we use the Matterport implementation of MaskRCNN, built with `Keras` on top of a `Tensorflow` backend. We modified the default MaskRCNN setup to provide additional augmentation at training time: all images to the network are randomly flipped vertically and horizontally, and they may additionally be rotated by 90, 180 or 270 degrees. Additional model configuration details can be found in Appendix A.

Our MaskRCNN is trained on a dataset of 1980 individual poststamp DEM images plus their masks. An additional 70 images and masks are

---

[1] https://github.com/facebookresearch/Detectron and https://github.com/matterport/Mask_RCNN

used as a validation set. All model layers are trained for 80 epochs, using the standard Stochastic Gradient Descent optimizer with a learning rate of 10^{-3}. The model with the best validation performance on the standard MS COCO[2] mAP50 metric (mean Average Precision at an intersection over union, or Jaccard Index, of 0.5) is kept for predictions on the test set. In essence this is a maximization of mask overlap that also penalizes against false detections.

Model weights are initialized with MS COCO pretrained values, so our training procedure can be seen as effectively fine-tuning a pretrained MaskR-CNN to our specific dataset.

Grey DEM images are fed to MaskRCNN as identically valued in each of the three RGB channels of the model. For better contrast, the images are pre-processed with the Contrast Limited Adaptive Histogram Equalization (CLAHE) algorithm, as implemented in the `OpenCV` library. In practice, we apply this contrast enhancement only on lightness (L-band) after an RGB to LAB color space transformation (see Appendix B for the detailed pre-processing algorithm). Note that prior to any other processing the source DEM is downsampled from 16-bit greyscale to 8-bit greyscale, as in the DeepMoon pipeline.

---

[2]http://cocodataset.org

2.3. Post-Processing

After visually inspecting our test set predictions, we found several (but rare) instances where the machine failed entirely, predicting a mask that covers almost the entire poststamp image. This is likely due to the complexity of the machine learning task of predicting the same object at very different size scales. We hence filter all masks whose pixel area is bigger than 1/3 of a 512x512 image. This corresponds to roughly 0.3% of the total unique craters.

While MaskRCNN predicts the position and shape of a crater through respectively the bounding box and mask it returns, we still need to retrieve the physical diameter and ellipticity from the boxed mask. Calculating the diameter is a straightforward task, as to first order it suffices to define the diameter (in pixels) as simply the length of the mask's bounding box, and then (knowing the poststamp image's location on the moon) convert this quantity to real diameter in kilometers. Any errors introduced by this method are small compared to the uncertainties we have on the measured crater diameters. Therefore, we are virtually getting the crater diameters for "free", compared to DeepMoon where a separate pipeline had to be implemented in order to calculate these values with a brute-force fitting algorithm. Note that we ignore any errors on the diameters introduced by the ellipticity of the craters resulting in some bounding boxes deviating from exact squares. This is justified by the relatively low number of highly elliptical craters.

Retrieving the ellipticity on the other hand is formally a computer vision task consisting of fitting the near-circular mask image to an ellipse template and then measuring its semi-major and semi-minor axes. We do this using using OpenCV's "fitEllipse" module (Bradski, 2000) implemented using the direct least-square fitting method of Pilu et al. (1996). This module takes the crater mask as input, then returns the center of the fitted ellipse, its major and minor axes, and an orientation angle. We finally define the ellipticity (ϵ) of the crater as the ratio of its major to minor axis.

We emphasis that we do not have a ground truth for crater ellipticity, the ground truth masks on which MaskRCNN is trained are all perfectly circular. Therefore the machine learns the non-circular shape deviations as a byproduct of detecting craters with various shapes, something referred to as "weak-supervision". In other words, we do not directly teach the machine to identify shapes, it learns them as part of its primary detection-focused task.

The results of section 3.3 should thus be taken with this in mind and viewed more as a guideline of what can be achieved by machine learning algorithms. The comparison between ellipticity data obtained by MaskR-CNN and literature studies also offers a test on how well the algorithm is learning to identify the features that define the shape of a crater.

Note that the number of craters with significant ellipticity is expected to be small so this is unlikely to have a noticeable effect on the basic task of crater identification.

After we determine crater properties for each poststamp image, additional calculations are required to derive a global crater catalog from them. Since we randomly crop poststamp images from a single global DEM to generate our dataset, the same crater can (and usually does) appear in multiple images. To generate accurate global crater size and shape distributions, we need to filter these duplicate craters. We hence use the same method as Silburt et al. (2019) (section 2.5), where we classify craters as duplicates if their longitudes $\mathcal{L}$, latitudes L, and radii R overlap within a certain tolerance factor:

$$\frac{\left((\mathcal{L}_i - \mathcal{L}_j)^2 \cos^2\left(\frac{\pi}{180°}\langle L \rangle\right) + (L_i - L_j)^2\right)}{C_{KD}^2 \min(R_i, R_j)^2} < D_{\mathcal{L},L} \tag{1}$$

$$\frac{\text{abs}(R_i - R_j)}{\min(R_i, R_j)} < D_R \tag{2}$$

with $C_{KD} = \frac{180°}{\pi R_{\text{Moon}}}$, $\langle L \rangle = \frac{1}{2}(L_i + L_j)$, $D_{\mathcal{L},L}$=2.08 and D$_R$=1.44. We checked the validity of this approach by creating a set of 300 randomly chosen pairs of craters classified as duplicates by the algorithm, then visually inspecting them. We estimated the false-duplicate error rate to be around $\sim 4\%$, with embedded craters being its main source.

11

3. Results

3.1. Crater Identification on the Moon

The first and most basic task our model needs to do is crater identification: in any given DEM, find all existing craters. The performance of such machines is measured via precision and recall, defined respectively as:

$$P = \frac{T_p}{T_P + F_p} \tag{3}$$

$$R = \frac{T_p}{T_P + F_n} \tag{4}$$

where T_p is the number of true positive craters, F_p are false positives and F_n are the false negatives. High precision implies that most craters that were found by the model are actually present in the ground truth, while high recall implies that the model was able to find most craters in the ground truth list (independent of false positive rate).

In table 3.1 we compare the precision and recall to DeepMoon[3]. We follow DeepMoon in calculating two sets of recall/precision: Post-CNN where the values are separately calculated for each of the individual poststamp images before being averaged over the entire test set; and Post-processing where they are calculated from the global unique crater distribution. Post-CNN recall allows us to quantify the performance of the model on individ-

---

[3]Note that DeepMoon's test set extended from +60 to +180 degrees longitude, while in this work the test set is between -60 and -180 degrees longitude.

ual small scale images (how many known craters per image MaskRCNN is finding), while Post-processing recall gives us the overall performance of the model on the entire test set (how many of all known test set craters it is finding). Note that by transforming from a set of individual images to a global catalog we are essentially doing significant ensembling, giving the same crater multiple chances of being detected in at least one of the images.

For post-CNN, we find that MaskRCNN has a 20% higher recall than DeepMoon (87.6% compared to 54%), with a 12% lower variance. This implies that MaskRCNN is significantly more robust in detecting craters in any given image, successfully identifying on average 87.6% of all known craters. After post-processing however, MaskRCNN has found in total 85.1% of all craters in the test set, while DeepMoon has found 92%. While this difference can be partially attributed to the different test set longitudes used by the two models, Fig. 4 shows clearly that DeepMoon is better at detecting intermediate sized craters than MaskRCNN, probably leading to the better overall recall.

The post-processing precision of MaskRCNN is 40.2%, compared to 56% for DeepMoon. However, as Silburt et al. (2019) argued, the low precision score for these models is due to the machines identifying a large number of new craters not present in the GT dataset. This is not surprising since the crater identification methodology of Povilaitis et al. (2018) was conservative, including only surface features definitively identified as craters.

MaskRCNN's lower precision can hence be interpreted as the machine identifying more new craters with respect to the number of ground truth craters than DeepMoon.

As an ultimate test for the performance of our model, we calculated the false positive rate of "new" craters identified but not present in the ground truth catalogues. This was done in the same fashion as Silburt et al. (2019) where a group of 3 astronomers were shown a representative random sample of data (30 poststamp images containing a total of 349 uncatalogued craters), and tasked with counting what they subjectively deemed as false detections. We then calculated the arithmetic mean of the 3 estimates to finally get a false positive rate of 24%.

We emphasise that the 'new' craters identified by the machine are not necessarily (and in most cases are unlikely to be) new to science and may well have been examined by previous studies, especially regional ones. The key point here is that they are not present in the ground truth data set and thus represent the machine applying the patterns it has learned from the craters contained in the ground truth and identifying additional surface features that satisfy them.

Finally, alongside the recall and precision that tell us how good the machine is at identifying the presence of a crater, we also wish to examine how well the machine reproduces the dimensions and location of the crater. Averaging over all detected craters that can be identified with a target crater in the ground truth we find that the fractional error in the centroid location

of the crater has a standard deviation of 10.5% of the crater radius in longitude and 7.5% of the crater radius in latitude. The difference between the longitude and latitude errors is likely due to the effects of the projection. For the radius we find that the standard deviation over the whole dataset is 7%. In addition, to test for any size or location dependent biases we show in Figure 2 the fractional error in the radius as a function of latitude and in Figure 3 the radius as predicted by the machine versus the radius in the ground truth. Figure 2 shows that there is no correlation between the fractional error in the radius and the latitude, reassuring us that in the $\pm 60°$ latitude range we examine there are no residual effects from the projection correction. The larger number of points at positive latitudes is simply due to a higher abundance of craters in the Northern hemisphere. From Figure 3 we can see that our model does have a slight tendency to underpredict the crater radius for craters larger than around 20 km, and that the dispersion is somewhat larger for these larger craters

3.2. Crater size distribution

The major success of DeepMoon was the accurate retrieval of the test set's crater size distribution. In this section we analyze the performance of MaskRCNN on this fundamental task. In Fig. 4 we show the crater size distribution predicted by MaskRCNN presented as a cumulative size-frequency distribution (NASA Technical Memorandum (79,730, 1978)). We show the separate distributions of the lunar highlands and mare, compared to the
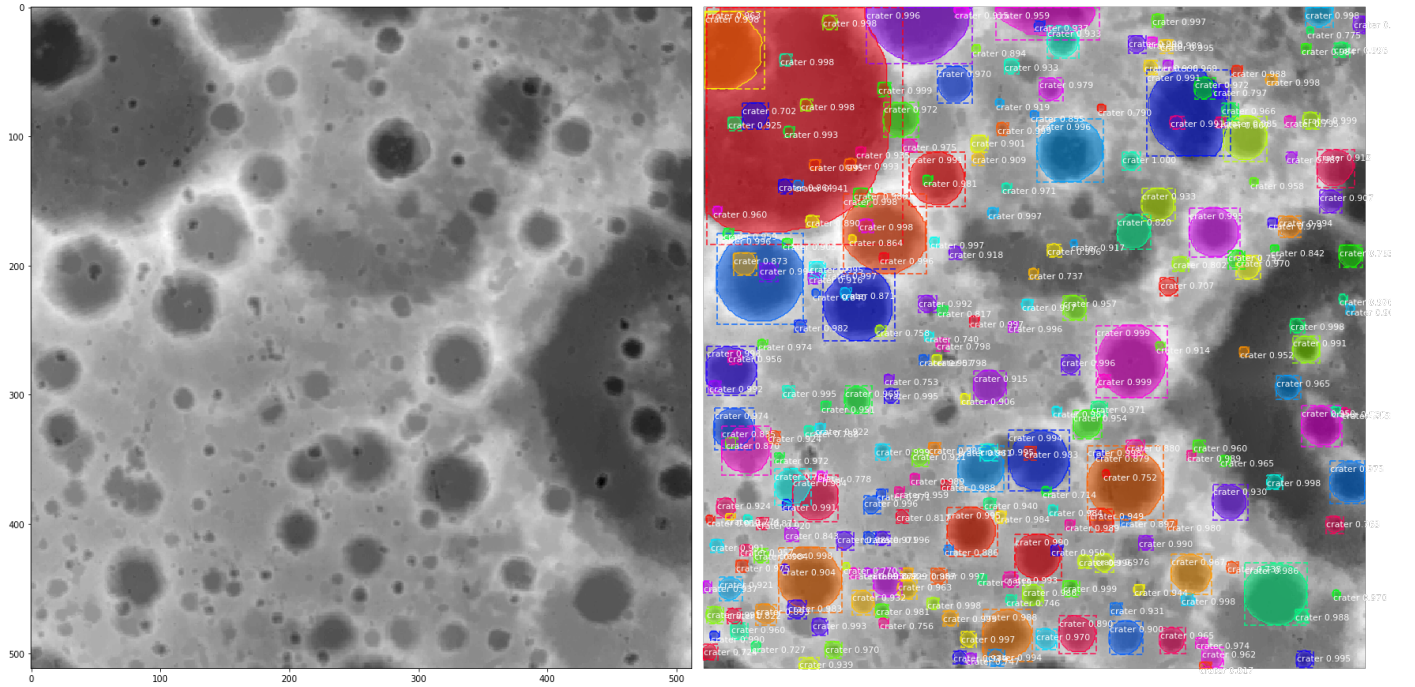
15

Figure 1: Poststamp (DEM) image in our test set showing the identified craters bounded by the detection boxes, and their associated shape masks. The numbers shown are the detection certainty. This image shows that our algorithm successfully identified tens of craters across a large size spectrum, including overlapping and embedded cases. It also visibly retrieved deviations from circularity in its prediction masks.

Accuracy Metric	MaskRCNN (Post-CNN)	MaskRCNN (Post-proc)	DeepMoon (Post-CNN)	DeepMoon (Post-proc)
Recall	$87.6 \pm 8\%$	85.1%	$57 \pm 20\%$	92%
Precision	$66.5 \pm 17\%$	40.2%	$80 \pm 15\%$	56%
F1	0.75	0.54	0.66	0.69
Frac. long. error	10.5%	-	10%	-
Frac. lat. error	7.5%	-	10%	-
Frac. rad. error	7%	-	8%	-

Table 1: Accuracy metrics table comparing the test set precision, recall, F1 score, and fractional error on craters coordinates of MaskRCNN to DeepMoon. Post-CNN is the average of the values calculated for each poststamp image, and is hence biased by craters present in more than one image. Post-processing is the value calculated over the entire test set once unique craters have been extracted, and their physical dimensions retrieved.
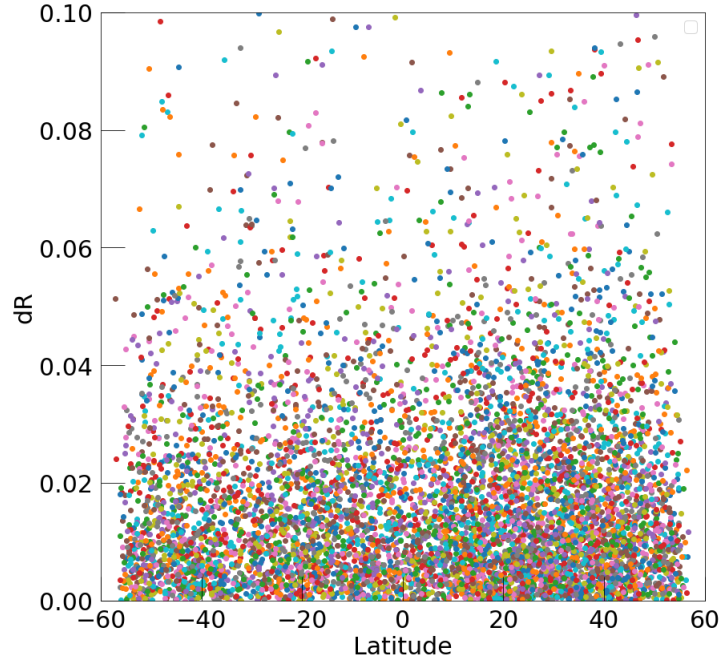

Figure 2: The fractional radius error between the ground truth and predictions defined as $abs(R_G - R_P)/\bar{R}$, as a function of the GT craters latitudes. Colors are for display purposes only. The uniformity of the scatter shows that the detection precision is uncorrelated with latitude.
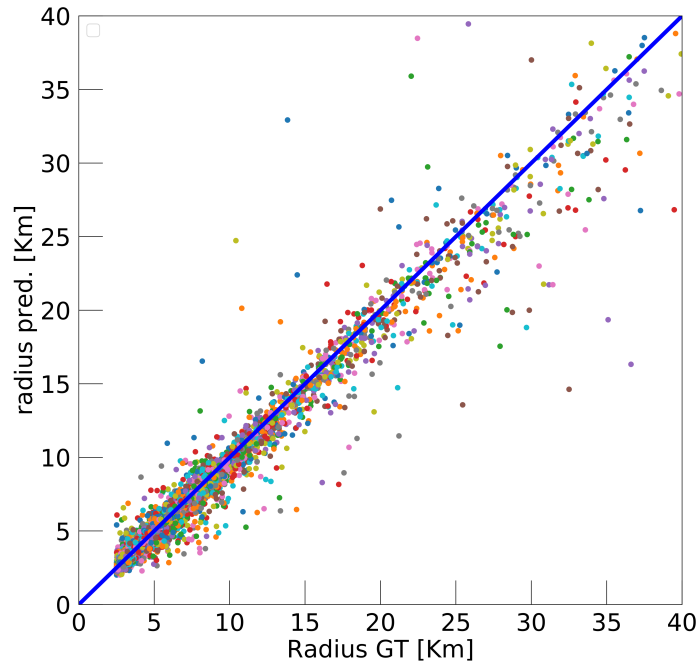
Figure 3: The predicted craters radii as a function of the GT radii. Colors are for display purposes only. Scatter points distribution is symmetric with respect to the blue diagonal for craters smaller than ~ 20 Km, implying that the fractional radius error is independent of the craters size in this range.

ground truth from our test set. The predicted distribution closely follows the GT for craters larger than 30 km, while increasing more steeply for smaller diameters. This is due to MaskRCNN "discovering" a large number of smaller craters not present in the GT. One of the shortcomings of Deep-Moon was the decrease in detection performance for craters larger than 20 km. Comparing MaskRCNN and DeepMoon curves in Fig. 4 we can see that although there is a slight drop-off in the performance of MaskRCNN at larger crater diameters it is a marked improvement over DeepMoon for these sizes. MaskRCNN however detects less intermediate sized craters with diameters between 15 and 40 km than DeepMoon (even though it is still finding more craters than in the ground truth), while detecting more for diameters smaller than 15 km.

This highlight the importance of models ensembling in machine learning, since different models can have better performance in different regimes. Note that the distributions inferred by both DeepMoon and MaskRCNN roll over for craters smaller than 5 Km, which is a sign of incompleteness at these scales.

In Fig. 4 we can also infer that the cumulative size distributions of highland and maria craters are very similar except at very large sizes where the maria distribution suffers from low number statistics, and very small sizes where the distributions plateau due to resolution limitations. This is demonstrated further with a 2-sample Kolmogorov-Smirnov test for craters larger than 5 km, returning a p-value of 0.30.
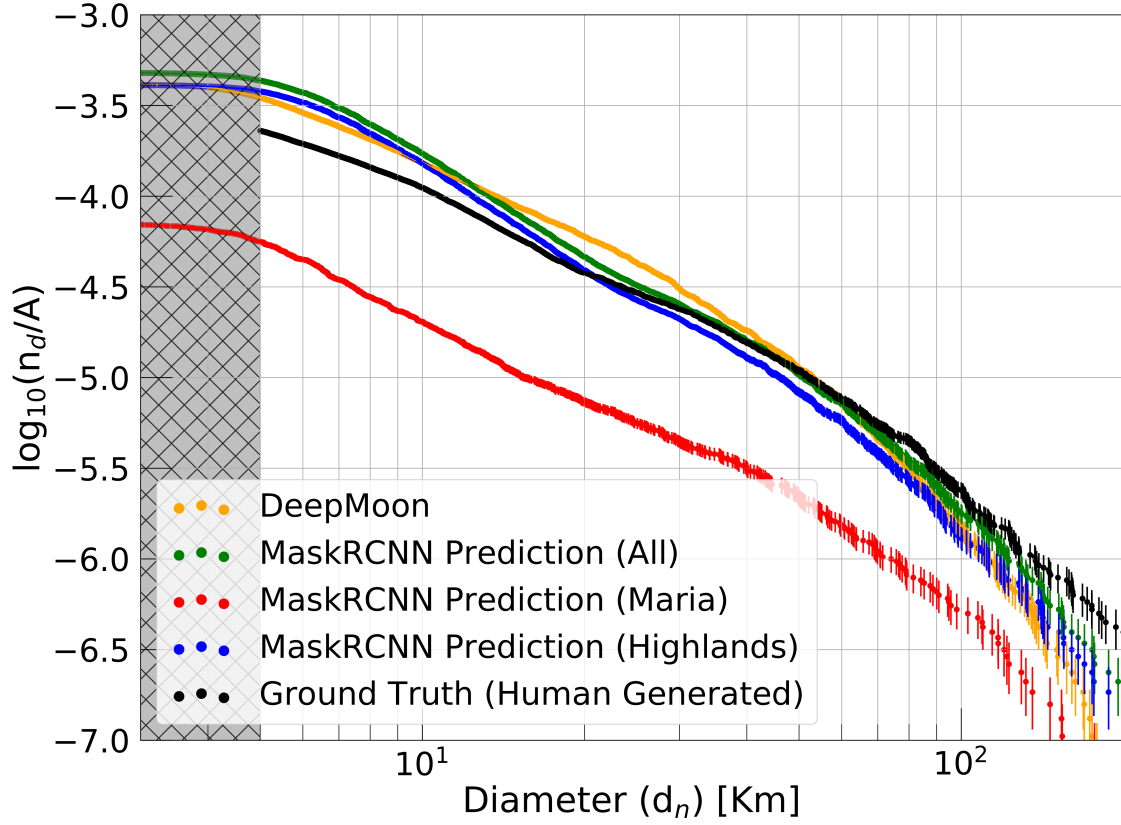
Figure 4: Lunar crater size-frequency distributions represented as CSFD plots, for our test set. Black is the human-generated ground truth, red and blue are MaskRCNN's predictions for the maria and highlands craters, respectively, and green is both simultaneously. The model predictions follow the GT closely for intermediate and large craters, before increasing more steeply for small craters where MaskRCNN is detecting new craters not present in the GT. Note that DeepMoon's distribution was generated for a different longitude range of the Moon.

3.3. Lunar Crater Ellipticity Distribution

Not all craters are exactly circular, impacts at a non-normal angle can result in elliptical craters, elongated along the direction of travel of the impactor (e.g. Gault and Wedekind, 1978). The expected distribution of impact angles is well known (Shoemaker, 1962), and peaks at an angle of $45°$, diminishing towards both perfectly vertical and perfectly grazing. Given a large set of impact craters however it is expected that a fraction of them will occur at shallow enough angles to generate significant ellipticity. The well defined distribution of expected impact angles allows the ellipticity distribution to be mapped onto the angle distribution to find the impact angle that results in a certain ellipticity. The threshold angle at which a given ellipticity is achieved varies between different surfaces and depends on the strength of the target material (e.g. Collins et al., 2011). Stronger materials more readily retain information about impact orientation and have a larger threshold angle for a given ellipticity. The ellipticity distribution on the surface of a body is thus of interest since it provides some access to information about the strength of the surface materials.

Ellipse fitting is easier with the mask approach that we use here compared with the binary rings used by Silburt et al. (2019) and so here we examine the crater shape distribution found by MaskRCNN. In Fig. 5 we show the ellipticity (ϵ) frequency distribution of craters in the test set. An issue for ellipse fitting with small craters is the quantisation of the crater dimensions into the image pixels. For example a crater that is 7 pixels along
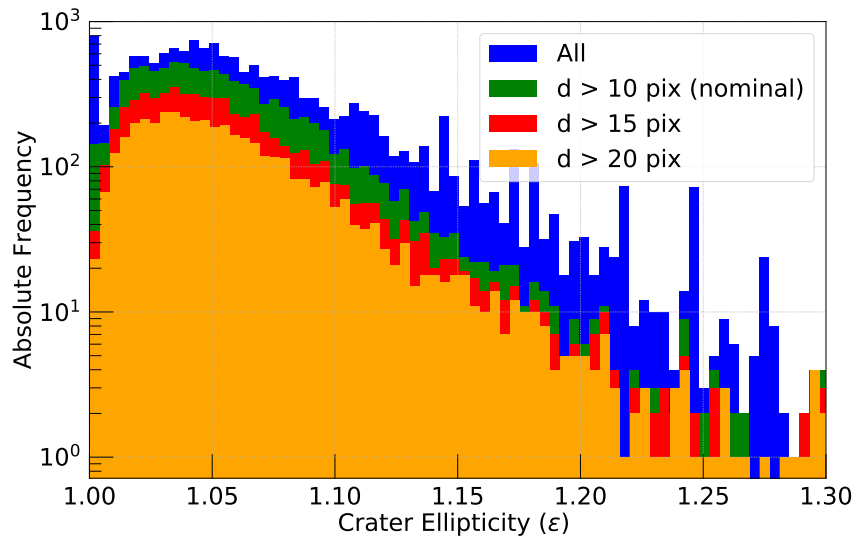
21

Figure 5: Crater ellipticity frequency distribution as predicted by our model for the test set, for different levels of minimum craters diameters in pixels. Notice the spikes when all craters are considered, due to inaccurate ellipse fitting of the smallest craters resolved with a small number of pixels. These disappear when we exclude craters smaller than 10 pixel in diameters.

the longest dimension can be 7 by 7 ($\epsilon = 1$), 6 by 7 ($\epsilon = 1.6$), etc, but not values in between. To deal with this issue we introduce a cut-off in the on-image uncalibrated diameter (in pixels) below which we consider a crater to be too small for accurate ellipse fitting and exclude it from the distribution. The blue histogram in Fig. 5 shows all of the craters with no cut-off and has distinct spikes at $\epsilon = 1$ and multiple values beyond around $\epsilon = 1.12$. The green, red and orange histograms then show the distribution with successively larger cut-off sizes of 10, 15 and 20 pixels. We can see that even the 10 pixel cut-off is sufficient to eliminate almost all of the spikes and the shape of the distribution is very similar for all three values. As such for the rest of this section we use a cut-off size of 10 pixels.

Compared to standard crater counts (that assume circularity) studies of elliptical craters are somewhat sparser. The earliest systematic observational studies specifically of elliptical craters focussed on Mars (e.g. Barlow, 1988), while early lunar studies examined unusual or anomalous craters more generally rather than specifically elliptical craters (e.g. Schultz and Lutz-Garihan, 1982). Surprisingly, the first study that explicitly examines elliptical lunar craters is probably that of Bottke et al. (2000). In that work, the authors surveyed the lunar maria for elliptical craters using Lunar Orbiter IV images. They measured a total of 932 craters between 2.3 and 89 km in diameter, and concluded that 50 of those ($\sim 5.4\%$) have an ellipticity higher than 1.2. This compares well with our data for which we find that $\sim 3\%$ (494 out of 16,664 craters) have $\epsilon > 1.2$.
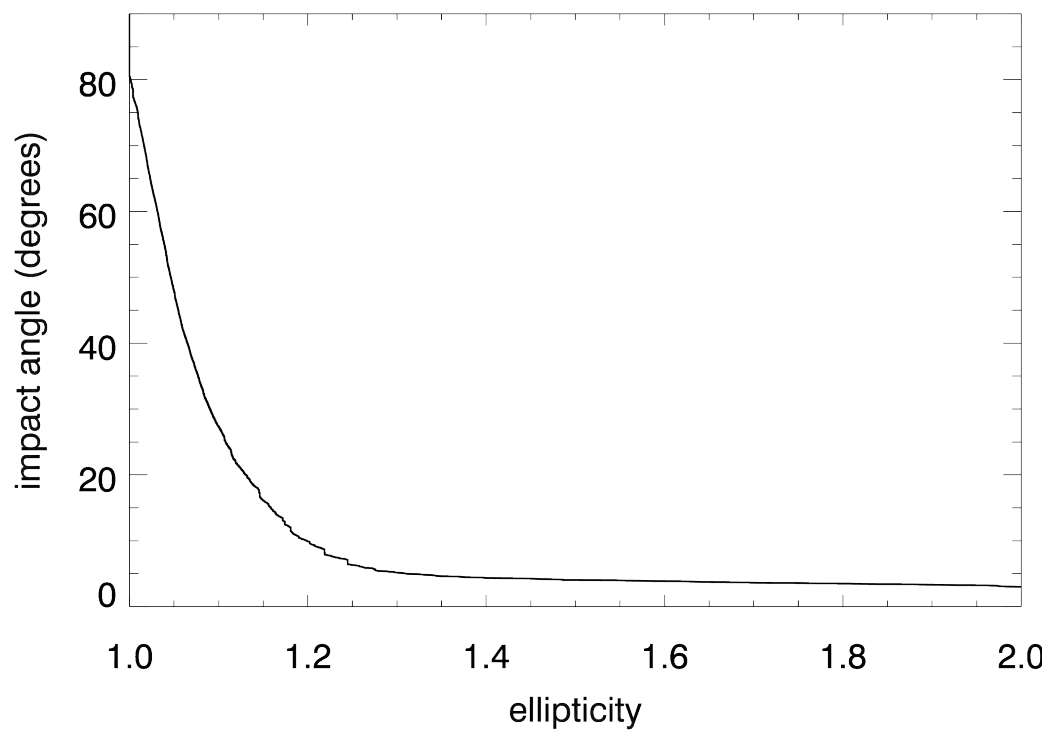
Figure 6: Ellipticity-impact angle distribution for the 16,664 craters in the test set, constructed assuming that the impact angle follows a standard $P(\theta)d\theta = 2\sin 2\theta d\theta$ distribution.

Previous studies like that of Bottke et al. (2000) have focussed on the proportion of craters that are more elliptical than a certain threshold, typically $\epsilon = 1.1$ or 1.2. This is partly because conducting an search for elliptical craters, fitting ellipses in a manual or only semi-automated fashion, is much more time-consuming than simple circular searches, especially for craters that are fairly close to circular. Concentrating on craters that are unambiguously elliptical dramatically reduces the number of craters that the human counter needs to examine. Our fully automated approach however allows us to examine many more craters, including the full distribution from circular to highly elliptical as illustrated in Fig. 5. In addition to converting the fraction with an ellipticity above a certain value into a threshold angle this allows us to construct a complete ellipticity-angle distribution as shown in Fig. 6. From this we can see that impacts at less than $27°$ result in $\epsilon > 1.1$, while impacts at less than $10°$ result in $\epsilon > 1.2$. There are around 6 times as many craters with ellipticities in the range 1.1-1.2 than there are with $\epsilon > 1.2$. This $27°$ threshold angle for $\epsilon > 1.1$ compares well with the results of Collins et al. (2011) for numerical simulations of intermediate and gravity regime impacts with fairly low cohesion. Note that while the expected impact angle distribution for primary craters is well known the distribution for secondary craters may differ from this. However we make no attempt to separate secondary craters and assume that all craters obey the expected impact angle distribution for primary craters in constructing Fig. 6.
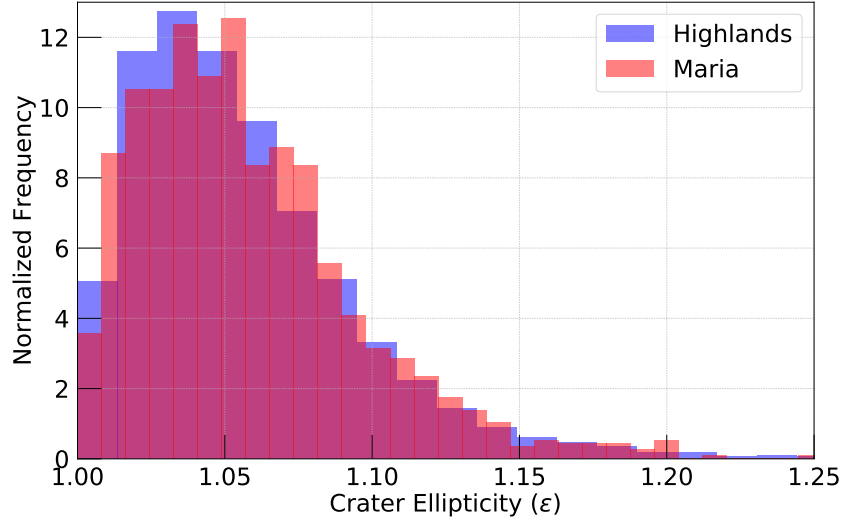
Figure 7: Crater ellipticity frequency distribution for the highlands and maria, normalized for ease of comparison. This reveals no notable differences between the two.

Note that there are a small number (less than 0.3%) of craters in our dataset that have ellipticities greater than 2, which may be a result of failures in the ellipse fitting, however excluding these craters does not significantly change the distribution at intermediate ellipticities. Similarly, around 3.5% of the craters have ellipticities of exactly 1. This is due to the limitations of pixel-based imaging noted above, even for a crater that is 100 pixels in the long dimension the lowest non-circular ellipticity it can have is 1.01, and is why the curve in Fig. 6 appears to intersect the y-axis at $80°$ rather than $90°$. As with the small tail of very high ellipticities excluding these does not significantly change the distribution at intermediate ellipticities.

As we have discussed, the relationship between impact angle and crater ellipticity depends on the strength of the target materials. As such it is interesting to compare the ellipticity distributions for different regions of the moon as differences may indicate differences in the properties of the surface materials. The most obvious distinction in surfaces on the Moon is between the mare and the highlands, and so in Fig. 7 we separate the crater populations of the mare and highlands and show the ellipticity frequency distributions for each. We can immediately see that the distributions are very similar except perhaps at the high ellipticities where the mare suffer from low number statistics. A statistical test confirms this, with a 2-sample Kolmogorov-Smirnov test returning $D = 0.146$, implying that both samples were probably drawn from the same distribution. This is not surprising since the highlands inevitably dominate any global crater distribution and we have already noted the compatibility between our total results and those of Bottke et al. (2000) that were derived only from mare craters.

The importance of the strength of the target materials relative to the gravity of the target body in determining the outcome of an impact depends on the scale of the impact, with material strength being more important at smaller scales. Collins et al. (2011) thus predicts that there should be some variation in the ellipticity distribution as a function of crater size. To examine this in Fig. 8 we show the ellipticity plotted against the crater diameter. Despite the large number of craters at diameters less than 20 km the proportion of craters with significant ellipticity appears lower than at larger
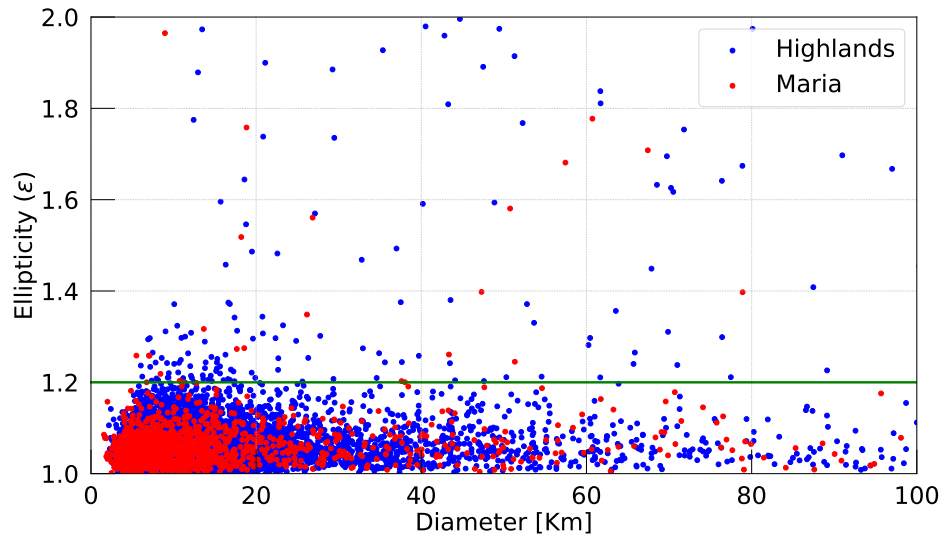
Figure 8: Ellipticity (ratio of major to minor axis) of the test set craters as a function of diameter, as predicted by MaskRCNN. We cut off this plot at 5 km, to limit the effects of artifically distorted small craters (with diameters comparable to pixel size) on the distribution. Red dots are for maria, and blue for highlands. The ratio of elliptical ($\epsilon > 1.2$) to non-elliptical craters predicted is consistent with the observational results of Bottke et al. (2000).
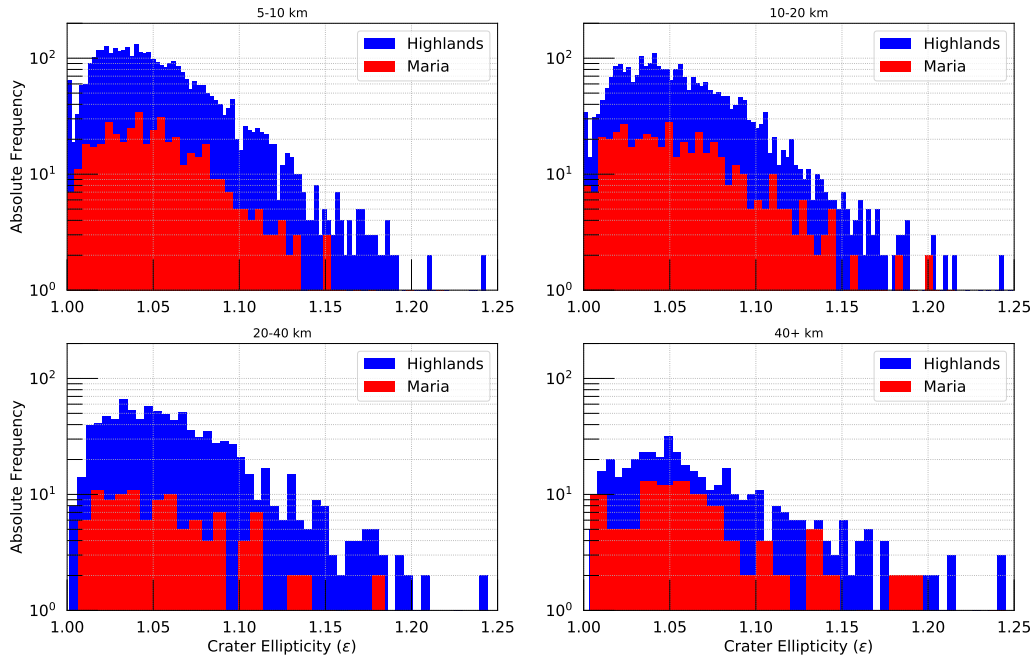
28

Figure 9: Histograms of ellipticity for craters in 4 different diameter ranges, 5-10 km, 10-20 km, 20-40 km and >40 km. The distributions for highlands craters are shown in blue, while the mare craters are shown in red. Note that the bin widths in the lower two panels are twice that in the upper panels due to the smaller numbers of craters at these larger sizes.

diameters. This is illustrated in more detail in Fig. 9, in which we show histograms of crater ellipticity for 4 different diameter ranges. We can see that as we move to larger diameters the distribution flattens, which indeed indicates that the proportion of elliptical craters increases with increasing crater diameter. This is different from the prediction of Collins et al. (2011) who suggest that on the Moon the proportion of elliptical craters should increase with increasing size for diameters larger than around 80 km, but that for craters between about 10 and 80 km the proportion of elliptical craters should be roughly flat or slightly decreasing with increasing size. This is clearly in need of further study and an excellent example of how the ability of machine learning to aid in the construction of large datasets can be useful.

3.4. Craters depth

A supplementary diagnostic for our model is the predicted crater depth distribution. This is possible to investigate since our dataset is generated from a global digital elevation map, and hence the brightness of the pixel is linear to the absolute elevation above or below a lunar reference radius. Since ultimately what the neural network is "seeing" is just the pixels brightness, it is of interest to check whether the machine has a bias to certain craters depth values (and hence pixel brightness gradient). We measured the depth by taking 4 profiles all the way across each crater, measuring the maximum depth difference after leveling the profiles slopes, and

taking the average of the four.

The ground truth and predicted craters depth frequency distributions are shown in Fig. 10. It is clear by looking at the overall shapes of two histograms that the craters found by our machine follow the same general depth distribution trends as the ground truth, with the two peaks coinciding at 0.5 km. The predicted distribution however has excesses of craters for most depths bins below 3 km, except around the central 0.5 km peak where the distribution plateaus. By examining Fig. 11b showing the depth vs diameter distribution of craters (explained below), we can associate craters shallower than 3 km to diameters less than 15 km, implying that the excess of craters in Fig. 10 is due to the "new" craters not present in the ground truth (discussed in section 3.2). The absence of an excess around 0.5 km however is puzzling. While it is conceivable that our machine is less efficient at detecting craters of this depth, this seems unlikely due to the presence of the shallower craters excess (hence with weaker pixel intensity gradients). We therefore prefer the alternative interpretation that the difference in the distributions shape is physical, and is due to the machine detecting a large number of flooded, eroded, or otherwise significantly altered craters not present in the GT. These are likely to have a different depth distribution than pristine craters, and the predicted distribution would be the sum of the two. Comparing detection biases as function of depth and crater flooding level is a major undertaking that we leave for future work. All of the above hint that our machine does not show

obvious bias with respect to depth, and is capable of detecting craters for a wide range of pixel brightness gradient. In consequence, our model can be used to constrain the depth distribution of unknown craters, in addition to their sizes and shapes.

The depth to diameter ratio can also provide insight into strength of the target material and the crater formation process. Figure 11 (left) shows our measurements for the depth versus diameter of craters in lunar highlands and mare. Shallow degraded and flooded craters are included in our data set, creating a wide range in depth for a particular diameter. Despite this, a clear transition in slope is observable near 18 km in diameter, corresponding with the transition from simple bowl shaped craters, to complex craters (Pike, 1977). Previous studies have found that simple craters have a depth to diameter slope of 0.1, while larger complex craters have have a depth to diameter slope of 0.05-0.005 (Pike, 1977). After filtering out degraded and flooded craters, we find a similar relationship (Fig. 11 (right)). It has also been predicted that the simple to complex crater transition occurs at different diameters in Mare (16 km) versus Highlands (21 km), due to differences in impact target strength (Pike, 1980). Our depth to diameter measurements suggest a transition diameter of $\sim$ 18 km in Maria and $\sim$ 20 km in Highlands consistent with (Pike, 1980). We also find that complex Mare craters are on average shallower than complex highland craters, confirming that target strength influences crater morphology.
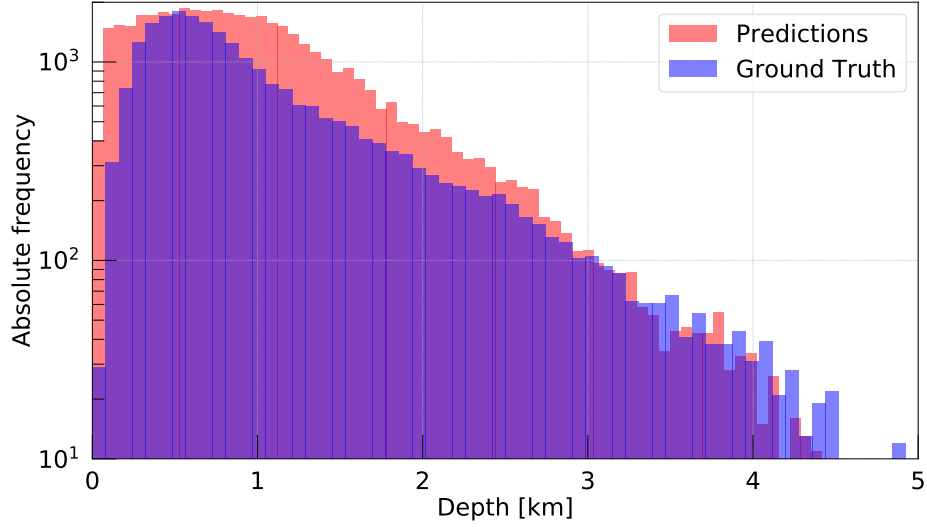
Figure 10: Depth frequency distribution for the ground truth and predictions of our test set (post-processing).

4. Summary & conclusions

In this work we trained the general computer vision framework MaskR-CNN to detect lunar craters in digital elevation maps, and extract their sizes and shapes accurately. Our model was able to detect on average 87% of all known craters in an image, and discovered $\sim 40\%$ more craters than previously catalogued. This method was moreover able to recover the observed size distribution of the test set at all diameters, improving over our previous DeepMoon algorithm for large craters. Finally, the main advantage of MaskRCNN was retrieving the crater shapes (through the instances masks) for free, without prior explicit training. Our retrieved shape distribution was found to be consistent with observational and numerical results, sug-
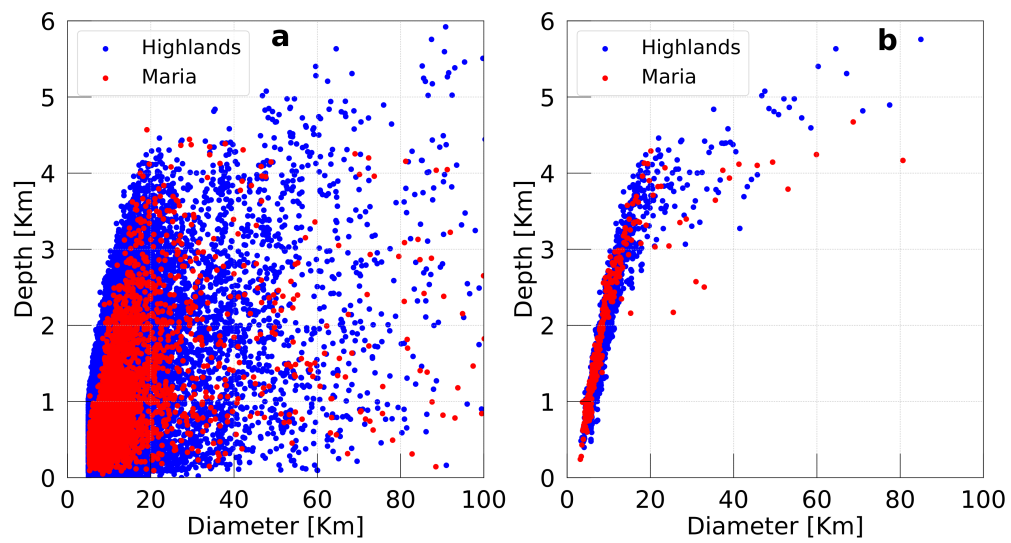
Figure 11: Depth-diameter diagram for the predicted craters, for highlands and maria separately. The knee at 20 Km diameters is due to the transition from simple to complex craters. Left hand side shows all of the craters, while RHS shows only craters we deemed pristine.

gesting that ∼2% of all craters have an ellipticity higher than 1.2. The efficiency of our approach allowed us to search for statistical differences in the size and shape distributions of highlands and maria craters, but we found these to be roughly identical. This is first such statistical study, thanks to automated shape retrieval. We finally found our model to perform well at all crater depths.

Acknowledgements

Appendix A.

We adopted the following Matterport MaskRCNN model parameters in our work:

IMAGES_PER_GPU = 2

IMAGE_MIN_DIM = 256

IMAGE_MAX_DIM = 512

RPN_ANCHOR_SCALES = (4, 8, 16, 32, 64)

TRAIN_ROIS_PER_IMAGE = 600

RPN_NMS_THRESHOLD = 0.7

MEAN_PIXEL = [165.32, 165.32, 165.32]

LEARNING_RATE = 1e-3

USE_MINI_MASK = True

MAX_GT_INSTANCES = 400

DETECTION_MAX_INSTANCES = 400

Appendix B.

Our Python preprocessing routine is provided here for reproductibility. It is applied to three identical channels in the present case.

```python
import cv2
grid_size = 8
def rgb_clahe_justl(in_rgb_img):
    bgr = in_rgb_img[:,:,[2,1,0]] # switch R and B (RGB → BGR)
    lab = cv2.cvtColor(bgr, cv2.COLOR_BGR2LAB)
    clahe = cv2.createCLAHE(clipLimit=2.0, tileGridSize=(grid_size,grid_size))
    return clahe.apply(lab[:,:,0])
```

References

Barlow, N. G., Aug. 1988. Crater size-frequency distributions and a revised Martian relative chronology. Icarus75, 285–305.

Bottke, W. F., Love, S. G., Tytell, D., Glotch, T., May 2000. Interpreting the Elliptical Crater Populations on Mars, Venus, and the Moon. Icarus145, 108–121.

Bradski, G., 2000. The OpenCV Library. Dr. Dobb's Journal of Software Tools.

Christoff, N., Manolova, A., Jorda, L., Mari, J., Sep. 2018. Morphological crater classification via convolutional neural network with application on mola data. In: ANNA '18; Advances in Neural Networks and Applications 2018. pp. 1–5.

Collins, G. S., Elbeshausen, D., Davison, T. M., Robbins, S. J., Hynek, B. M., Oct. 2011. The size-frequency distribution of elliptical impact craters. Earth and Planetary Science Letters 310, 1–8.

Elbeshausen, D., Wünnemann, K., Collins, G. S., Dec. 2009. Scaling of oblique impacts in frictional targets: Implications for crater size and formation mechanisms. Icarus204, 716–731.

Elbeshausen, D., Wünnemann, K., Collins, G. S., Nov. 2013. The transition from circular to elliptical impact craters. Journal of Geophysical Research (Planets) 118, 2295–2309.

Gault, D. E., Wedekind, J. A., 1978. Experimental studies of oblique impact. In: Lunar and Planetary Science Conference Proceedings. Vol. 9 of Lunar and Planetary Science Conference Proceedings. pp. 3843–3875.

Girshick, R., Donahue, J., Darrell, T., Malik, J., Nov. 2013. Rich feature hierarchies for accurate object detection and semantic segmentation. arXiv e-prints.

Harris, A., 09 2018. Lunar crater identification using convolutional neural net with eigenfunction expansion pooling layer.

He, K., Gkioxari, G., Dollár, P., Girshick, R., Mar. 2017. Mask R-CNN. arXiv e-prints.

Head, J. W., Fassett, C. I., Kadish, S. J., Smith, D. E., Zuber, M. T., Neumann, G. A., Mazarico, E., Sep. 2010. Global Distribution of Large Lunar Craters: Implications for Resurfacing and Impactor Populations. Science 329, 1504.

Herrick, R. R., Forsberg-Taylor, N. K., Nov. 2003. The shape and appearance of craters formed by oblique impact on the Moon and Venus. Meteoritics and Planetary Science 38, 1551–1578.

Holo, S. J., Kite, E. S., Robbins, S. J., Aug. 2018. Mars obliquity history constrained by elliptic crater orientations. Earth and Planetary Science Letters 496, 206–214.

Liu, Z., Yue, Z., Michael, G., Gou, S., Di, K., Sun, S., Liu, J., Sep. 2018. A global database and statistical analyses of (4) Vesta craters. Icarus311, 242–257.

LOLA Team, Kaguya Team, 2015. LRO LOLA and Kaguya Terrain Camera DEM merge 60N60S 512ppd (59m). `https://astrogeology.usgs.gov/search/map/Moon/LRO/LOLA/Lunar_LRO_LrocKaguya_DEMmerge_60N60S_512ppd`.

Melosh, H. J., 1989. Impact cratering: A geologic process.

Pike, R. J., 1977. Apparent depth/apparent diameter relation for lunar craters. In: Merril, R. B. (Ed.), Lunar and Planetary Science Conference Proceedings. Vol. 8 of Lunar and Planetary Science Conference Proceedings. pp. 3427–3436.

Pike, R. J., Jul. 1980. Formation of complex impact craters - Evidence from Mars and other planets. Icarus43, 1–19.

Pilu, M., Fitzgibbon , A., Fisher, R., 1996. Ellipse-specific direct least-square fitting. Proceedings of 3rd IEEE International Conference on Image Processing.

Povilaitis, R. Z., Robinson, M. S., van der Bogert, C. H., Hiesinger, H., Meyer, H. M., Ostrach, L. R., Nov. 2018. Crater density differences: Exploring regional resurfacing, secondary crater populations, and crater saturation equilibrium on the moon. Planet. Space Sci.162, 41–51.

Robbins, S. J., Apr. 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.

Robbins, S. J., Hynek, B. M., May 2012. A new global database of Mars impact craters $\geq$1 km: 1. Database creation, properties, and parameters. Journal of Geophysical Research (Planets) 117 (E5), E05004.

Schultz, P. H., Lutz-Garihan, A. B., 1982. Grazing impacts on mars: A record of lost satellites. Journal of Geophysical Research: Solid Earth 87 (S01), A84–A96.
URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/JB087iS01p00A84`

Shoemaker, E. M., 1962. Interpretation of lunar craters. In: Kopal, Z. (Ed.), Physics and Astronomy of the Moon. Academic Press, New York, pp. 283 – 359.

Silburt, A., Ali-Dib, M., Zhu, C., Jackson, A., Valencia, D., Kissin, Y., Tamayo, D., Menou, K., Jan. 2019. Lunar crater identification via deep learning. Icarus 317, 27–38.